RENDICONTI DEL SEMINARIO MATEMATICO

Università e Politecnico di Torino

Splines, Radial Basis Functions and Applications

CONTENTS

Volume 61, N. 3 **2003**

Preface

This special issue of *Rendiconti del Seminario Matematico dell'Università e del Politecnico di Torino* contains the invited papers presented at the Workshop *Giornate di studio su funzioni spline e funzioni radiali*, held at the University of Torino, on February 5-7, 2003, in conclusion of the GNCS project *Funzioni spline e funzioni radiali: applicazioni a problemi integrali e differenziali*, Catterina Dagnino coordinator. The Workshop was supported by GNCS *(Gruppo Nazionale per il Calcolo Scientifico)*.

As far as the program was concerned, there were three invited speakers: Armin Iske (Munich University of Technology, Germany), George Micula (Babes-Bolyai University, Cluj-Napoca, Romania) and Paul Sablonnière (Institut National des Sciences Appliquées, Rennes, France).

In addition, speakers from several Italian Universities provided papers for inclusion in the current issue, which were all refereed.

The contributions collected here deal with different aspects of numerical approximation based on splines and radial basis functions: a variational approach to spline functions theory, quadratic spline quasi-interpolants on bounded domains of $\mathbb{R}^d$, $d = 1, 2, 3$, basics and advanced topics on radial basis functions and meshfree methods for transport problems, a deficient spline collocation method for certain differential and integral equations with delay, some recent results on a new class of bivariate refinable functions, optimal nodal splines and their applications to integral problems, a local hybrid scheme for scattered data approximation, optimal center locations for radial basis function interpolation, a collocation method for linear fourth order boundary value problems and closed spline curves bounding maximal area.

The invited papers appear according to their order during the oral presentation, the other ones in alphabetical order according to the author's name.

The guest editors are deeply grateful to the authors who contributed to this issue. Moreover they thank *Seminario Matematico*, for taking care of the publication and GNCS for its support.

Catterina Dagnino, Vittoria Demichelis

Rend. Sem. Mat. Univ. Pol. Torino
Vol. 61, 3 (2003)
Splines and Radial Functions

G. Micula*

A VARIATIONAL APPROACH TO SPLINE FUNCTIONS THEORY

Abstract. Spline functions have proved to be very useful in numerical analysis, in numerical treatment of differential, integral and partial differential equations, in statistics, and have found applications in science, engineering, economics, biology, medicine, etc. It is well known that interpolating polynomial splines can be derived as the solution of certain variational problems. This lecture presents a variational approach to spline interpolation. By considering quite general variational problems in abstract Hilbert spaces setting, we derive the concept of "*abstract splines*". The aim of this leture is to present a sequence of theorems and results starting with Holladay's classical results concerning the variational property of natural cubic splines and culminating in some general variational approach in abstract splines results. The whole exposition of this lecture is based on the papers of Champion, Lenard and Mills [24], [25].

1. Introduction

It is more than 50 years since I. J. Schoenberg ([56], 1946) introduced "spline functions" to the mathematical literature. Since then, splines, have proved to be enormously important in various brances of mathematics such as approximation theory, numerical analysis, numerical treatment of differential, integral and partial differential equations, and statistics. Also, they have become useful tools in field of applications, especially CAGD in manufacturing, in animation, in tomography, even in surgery.

Our aim is to draw attention to a variational approach to spline functions and to underline how a beautiful theory has evolved from a simple classical interpolation problem. As we will show, the variational approach gives a new way of thinking about splines and opens up directions for theoretical developments and new applications.

Despite of so many results, this topics is not mentioned in many relevant texts on numerical analysis or approximation theory: even books on splines tend to mention the variational approach only tangentially or not at all.

Even though, there are recently published a few papers which underline the vari-

*The author is very grateful to Professor Paul Sablonière for helpful comments and additional references during the preparation of final form of this lecture. Extremly indebted is the author to Professors R. Champion, C. T. Lenard and T. M. Mills for the kindness to send him the papers [24] and [25] on which basis this lecture has been preparated.

ational aspects of splines, and we mention the papers of Champion, Lenard and Mills ([25], 2000, [24], 1996) and of Beshaev and Vasilenko ([17], 1993).

The contain of this lecture is a completation of the excellent expository paper of Champion, Lenard and Mills [25]. We shall keep also the definitions and the notation from these papers.

The theorems and results of increasing generality or complexity which culminate in some general and elegant abstract results are not necessarily chronological.

2. Preliminaries

Notations:

$\mathbb{R}$ – the set of real numbers

$I : [a, b] \subset \mathbb{R}$

$\mathcal{P}_m := \{p \in \mathbb{R} \to \mathbb{R},\ p$ is real polynomial of degree $\leq m,\ m \in \mathbb{N}\}$

$H^m(I) := \{x : I \to \mathbb{R},\ x^{(m-1)}$ abs. cont. on $I,\ x^{(m)} \in L^2(I),\ m \in \mathbb{N},$ given$\}$

If we define an inner product on $H^m(I)$ by

$$(x_1, x_2) := \int_I \sum_{j=0}^{m} x_1^{(j)}(t) x_2^{(j)}(t) dt$$

then $H^m(I)$ becomes a Hilbert space.

If X is a linear space, then θ_X will denote the zero element of X.

DEFINITION 1. *Let $a = t_0 < t_1 < \cdots < t_n < t_{n+1} = b$ be a partition of I. The function $s : I \to \mathbb{R}$ is a polynomial spline of degree m with respect to this partition if*

- $s \in C^{m-1}(I)$
- *for each* $i \in \{0, 1, \ldots, n\}$, $s|_{[t_i, t_{i+j}]} \in \mathcal{P}_m$

The interior points $\{t_1, t_2, \ldots, t_n\}$ *are known as "knots".*

Natural cubic splines

Suppose that $t_1 < t_2 < \cdots < t_n$ and $\{z_1, z_2, \ldots, z_n\} \subset \mathbb{R}$ are given. The classical problem of interpolation is to find a "suitable" function Φ which interpolates the data point (t_i, z_i), $1 \leq i \leq n$, that is:

$$\Phi(t_i) = z_i, \quad 1 \leq i \leq n.$$

Classical approaches developed by Lagrange, Hermite, Cauchy and others rely on choosing Φ to be some suitable *polynomial*. But are there better functions for solving this interpolation problem? The first answer to this question can be found in a result which was proved by Holladay [38] in 1957.

THEOREM 1 (HOLLADAY, 1957). *If*

- $X := H^2(I)$,
- $a \leq t_1 < \cdots < t_n \leq b;\ n \geq 2$,
- $\{z_1, z_2, \ldots, z_n\} \subset \mathbb{R}$, *and*
- $I_n := \{x \in X : x(t_i) = z_i,\ 1 \leq i \leq n\}$,

then $\exists!\ \sigma \in I_n$ *such that*

$$\int_I [\sigma^{(2)}(t)]^2 dt = \min \left\{ \int_I [x^{(2)}(t)]^2 dt : x \in I_n \right\} \tag{1}$$

Furthermore,

- $\sigma \in C^2(I)$,
- $\sigma|_{[t_i, t_{i+1}]} \in \mathcal{P}_3$ *for* $1 \leq i \leq n-1$,
- $\sigma|_{[a, t_1]} \in \mathcal{P}_1$ *and* $\sigma|_{[t_n, b]} \in \mathcal{P}_1$.

From (1) we conclude that σ is an optimal interpolating function – "optimal", in the sense that it minimize the functional $\int_I [x^{(2)}(t)]^2 dt$ over all functions in I_n. The theorem goes on to state that σ *is a cubic spline function in the meaning of Schoenberg definition (1946)*. As σ is linear outside $[t_1, t_n]$ it is called "*natural cubic spline*".

So, in a technical sense, we have found functions which are better than polynomials for solving the interpolation problem. Holladay's theorem is most surprising not only because its proof is quite elementary, relying on nothing more complicated than integration by parts, but it shows the intrinsec aspect of splines as solution of a variational problem (1) that has been a starting point to develop a variational approach to splines.

It is natural to ask: "*Why would one choose to minimize* $\int_I [x^{(2)}(t)]^2 dt$*?*"

For three reasons:

i) The curvature of function σ is $\sigma^{(2)}/(1 + \sigma'^2)^{3/2}$ and so the natural cubic spline is the best in the sense that it approximates the interpolating function with minimum total curvature if σ' is small.

ii) The second justification is that the natural cubic spline approximates the solution of a problem in physics, in which a uniform, thin, elastic, linear bar is deformed to interpolate the knots specified in absence of external forces. This shape of such a bar is governed by a minimum energy in this case minimum elastic potential energy. The first order approximation to this energy is proportional to the functional (1). Hence the term natural spline is borrowed the term "spline" from the drafting instrument also known as a spline.

iii) When presented with a set of data points (t_i, z_i), $1 \le i \le n$, a statistician can find a regression line which is the line of best fit in the least squares sense. This line is close to the data points Holladay's theorem shows that σ minimizes $\int_I [x^{(2)}(t)]^2 dt$ while still interpolating the data. We could say that σ is an interpolating function which is "close to a straight lines" in that it minimizes this integral.

Thus, linear regression gives us

a straight line passing close to the points

whereas Holladay's result gives a curve σ which is

close to a straight line but passing through the points.

3. More splines

As we shall see, the Holladay's theorem was the starting point in developing the variational approach to splines. In what follows we shall describe a few of the many important generalizations and extensions of Holladay's theorem.

D^m-splines

The next step was taken in 1963 by Carl de Boor [18] with the following result.

THEOREM 2 (C. DE BOOR, 1963). *If*

- $X := H^m(I)$,
- $a \le t_1 < t_2 < \cdots < t_n \le b;\ n \ge m$,
- $\{z_1, z_2, \ldots, z_n\} \subset \mathbb{R}$ *and*
- $I_n := \{x \in X :\ x(t_i) = z_i,\ 1 \le i \le n\}$

then $\exists!\ \sigma \in I_n$ *such that*

$$\int_I [\sigma^{(m)}(t)]^2 dt = \min \left\{ \int_I [x^{(m)}(t)]^2 dt :\ x \in I_n \right\}$$

Furthermore,

- $\sigma \in C^{2m-2}(I)$,
- $\sigma|_{[t_i, t_{i+1}]} \in \mathcal{P}_{2m-1},\ 1 \le i \le n-1$, *and*
- $\sigma|_{[a,t_1]} \in \mathcal{P}_{m-1}$ *and* $\sigma|_{[t_n,b]} \in \mathcal{P}_{m-1}$.

The function σ was called D^m*-spline* because it minimizes $\int_I (D^m x)^2 dt$, as x varies over I_n. The function σ is called the interpolating *natural spline function of odd degree*.

Clearly if we let $m = 2$ in de Boor result, then we obtain Holladay result. For the even degree splines, such result was given by P. Blaga and G. Micula in 1993 [49], following the ideas of Laurent [44].

Trigonometric splines

In 1964, Schoenberg [57] changed the setting of the interpolation problem from the interval $[a, b]$ to the unit circle: that is, from a non-periodic setting to a periodic setting.

Similarly, let $H^k_{2\pi}([0, 2\pi))$ denote the following space of 2π-periodic functions:

$$H^k_{2\pi}([0, 2\pi)) := \{x : [0, 2\pi) \to \mathbb{R} : \; x - 2\pi - \text{periodic},$$

$$x^{(k-1)} \text{ abs. cont. on } [0, 2\pi), \; x^{(k)} \in L^2_{2\pi}([0, 2\pi))\}.$$

THEOREM 3 (SCHOENBERG, 1964). *If*

- $X := H^{2m+1}_{2\pi}([0, 2\pi))$
- $0 \leq t_1 < t_2 < \cdots < t_n < 2\pi, \; n > 2m + 1$
- $\{z_1, z_2, \ldots, z_n\} \subset \mathbb{R}$ *and*
- $T : X \to L^2_{2\pi}([0, 2\pi))$, *where* $T := D(D^2 + 1^2) \ldots (D^2 + m^2)$,

then $\exists! \; \sigma \in I_n$ *such that*

$$\int_0^{2\pi} [T(\sigma)(t)]^2 dt = \min \left\{ \int_0^{2\pi} [T(x)(t)]^2 dt : \; x \in I_n \right\}.$$

The optimal interpolating function σ is called the *trigonometric spline*. Schoenberg defined a trigonometric spline as a smooth function which in a particular piecewise trigonometric polynomial manner. He shows that trigonometric splines, so defined, provide the solution of this variational problem.

Note that the differential operator T has as $KerT$ all the trigonometric polynomials of order m, that is, of the form:

$$x(t) = a_0 + \sum_{j=1}^{m} (a_j \cos jt + b_j \sin jt).$$

g-splines

Just over 200 years ago in 1870 Lagrange has constructed the polynomial of minimal degree such that the polynomial assumed prescribed values at given nodes and the derivatives of certain orders of the polynomial also assumed prescribed values at the nodes.

In 1968, Schoenberg [58] extended the idea of Hermite for splines. To specify that the orders of the derivatives specified may vary from node to node we introduce an *incidence matrix* E. As usual, let $I := [a, b]$ be an interval partitioned by the nodes $a \le t_1 < t_2 < \cdots < t_n \le b$. Let l be the maximum of the orders of the derivatives to be specified at the nodes. The incidence matrix E is defined by:

$$E := (e(i, j) :\ 1 \le i \le n,\ 0 \le j \le l) =: (e(i, j))$$

where each $e(i, j)$ is 0 or 1. Assume also that each row of E and the last column of E contain a 1.

DEFINITION 2. *If $m \ge 1$ is an integer, we will say that the incidence matrix $E = (e(i, j))$ is m-poised with respect to $t_1 < t_2 < \cdots < t_n$ if*

- $P \in \mathcal{P}_{m-1}$ *and*
- $e(i, j) = 1 \Rightarrow P^{(j)}(t_i) = 0$

together imply that $P \equiv 0$.

Now we can state Schoenberg's result.

THEOREM 4 (SCHOENBERG, 1968). *If*

- $X := H^m(I)$
- $a \le t_1 < t_2 < \cdots < t_n \le b$
- *E is an m-poised incidence matrix of dimension* $n \times (l + 1)$
- $l < m \le \sum_i \sum_j e(i, j)$
- $\{z_{ij} :\ e(i, j) = 1\} \subset \mathbb{R}$ *and*
- $I_n := \{x \in X :\ x^{(j)}(t_i) = z_{ij}$ *if* $e(i, j) = 1\}$

then $\exists!\ \sigma \in I_n$ *such that*

$$\int_I [\sigma^{(m)}(t)]^2 dt = \min \left\{ \int_I [x^{(m)}(t)]^2 dt :\ x \in I_n \right\}.$$

Schoenberg called the function σ as g-spline from "generalized-splines". Better may have been H-splines after Hermite or HB-splines after Hermite and Birkhoff.

Again, Schoenberg has defined g-splines as smooth piecewise polynomials where the smoothness is governed by E and then he proved that g-splines solves the above variational problem.

L-splines

In 1967, Schultz and Varga [59] gave a major extension of the D^m-splines. Instead of the m-order derivative, operator D^m they considered a linear differential operator L creating a theory of so called L-splines. We shall state only one simple consequence of the many results of Schultz and Varga.

THEOREM 5 (SCHULTZ AND VARGA, 1967). *If*

- $X := H^m(I)$
- $a \leq t_1 < t_2 < \cdots < t_n \leq b;\ n \geq m$
- $\{z_1, z_2, \ldots, z_n\} \subset \mathbb{R}$
- $I_n := \{x \in X : x(t_i) = z_i,\ 1 \leq i \leq n\}$
- $L : X \to L^2(I)$*, so that* $L[x](t) := \sum_{j=0}^{m} a_j(t) D^j x(t)$*, where* $a_j \in C^j(I)$*,* $0 \leq j \leq m$*, and* $\exists\, \omega > 0$ *such that* $a_m(t) \geq \omega > 0$ *on* I *and*
- L *has Pólya's property* W *on* I

then $\exists !\, \sigma \in I_n$ *such that*

$$\int_I [L[\sigma](t)]^2 dt = \min \left\{ \int_I [L[x](t)]^2 dt : x \in I_n \right\}.$$

Clearly complexity is increasing with generality.

We note that L has Pólya's property W on I if $L[x] = 0$ has m solutions $x_1, x_2, \ldots, x_m$ such that, for all $t \in I$ and for all $k \in \{1, 2, \ldots, m\}$

$$\det \begin{bmatrix} x_1(t) & x_2(t) & \ldots & x_m(t) \\ Dx_1(t) & Dx_2(t) & \ldots & Dx_m(t) \\ \ldots & \ldots & \ldots & \ldots \\ D^{k-1}x_1(t) & D^{k-1}x_2(t) & \ldots & D^{k-1}x_m(t) \end{bmatrix} \neq 0.$$

The relevance of Pólya's property W is contained in the following sentence. To say that L has Pólya's property W on I implies that, if $L[x] = 0$ and x has m or more zeros on I, then $x \equiv 0$.

The optimal function σ is known as an *L-spline*.

If $L \equiv D^m$ we obtain the D^m-spline: so this is a major extension of previously stated results.

Schultz and Varga have defined an L-spline to be a smooth function constructed in a piecewise manner, where each piece is a solution of the differential equation $L^*Lx = 0$ where L^* is the formal adjoint of the operator L.

A consequence of their paper is that L-spline provide a solution of the above variational problem.

REMARK 1. - The result of Schultz and Varga was proved in 1964 by Ahlberg, Nilson and Walsh [2]. They called σ a "generalized splines".

- The above result also follows from a paper of de Boor and Lynch [20] published in 1966.

- Perhaps the first paper along these lines of replacing the operator D^m by a more general differential operator was given by Greville [36] also in 1964. Unfortunately this often cited technical report was never published. Greville illustrates his method with an application to the classical numerical problem of interpolating mortality tables. Schultz and Varga applied their ideas to the numerical analysis of nonlinear two-point boundary value problems.

- Prenter [53] and Micula [50] are two of the few text books which touch this topic.

Lg-splines

Schoenberg extended the concept of D^m-splines to allow interpolation conditions of the Hermite type: this leads to g-splines. Schultz and Varga (and others) extended the concept of D^m-spline in a different direction by replacing the differential operator D^m by a more general operator: this leads to L-splines. The question is if one could combine both these extensions. In 1969 Jerome and Schumaker [31] combined these two extensions together in a very effective manner. One of their results is the following:

THEOREM 6 (JEROME AND SCHUMAKER, 1969). *If*

- $X := H^m(I)$
- $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ *is a set of linearly independent, continuous linear functionals on X*
- $\{z_1, z_2, \ldots, z_n\} \subset \mathbb{R}$
- $I_n := \{x \in X : \lambda_i(x) = z_i,\ 1 \le i \le n\}$
- $L : X \to L^2(I)$ *so that* $L[x](t) = \sum_{j=0}^{m} a_j(t) D^j x(t)$, $a_j \in C^j(I)$, $0 \le j \le m$, *and* $\exists\ \omega > 0$ *such that* $a_m(t) \ge \omega > 0$ *on* I *and*
- $ker L \cap \{x \in X : \lambda_i(x) = 0,\ 1 \le i \le n\} = \{\theta_X\}$

then $\exists!\ \sigma \in I_n$ *such that*

$$\int_I [L[\sigma](t)]^2 dt = \min \left\{ \int_I [L[x](t)]^2 dt : x \in I_n \right\}.$$

The optimal function σ is called the *Lg-spline*. The hypothesis about Pólya's property W in Theorem 5 has with the more functional-analytic flavour. Jerome and Schumaker allow interpolation conditions for the more general form $\lambda_i(x) = z_i$, $1 \le i \le n$, where λ_i $(1 \le i \le n)$ are continuous linear functionals on X. This idea could cover also others conditions like $\int_{t_i}^{t_{i+1}} x(t)dt = z_i$, $1 \le i \le n$. We note also that they and Laurent [44], pp. 225-226 replace the conditions $\lambda_i(x) = z_i$ by $\underline{z}_i \le \lambda_i(x) \le \overline{z}_i$, where $\underline{z}_i$ and $\overline{z}_i$ $(i = 1, 2, \dots, n)$ are given real numbers with $\underline{z}_i \le \overline{z}_i$.

pLg-splines

For $1 < p < \infty$ we define the space $H^m(I^p)$ of functions by:

$$H^{m,p}(I) := \{x : I \to \mathbb{R} : \; x^{(m-1)} \text{ abs. cont.}, x^{(m)} \in L^p(I)\}$$

With a norm on $H^{m,p}(I)$ defined by:

$$\|x\|_{m,p} := \sum_{j=0}^{m} |x^{(j)}(a)| + \left(\int_I |x^{(m)}(t)|^p dt \right)^{1/p}$$

the $H^{m,p}(I)$ is a Hilbert space.

In 1978 Copley and Schumaker [26] established the following result:

THEOREM 7 (COPLEY AND SCHUMAKER, 1978). *If*

- $X := H^{m,p}(I), \;\; p > 1$
- $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ *is a set of linearly independent continuous linear functionals on* X
- $\{z_1, z_2, \dots, z_n\} \subset \mathbb{R}$
- $I_n := \{x \in X : \; \lambda_i(x) = z_i, \;\; 1 \le i \le n\} \neq \emptyset$
- $L : X \to L^p(I)$ *so that* $L[x](t) = \sum_{j=0}^{m} a_j(t) D^j x(t)$, $a_j \in C^j(I)$, $0 \le j \le m$ *and* $\exists\, \omega > 0$ *such that* $a_m(t) \ge \omega > 0$ *on* I, *and*
- $ker L \cap \{x \in X : \; \lambda_i(x) = 0, \;\; 1 \le i \le n\} = \{\theta_X\}$

then $\exists!\, \sigma \in I_n$ *such that:*

$$\int_I |L[\sigma](t)|^p dt = \min \left\{ \int_I |L[x](t)|^p dt \, : \; x \in I_n \right\}.$$

The optimal function σ is called a *pLg-spline*. For the first time, in this paper Copley and Schumaker have defined a pLg-spline to be a solution of the variational interpolation problem. One of the main problems that they investigated is to determine the structure of such splines. Can they be constructed in a piecewise manner? The complexity of their answer compensates the simplicity of their definition on a pLg-spline. In fact, Copley and Schumaker investigated more general interpolation problems. For example, they consider sets of linear functionals $\{\lambda_\alpha : \alpha \in A\}$ where the index set A may be infinite, and also many extremly important examples.

Vector-valued Lg-splines

The following extension have come from researches in electrical engineering. In 1979 Sidhu and Weinert [60] consider the problem of simultaneous interpolation, that is, a method by which one could interpolate several functions at once.

THEOREM 8 (SIDHU AND WEINERT, 1979). *If*

- $r \geq 1,\ n_1 \geq 0, \ldots, n_r \geq 0$ *are fixed integers*
- $X := H^{n_1}(I) \times H^{n_2}(I) \times \cdots \times H^{n_r}(I)$
- $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ *is a set of linearly independent continuous linear functionals on* X
- $\{z_1, z_2, \ldots, z_n\} \subset \mathbb{R}$
- $I_n := \{x \in X : \lambda_i(x) = z_i,\ 1 \leq i \leq n\}$
- $L : X \to L^2(I) \times \cdots \times L^2(I)$ *(an r-fold product), where*

$$L[x](t) := \left(\sum_{j=1}^{r} L_{ij}[x_j](t) : i = 1, 2, \ldots, r \right)',$$

$$L_{ij} := \sum_{k=0}^{n_j} a_{ijk}(t) D^k;\ a_{ijn_j} = \delta_{ij};\ a_{ijk} \in C^k(I),\ 0 \leq k \leq n_j, \text{ and}$$

- $ker L \cap \{x \in X : \lambda_i(x) = 0,\ 1 \leq i \leq n\} = \{\theta_X\}$

then $\exists!\ \sigma \in X$ *such that:*

$$\int_I (L[\sigma](t))' L[\sigma](t) dt = \min \left\{ \int_I (L[x](t))' L[x](t) dt : x \in I_n \right\}.$$

(Here A' indicates the transpose of the matrix or vector A.)

The optimal interpolating vector σ is known as a *vector-valued Lg-spline*. The authors have defined a vector-valued Lg-spline to be the solution of a variational interpolation problem, proved the existence-uniqueness theorem and then discussed an algorithm for calculating such splines in the special case that the functional λ_i are of extended Hermite-Birkhoff type.

Thin plate splines

So far we have been considering the problem of interpolating functions of a single variable. In 1976, Jean Duchon [29], [23] - [27] - [28] developed a variational approach to interpolating functions of several variables. We will state his result only for functions of two variables. We denote an arbitrary element of $\mathbb{R}^2$ by $t = (\xi_1, \xi_2)$, $\|t\|^2 := \xi_1^2 + \xi_2^2$ and the set of linear polynomials by:

$$\mathcal{P}_1 := \{p_1(t) = a_0 + a_1\xi_1 + a_2\xi_2 : \{a_0, a_1, a_2\} \subset \mathbb{R}\}.$$

THEOREM 9 (DUCHON, 1976). *If*

- $X := H^2(\mathbb{R}^2)$,
- $\{t_1, t_2, \dots, t_n\} \subset \mathbb{R}^2$ *such that if* $p_1 \in \mathcal{P}_1$ *and* $p_1(t_1) = \cdots = p_1(t_n) = 0$, *then* $p_1 \equiv 0$,
- $\{z_1, z_2, \dots, z_n\} \subset \mathbb{R}$,
- $I_n := \{x \in X : x(t_i) = z_i, \ 1 \le i \le n\}$ *and*
- $J : X \to \mathbb{R}$ *such that*

$$J(x) := \iint\limits_{\mathbb{R}^2} \left[\left(\frac{\partial^2 x}{\partial \xi_1^2} \right) + 2 \left(\frac{\partial^2 x}{\partial \xi_1 \partial \xi_2} \right)^2 + \left(\frac{\partial^2 x}{\partial \xi_2^2} \right) \right] d\xi_1 d\xi_2$$

then $\exists! \, \sigma \in I_n$ *such that*

$$J(\sigma) = \min\{J(x) : x \in I_n\}.$$

Furthermore, $\forall \, t \in \mathbb{R}^2$

$$\sigma(t) = \sum_{j=1}^{n} \mu_i \|t - t_i\|^2 \ln \|t - t_i\| + p_1(t)$$

where $p_1 \in \mathcal{P}_1$ *and* $(\forall \, q \in \mathcal{P}_1)$, $\left(\sum_{i=1}^{n} \mu_i q(t_i) = 0 \right)$.

The optimal function σ is known as a "*thin plate spline*". The dramatic aspect of this result is the form of the spline σ: it is no a piecewise polynomial function.

This two-dimensional result appeared almost 20 years after Holladay's one-dimensional result. The delay is not so surprising. Holladay's proof involves nothing more complicated than integration by parts whereas Duchon's paper uses tempered distribution, Radon measure and other tools from functional analysis.

REMARK 2. i) A more elementary approach to Duchon's result is outlined in Powell [52].

ii) Duchon was not the first person to investigate the multivariate problem. In 1972 the work of two aircraft engineers Harder and Desmarais [37] approached this problem from an applied point of view. In 1974 Fisher and Jerome [32] addressed the multivariate problem. In 1970, J. Thomann [62] in his doctoral thesis considered a variational approach to interpolation on a rectangle or on a disk in $\mathbb{R}^2$. The book by Ahlberg, Nilson and Walsh [3] also deals with multivariate problems, but from a point of view which is essentially univariate.

Yet more splines

The overture of splines could be continued. There are other many splines associated with some variational interpolation problems and for each case we could state a theorem similar to those above. We shall only nominate they:

Λ-*splines* (1972, Jerome and Pierce [41])

LMg-splines (1979, R. J. P. de Figueiredo [30])

ARMA-splines (1979, Weinert, Sesai and Sidhu [67])

Spherical splines (1981, Freeden, Scheiner and Franke [33])

PDLg-splines (1990, R. J. P. de Figueiredo and Chen [31])

Polyharmonic splines (1990, C. Rabut [54])

Vector splines (1991, Amodei and Benbourhim [5])

Hyperspherical splines (1994, Taijeron, Gibson and Chandler [61]).

4. Abstract splines

The statements of the above theorems were becoming quite long and complicated. But, there is a general abstract result which captures the essence of most of them. The following result is attributed to M. Atteia [10], [11],[12] - [15], and it relates to following diagram:

$$\begin{array}{ccc} X & \xrightarrow{T} & Y \\ {\scriptstyle A}\downarrow & & \\ Z & & \end{array}$$

THEOREM 10 (ATTEIA, 1992). *If*

- *X, Y, Z are Hilbert spaces,*
- *T, A are continuous linear surjections,*
- $z \in Z$
- *$ker T + ker A$ is closed in X,*
- *$ker T \cap ker A = \{\theta_X\}$ and*

- $I(z) = \{x \in X : \; Ax = z\}$

then $\exists! \; \sigma \in I(z)$ *such that:*

$$\|T\sigma\|_Y = \min\{\|Tx\|_Y : \; x \in I(z)\}.$$

The optimal σ is known as a *variational interpolating spline*.

To illustrate that this theorem reflects the essence of the most above results, let us see how it generalizes Theorem 1 of Holladay. Put $X = H^2(I)$, $Y = L^2(I)$, $Z = \mathbb{R}^n$, $Tx := x^{(2)}$, $Ax := (x(t_1), x(t_2), \ldots, x(t_n))$. All the hypotheses of Atteia's theorem are satisfied. Atteia's theorem does not cover all the above results, e.g. Theorem 7 which deals with pLg-splines.

- An equivalent result to Atteia's theorem is found in the often cited, but unfortunately never published, report by Golomb [34] in 1967.
- The essential ideas also can be found in Anselone and Laurent [6] in 1968 and in the classic book by Laurent [44], entitled Approximation et Optimisation (Herman, Paris, 1972).

There are important remarks to be made about this theorem.

1. The role of the condition about $kerT + kerA$ is to ensure the existence of σ whereas the role of the condition $kerT \cap kerA$ is to ensure the uniqueness of σ. This separation was made clear by Jerome and Schumaker [42] in 1969.
2. The challenge of any abstract theory is to generalize a wide variety of particular cases, and simultaneously, preserve as much of the detail as possible. To a large extent, Atteia and others have, over many years, being doing this in the case that X is a reproducing kernel Hilbert space. Details of this theory can be found in the excellent monographs of Atteia ([11], 1992) and Bezhaev and Vasilenko ([17], 1993). The origins of this program can be found in 1959 paper by Golomb and Weinberger [35], in Habilitation Thesis of Atteia ([10], 1966) and in 1966 paper by de Boor and Lynch [20].
3. The above general theorem can itself be generalized in many directions.

One generalization enables us to consider constrained interpolation problems which are very important in contemporary mathematics. It is due to Atteia [10] and Utreras, [63] in 1987 and relates to the following diagram

$$\begin{array}{ccc} C \subset X & \xrightarrow{T} & Y \\ {\scriptstyle A}\downarrow & & \\ z \in Z & & \end{array}$$

THEOREM 11 (UTRERAS, 1987). *If*

- X, Y, Z *are Hilbert spaces,*
- C *is a closed, convex subset of* X*,*
- $z \in Z$
- A, T *are continuous, linear surjections,*
- $w \in I(C, z) := \{x \in C : \ Ax = z\}$
- $ker T + (ker A \cap (C - w))$ *is closed in* X *and*
- $ker A \cap ker T = \{\theta_X\}$

then $\exists!\ \sigma \in I(C, z)$ *such that*

$$\|T\sigma\|_Y = \min\{\|Tx\|_Y : \ x \in I(C, z)\}.$$

If we put $C = X$ then we obtain Theorem 10 of Atteia. Utreras' theorem is useful if, for example, we want to interpolate positive data by positive functions. In this case we have $X = H^m(I)$ and C is the set of positive function in X.

Other generalizations have extended Atteia's theorem to *Banach spaces settings*, rather than Hilbert spaces. So that are known the following new splines in Banach spaces:

R-splines (1972, Holmes [40])

M-splines (1972, Lucas [47], 1985 Abraham [1])

Lf-splines (1983, Pai [51])

Tf-splines (1993, Benbourhim and Gaches [16]).

A key work in the Banach space setting is the 1975 paper of Fischer and Jerome [32], where the perfect splines are very important in this contex.

5. Conclusions and comments

The book of Laurent ([44], 1972) was perhaps the first book which emphasized the variational approach to splines.

Atteia's book ([11], 1992) is the key work in this area, especially for those interested in functional analysis.

Whaba ([66], 1990) is the first book describing applications of these ideas (in smoothing rather the interpolation) to statistics.

Bezhaev and Vasilenko ([17], 1993) published in Novosibirsk entitled "Variational Spline Theory" contains the most abstracts and rigorous results in this field.

To close this presentation there are three conclusions to be underlined.

1. Splines may be defined as solution of variational problems rather than functions constructed in some piecewise manner. We have seen that these variational problems have become increasingly abstract and hence the concept of "splines" has became increasingly abstract. This may not be everyone's liking, at least, initially. For example, in 1966 in [20] de Boor and Lynch have written: "*in order not to dilute the notion of spline functions too much, we prefer to follow Greville's definition of a general spline function*" – which is based on a piecewise, constructive approach. In any case, the variational theory gives us a new appreciation of the concept of a "*spline*".

2. The variational approach facilitates a natural, attractive way to extend the classical theory of interpolating splines, especially to multivariate situations. The works of Duchon [29], [23] - [27] - [28], in 1976 and Whaba [65] in 1981 illustrate this conclusion. More recently, in 1993, de Boor [19] changing his earlier opinion wrote: "*I am convinced that the variational approach to splines will play a much greater role in multivariate spline theory that it did or should have in univariate theory*".

3. The theory of variational splines demonstrates the power of functional analysis to yield a unified approach to computational problems in interpolation. As S. Sobolev [45], one year before his dead has been quoted: "*It is impossible to image the theory of computations with no Banach spaces*".

References

[1] ABRAHAM A., *On the existence and uniqueness of M-splines*, J. Approx. Theory **43** (1985), 36–42.

[2] AHLBERG J.H., NILSON E.N. AND WALSH J.L., *Fundamental properties of generalized splines*, Proc. Nat. Acad. Sci. USA **52** (1964), 1412–1419.

[3] AHLBERG J.H., NILSON E.N. AND WALSH J.L., *The theory of splines and their applications*, Academic Press, New York 1967.

[4] AMODEI L., *Étude d'une classe de fonctions splines vectorielles en vue de l'approximation d'un champ de vitesse. Applications à la météorologie*, Thèse, Université Paul Sabatier, Toulouse 1993.

[5] AMODEI L. AND BENBOURHIM M.N., *A vector spline approximation*, J. Approx. Theory **67** (1991), 51–79.

[6] ANSELONE P.M. AND LAURENT P.J., *A general method for the construction of interpolating or smoothing spline-functions*, Numer. Math. **12** (1968), 66–82.

[7] APPRATO D., ARCANGELI R. AND GACHES J., *Fonctions spline par moyennes locales sur un ouvert borne de* IR^n, Ann. Facul. des Sci. Toulouse **V** (1983), 61–87.

[8] APPRATO D., ARCANGELI R. AND GACHES J., *Fonctions spline par moyennes locales sur un ouvert borne de IR^n*, Ann. Facul. des Sci. Toulouse **V** (1983), 61–87.

[9] ATTEIA M., *Généralisation de la définition et des propriétés des spline fonctions*, C. R. Acad. Sc. Paris **260** (1965), 3550–3553.

[10] ATTEIA M., *Étude de certains noyaux et théorie des fonctions spline en analyse numérique*, Thèse d'Etat, Instit. Math. Appl., Grenoble 1966.

[11] ATTEIA M., *Hilbertian kernels and spline functions*, North Holland, Amsterdam 1992.

[12] ATTEIA M., *Fonctions splines généralisées*, Sci. Paris **261** (1965), 2149–2152.

[13] ATTEIA M., *Existence et détermination des fonction - spline à plusieurs variables*, C.R. Acad. Sci. Paris, **262** (1966), 575–578.

[14] ATTEIA M., *Fonctions - splines et noyaux reproduisants d' Aronszajn - Bergman*, Re. Franc. d'Inf. et de Rech. Oper. **R3** (1970), 31–43.

[15] ATTEIA M., *Fonctions - splines definies dans un ensemble convexe*, Num. Math. **12** (1968), 192–210.

[16] BENBOURHIM N.M. AND GACHES J., *T_f-splines et approximation par T_f-prolongement*, Studia Math. **106** (1993), 203–211.

[17] BEZHAEV A.YU. AND VASILENKO V.A., *Variational spline heory*, Novosibirsk Computing Center, Novosibirsk 1993.

[18] DE BOOR C., *Best approximation properties of spline functions of odd degree*, J. Math. Mech. **12** (1963), 747–749.

[19] DE BOOR C., *Multivariate piecewise polynomials*, Acta Numerica (1993), 65–109.

[20] DE BOOR C. AND LYNCH R.E., *On spline and their minimum properties*, J. Math. Mech. **15** (1966), 953–969.

[21] BOUHAMIDI A., *Interpolation et approximation par des fonctions splines radiales à plusieurs variables*, Thèse, Univ. de Nantes, Nantes 1992.

[22] BOUHAMIDI A. AND LE MÉHAUTÉ A., *Splines curves et surfaces under tension*, in: "Wavelets, Image and Surface Fitting" (Eds. Laurent P.J., Le Méhauté A., Schumaker L.L. and Peters A.K.), Wellesley 1994, 51–58.

[23] BOUHAMIDI A. AND LE MÉHAUTÉ A., *Hilbertian approach for univariate spline with tension*, ATA **17** 4 (2001), 36–57.

[24] CHAMPIN R., LENARD C.T. AND MILLS T.M., *An introduction to abstract splines*, Math. Scientist **21** (1996), 8–26.

[25] CHAMPIN R., LENARD C.T. AND MILLS T.M., *A variational approach to splines*, The ANZIAM J. **42** (2000), 119–135.

[26] COPLEY P. AND SCHUMAKER L.L., *On p Lg-splines*, J. Approx. Theory **23** (1978), 1–28.

[27] DUCHON J., *Splines minimizing rotation-invariant semi-norms in Sobolev spaces*, LNM **571**, Springer Verlag 1977, 85–100.

[28] DUCHON J., *Sur l'erreur d'interpolation des fonctions de plusieurs variables par D^m - splines*, RAIRO Anal. Numer. **12** 4 (1978), 325–334.

[29] DUCHON J., *Interpolation des fonctions de deux variables suivant le principle de la flexion des plaques minces*, R.A.I.R.O. Analyse Numérique **10** (1976), 5–12.

[30] DE FIGUEIREDO R.J.P., *Lm-g-splines*, J. Approx. Theory **19** (1979), 332–360.

[31] DE FIGUEIREDO R.J.P. AND CHEN G., PDL_g *splines defined by partial differential operators with initial and boundary value condiitons*, SIAM J. Numer. Anal. **27** (1990), 519–528.

[32] FISHER S.D. AND JEROME J.W., *Elliptic variational problems in L^2 in L^∞*, Indiana Univ. Math. J. **23** (1974), 685–698.

[33] FREEDEN W., SCHREINER M. AND FRANKE R., *A survey of spherical spline approximation*, Surveys on Mathematics for Industry, 1992.

[34] GOLOMB M., *Splines, n-widths and optimal approximations*, MRC Technical Summary Report **784**, Mathematics Research Center, US Army, Wisconsin 1967.

[35] GOLOMB M. AND WEINBERGER H.F., *Optimal approximation and error bounds*, in: "On Numerical Approximation" (Ed. Langer R.E.), Univ. Wisconsin Press, Madison 1959, 117–190.

[36] GREVILLE T.N.E., *Interpolation by generalized spline functions*, MRC Technical Summary Report **476**, Mathematics Research Center, US Army, Madison, Wisconsin 1964.

[37] HARDER R.L. AND DESMARAIS R.N., *Interpolation using surface splines*, J. Aircraft **9** (1972), 189–191.

[38] HOLLADAY J.C., *A smoothest curve approximation*, Math. Tables Aids Comput. **11** (1957), 233–243.

[39] HOLLAND A.S.B. AND SAHNEY B.N., *The general problem of approximation and spline functions*, Robert E. Krieger Publ., Huntington 1979.

[40] HOLMES R., *R-splines in Banach spaces: I. Interpolation of linear manifolds*, J. Math. Anal. Appl. **40** (1972), 574–593.

[41] JEROME J.W. AND PIERCE J., *On spline functions determined by singular self-adjoint differential operators*, J. Approx. Theory **5** (1972), 15–40.

[42] JEROME J.W. AND SCHUMAKER L.L., *On Lg-splines*, J. Approx. Theory **2** (1969), 29–49.

[43] JEROME J.W. AND VARGA R.S., *Generalizations of spline functions and applications to nonlinear boundary value and eigenvalue problems*, in: "Theory and Applications of Spline Functions" (Ed. Greville T.N.E.), Academic Press, New York 1969, 103–155.

[44] LAURENT P.J., *Approximation et optimisation*, Hermann, Paris 1972.

[45] LEBEDEV V.I., *An introduction to functional analysis and computational mathematics*, Birkhäuser, Boston 1997.

[46] LUCAS T.R., *A theory of generalized splines with applications to nonlinear boundary value problems*, Ph. D. Thesis, Georgia Institute of Technology 1970.

[47] LUCAS T.R., *M-splines*, J. Approx. Theory **5** (1972), 1–14.

[48] MICULA G., *Funcţii spline şi aplicaţii*, Tehnică, Bucureşti 1978.

[49] MICULA G. AND BLAGA P., *Natural spline function of even degree*, Studia Univ. Babeş-Bolyai Cluj-Napoca, Series Mathematica **38** (1993), 31–40.

[50] MICULA G. AND MICULA S., *Handbook of Splines*, Kluwer Academic Publishers, Dordrecht-Boston-London 1999.

[51] PAI D.V., *On nonlinear minimization problems and Lf-splines. I*, J. Approx. Theory **39** (1983), 228–235.

[52] POWELL M.J.D., *The theory of radial basis function approximation*, Adv. Num. Anal. Vol.II (Ed. Light W.), OUP, Oxford 1992.

[53] PRENTER P.M., *Splines and variational methods*, John Wiley and Sons, New York 1975.

[54] RABUT C., *B-splines polyharmoniques cardinales: interpolation, quasi-interpolation, filtrage*, Thèse, Université Paul Sabatier, Toulouse 1990.

[55] SCHABACK R., *Konstruktion und algebraische Eigenschaften von M-Spline-Interpolierenden*, Numer. Math. **21** (1973), 166–180.

[56] SCHOENBERG I.J., *Contribution to the problem of approximation of equidistant data by analytic functions*, Parts A and B, Quart. Appl. Math. **4** (1946), 45–99, 112–141.

[57] SCHOENBERG I.J., *On trigonometric spline interpolation*, J. Math. Mech. **13** (1964), 795–825.

[58] SCHOENBERG I.J., *On the Ahlberg-Nilson extension of spline interpolation: The g-splines and their optimal properties*, J. Math. Anal. Appl. **21** (1968), 207–231.

[59] SCHULTZ M.H. AND VARGA R.S., *L-splines*, Numer. Math. **10** (1967), 345–369.

[60] SIDHU G.S. AND WEINERT H.L., *Vector-valued Lg-splines. I. Interpolating splines*, J. Math. Anal. Appl. **70** (1979), 505–529.

[61] TAIJERON H.J., GIBSON A.G. AND CHANDLER C., *Spline interpolation and smoothing on hyperspheres*, SIAM J. Sci. Comput. **15** (1994), 1111–1125.

[62] THOMANN J., *Détermination et construction de fonctions spline à deux variables définies sur un domaine rectangulaire ou circulaire*, Thèse, Université de Lille, Lille 1970.

[63] UTRERAS F.I., *Convergence rates for constrained spline functions*, Rev. Mat. Appl. **9** (1987), 87–95.

[64] VARGA R.S., *Functional analysis and approximation theory*, SIAM, Philadelphia 1971.

[65] WAHBA G., *Spline interpolation and smoothing on the sphere*, SIAM J. Sci. Stat. Comp. **2** (1981), 5–16.

[66] WAHBA G., *Spline models for observational data*, SIAM, Philadelphia 1990.

[67] WEINERT H.L., SESAI U.B. AND SIDHU G.S., *Arma splines, system inverses, and least-squares estimates*, SIAM J. Control Optim. **17** (1979), 525–536.

AMS Subject Classification: 65D07, 41A65, 65D02, 41A02.

Gheorghe MICULA
Babeş-Bolyai University
Department of Applied Mathematics
3400 Cluj-Napoca, ROMANIA
e-mail: `ghmicula@math.ubbcluj.rott`

Rend. Sem. Mat. Univ. Pol. Torino
Vol. 61, 3 (2003)
Splines and Radial Functions

P. Sablonnière*

QUADRATIC SPLINE QUASI-INTERPOLANTS ON BOUNDED DOMAINS OF $\mathbb{R}^d$, $d = 1, 2, 3$

Abstract. We study some C^1 quadratic spline quasi-interpolants on bounded domains $\Omega \subset \mathbb{R}^d, d = 1, 2, 3$. These operators are of the form $Qf(x) = \sum_{k\in K(\Omega)} \mu_k(f)B_k(x)$, where $K(\Omega)$ is the set of indices of B-splines B_k whose support is included in the domain Ω and $\mu_k(f)$ is a discrete linear functional based on values of f in a neighbourhood of $x_k \in supp(B_k)$. The data points x_j are vertices of a uniform or nonuniform partition of the domain Ω where the function f is to be approximated. Beyond the simplicity of their evaluation, these operators are uniformly bounded independently of the given partition and they provide the best approximation order to smooth functions. We also give some applications to various fields in numerical approximation.

1. Introduction and notations

In this paper, we continue the study of some C^1 quadratic (or d-quadratic) spline *discrete quasi-interpolant* (dQIs) on bounded domains $\Omega \subset \mathbb{R}^d, d = 1, 2, 3$ initiated in [36]. These operators are of the form $Qf(x) = \sum_{k\in K(\Omega)} \mu_k(f)B_k(x)$, where $K(\Omega)$ is the set of indices of B-splines B_k whose support is included in the domain Ω and $\mu_k(f)$ is a *discrete linear functional* $\sum_{i\in I(r)} \lambda_k(i) f(x_{i+k})$, with $I(r) = \mathbb{Z}^d \cap [-r, r]^d$ for $r \in \mathbb{N}$ fixed (and small). The data points x_j are vertices of a uniform or nonuniform partition of the domain Ω where the function f is to be approximated. Such operators have been widely studied in recent years (see e.g. [4], [6]-[11],[14], [23], [24], [31], [38], [40]), but in general, except in the univariate or multivariate tensor-product cases, they are defined on the *whole space* $\mathbb{R}^d$: here we restrict our study to *bounded domains* and to C^1 *quadratic spline* dQIs. Their main interest lies in the fact that they provide approximants having the best approximation order and small norms while being easy to compute. They are particularly useful as initial approximants at the first step of a multiresolution analysis. First, we study univariate dQIs on uniform and non-uniform meshes of a bounded interval of the real line (Section 2) or on bounded rectangles of the plane with a uniform or non-uniform criss-cross triangulation (Section 3). We use

*The author thanks very much prof. Catterina Dagnino, from the Dipartimento di Matematica dell'Università di Torino, and the members of the italian project GNCS on spline and radial functions, for their kind invitation to the *Giornate di Studio su funzioni spline e funzioni radiali*, held in Torino in February 6-7, 2003, where this paper was presented by the author.

quadratic B-splines whose Bernstein-Bézier (abbr. BB)-coefficients are given in technical reports [37], [38] and which extend previous results given in [12]. In the same way, in section 4, we complete the study of a bivariate blending sum of two univariate dQIs of Section 1 on a rectangular domain. Finally, in Section 5, we do the same for a trivariate blending sum of a univariate dQI (Section 1) and of the bivariate dQI described in Section 2. For blending and tensor product operators, see e.g. [2], [3], [16], [18], [19], [20], [21], [30]. For some of these operators, we improve the estimations of infinite norms which are *bounded independently of the given partition* of the domain. Using the fact that the dQI S is exact on the space $\mathbb{P}_2 \in \mathcal{S}_2$ of quadratic polynomials and a classical result of approximation theory: $\|f - Sf\| \le (1+\|S\|)d(f, \mathcal{S}_2)$ (see e.g. [15], chapter 5), we conclude that $f - Sf = O(h^3)$ for f smooth enough, where h is the maximum of diameters of the elements (segments, triangles, rectangles, prisms) of the partition of the domain. But we specify upper bounds for some constants occuring in inequalities giving error estimates for functions and their partial derivatives of total order at most 2. Finally, in Section 6, we present some applications of the preceding dQIs, for example to the computation of multivariate integrals, to the approximate determination of zeros of functions, to spectral-type methods and to the solution of integral equations. They are still in progress and will be published elsewhere.

2. Quadratic spline dQIs on a bounded interval

Let $X = \{x_0, x_1, \dots, x_n\}$ be a partition of a bounded interval $I = [a, b]$, with $x_0 = a$ and $x_n = b$. For $1 \le i \le n$, let $h_i = x_i - x_{i-1}$ be the length of the subinterval $I_i = [x_{i-1}, x_i]$. Let $\mathcal{S}_2(X)$ be the $n+2$-dimensional space of C^1 quadratic splines on this partition. A basis of this space is formed by quadratic B-splines $\{B_i, 0 \le i \le n+1\}$. Define the set of evaluation points

$$\Theta_n = \{\theta_0 = x_0,\ \ \theta_i = \frac{1}{2}(x_{i-1} + x_i),\ \ for\ \ 1 \le i \le n,\ \ \theta_{n+1} = x_n\}.$$

The simplest dQI associated with Θ_n is the *Schoenberg-Marsden* operator (see e.g. [25], [36]):

$$S_1 f := \sum_{i=0}^{n+1} f(\theta_i) B_i$$

This operator is exact on $\mathbb{P}_1$. Moreover $S_1 e_2 = e_2 + \sum_{i=1}^{n} \frac{1}{4} h_i^2 B_i$. We have studied in [1] and [36] the unique dQI of type

$$S_2 f = f(x_0) B_0 + \sum_{i=1}^{n} \mu_i(f) B_i + f(x_n) B_{n+1}$$

whose coefficient functionals are of the form

$$\mu_i(f) = a_i f(\theta_{i-1}) + b_i f(\theta_i) + c_i f(\theta_{i+1}),\ \ 1 \le i \le n$$

and which is exact on the space $\mathbb{P}_2$ of quadratic polynomials. Using the following notations and the convention $h_0 = h_{n+1} = 0$, we finally obtain, for $1 \le i \le n$:

$$\sigma_i = \frac{h_i}{h_{i-1} + h_i}, \quad \sigma'_i = \frac{h_{i-1}}{h_{i-1} + h_i} = 1 - \sigma_i,$$

$$a_i = -\frac{\sigma_i^2 \sigma'_{i+1}}{\sigma_i + \sigma'_{i+1}}, \quad b_i = 1 + \sigma_i \sigma'_{i+1}, \quad c_i = -\frac{\sigma_i (\sigma'_{i+1})^2}{\sigma_i + \sigma'_{i+1}}.$$

Defining the fundamental functions of S_2 by

$$\begin{aligned}
\tilde{B}_0 &= B_0 + a_1 B_1, \\
\tilde{B}_i &= c_{i-1} B_{i-1} + b_i B_i + a_{i+1} B_{i+1}, \; 1 \le i \le n, \\
\tilde{B}_{n+1} &= c_n B_n + B_{n+1},
\end{aligned}$$

we can express $S_2 f$ in the following form

$$S_2 f = \sum_{i=0}^{n+1} f(\theta_i) \tilde{B}_i.$$

In [26] (see also [22] and [32], chapter 3), Marsden proved the existence of a unique *Lagrange interpolant* Lf in $\mathcal{S}_2(X)$ satisfying $Lf(\theta_i) = f(\theta_i)$ for $0 \le i \le n+1$. He also proved the following

THEOREM 1. *For f bounded on I and for any partition X of I, the Chebyshev norm of the Lagrange operator L is uniformly bounded by* 2.

Now, we will prove a similar result for the dQI S_2 defined above. It is well known that the infinite norm of S_2 is equal to the Chebyshev norm of the Lebesgue function $\Lambda_2 = \sum_{i=0}^{n+1} |\tilde{B}_i|$ of S_2.

THEOREM 2. *For f bounded on I and for any partition X of I, the infinite norm of the dQI S_2 is uniformly bounded by* 2.5.

Proof. Each function $|\tilde{B}_i|$ being bounded above by the continuous quadratic spline $\bar{B}_i$ whose BB-coefficients are absolute values of those of $\tilde{B}_i$, we obtain $\Lambda_2 \le \bar{\Lambda}_2 = \sum_{i=0}^{n+1} \bar{B}_i$. So, we have to find an upper bound of $\bar{\Lambda}_2$. First, we need the BB-coefficients of the fundamental functions: they are computed as linear combinations of the BB-coefficients of B-splines. In order to avoid complicated notations, we denote by $[a, b, c]$ the triplet of BB-coefficients of the quadratic polynomial $a(1-u)^2 + 2bu(1-u) + cu^2$ for $u \in [0, 1]$. Any function $g \in \mathcal{S}_2(X)$ can be written in this form on each interval $[x_{i-1}, x_i]$, $1 \le i \le n$, with the change of variable $u = (x - x_{i-1})/h_i$. So, the BB-coefficients of g consist of a list of n triplets. Let us denote by $L(i)$ the list associated with the function $\tilde{B}_i$ (we do not write the triplets of null BB-coefficients). Setting, for $1 \le i \le n-1$:

$$d_i = c_i \sigma_{i+1} + b_{i+1} \sigma'_{i+1}, \quad e_i = b_i \sigma_{i+1} + a_{i+1} \sigma'_{i+1},$$

we obtain for the three first functions $\tilde{B}_0$, $\tilde{B}_1$, $\tilde{B}_2$:

$$L(0) = [1, a_1, a_1\sigma_2], [a_1\sigma_2, 0, 0]$$

$$L(1) = [0, b_1, e_1], [e_1, a_2, a_2\sigma_3], [a_2\sigma_3, 0, 0]$$

$$L(2) = [0, c_1, d_1], [d_1, b_2, e_2], [e_2, a_3, a_3\sigma_4], [a_3\sigma_4, 0, 0]$$

For $3 \le i \le n-2$ (general case), we have $supp(\tilde{B}_i) = [x_{i-3}, x_{i+2}]$ and

$$L(i) = [0, 0, c_{i-1}\sigma'_{i-1}], [c_{i-1}\sigma'_{i-1}, c_{i-1}, d_{i-1}], [d_{i-1}, b_i, e_i],$$

$$[e_i, a_{i+1}, a_{i+1}\sigma_{i+2}], [a_{i+1}\sigma_{i+2}, 0, 0]$$

Finally, for the three last functions $\tilde{B}_{n-1}$, $\tilde{B}_n$, $\tilde{B}_{n+1}$, we get:

$$L(n-1) = [0, 0, c_{n-2}\sigma'_{n-2}], [c_{n-2}\sigma'_{n-2}, c_{n-2}, d_{n-2}], [d_{n-2}, b_{n-1}, e_{n-1}], [e_{n-1}, a_n, 0]$$

$$L(n) = [0, 0, c_{n-1}\sigma'_{n-1}], [c_{n-1}\sigma'_{n-1}, c_{n-1}, d_{n-1}], [d_{n-1}, b_n, 0]$$

$$L(n+1) = [0, 0, c_n\sigma'_n], [c_n\sigma'_n, c_n, 1]$$

We see that $d_i \ge 0$ (resp. $e_i \ge 0$), for it is a convex combination of c_i and b_{i+1} (resp. of b_i and a_{i+1}), with $b_i \ge 1$ and $|c_i|$ and $|a_i| \le 1$ for all i. Therefore, the absolute values of the above BB-coefficients (i.e. the BB-coefficients of the $\bar{B}'_i s$) are easy to evaluate. Now, it is easy to compute the BB-coefficients of the continuous quadratic spline $\bar{\Lambda}_2 = \sum_{i=0}^{n+1} \bar{B}_i$. On each interval $[x_{i-1}, x_i]$, for $2 \le i \le n-1$, we obtain

$$[\lambda_{i-1}, \mu_i, \lambda_i] = [-a_{i-1}\sigma_i + d_{i-1} + e_{i-1} - c_i\sigma'_i, b_i - a_i - c_i, -a_i\sigma_{i+1} + d_i + e_i - c_{i+1}\sigma'_{i+1}]$$

For the first (resp. the last) interval, we have $\lambda_0 = 1$ (resp. $\lambda_n = 1$) For the central BB-coefficient, we get, since σ_i and σ'_i are in $[0, 1]$ for all indices:

$$\mu_i = b_i - (a_i + c_i) = 2b_i - 1 = 1 + 2\sigma_i\sigma'_{i+1} \le 3$$

For the extreme BB-coefficients, we have, since $a_i + b_i + c_i = 1$:

$$\lambda_i = (1 - 2a_i)\sigma_{i+1} + (1 - 2c_{i+1})\sigma'_{i+1} = 1 + \frac{2(\sigma_i)^2\sigma_{i+1}\sigma'_{i+1}}{\sigma_i + \sigma'_{i+1}} + \frac{2\sigma_{i+1}\sigma'_{i+1}(\sigma'_{i+2})^2}{\sigma_{i+1} + \sigma'_{i+2}}.$$

Let us consider the rational function f defined by $\lambda_i = 1 + f(\sigma_i, \sigma_{i+1}, \sigma_{i+2})$:

$$f(x, y, z) = \frac{2x^2y(1-y)}{1+x-y} + \frac{2y(1-y)(1-z)^2}{1+y-z},$$

the three variables x, y, z lying in the unit cube. Its maximum is attained at the vertices $\{(0, 1, 0), (1, 0, 0), (1, 0, 1), (1, 1, 0)\}$ and it is equal to 1. This proves that $\lambda_i \le 2$ for all i. Therefore, in each subinterval (after the canonical change of variable), $\bar{\Lambda}_2$ is bounded above by the parabola:

$$\pi_2(u) = 2(1-u)^2 + 6u(1-u) + 2u^2$$

whose maximum value is $\pi_2(\frac{1}{2}) = \frac{5}{2} = 2.5$.

□

Now, we consider the case of a *uniform partition*, say with integer nodes for simplification (e.g. $I = [0, n]$, $X = \{0, 1, \ldots, n\}$). In that case, we have

$$\sigma_1 = 1, \;\; \sigma_1' = 0; \;\; \sigma_i = \sigma_i' = \frac{1}{2} \;\; for \;\; 2 \le i \le n; \;\; \sigma_{n+1} = 0, \;\; \sigma_{n+1}' = 1,$$

from which we deduce:

$$a_1 = c_n = -\frac{1}{3}, \;\; b_1 = b_n = \frac{3}{2}, \;\; c_1 = a_n = -\frac{1}{6},$$

and, for $2 \le i \le n-1$:

$$a_i = c_i = -\frac{1}{8}, \;\; b_i = \frac{5}{4}.$$

It is easy to see that, in order to compute $\|S_2\|_\infty$, it suffices to evaluate the maximum of the Lebesgue function on the subinterval $J = [0, 4]$. Here are the lists $L(i)$ of the BB-coefficients of the fundamental functions $\{\tilde{B}_i, 0 \le i \le 6\}$ whose supports have at least a common subinterval with J. As in the nonuniform case, we only give the triplets associated with subintervals of $supp(\tilde{B}_i) \cap J$:

$$\begin{aligned}
supp(\tilde{B}_0) \cap J = [0,2], \quad L(0) &= \left[1, -\frac{1}{3}, -\frac{1}{6}\right], \left[-\frac{1}{6}, 0, 0\right] \\
supp(\tilde{B}_1) \cap J = [0,3], \quad L(1) &= \left[0, \frac{3}{2}, \frac{11}{16}\right], \left[\frac{11}{16}, -\frac{1}{8}, -\frac{1}{16}\right], \left[-\frac{1}{16}, 0, 0\right], \\
supp(\tilde{B}_2) \cap J = [0,4], \quad L(2) &= \left[0, -\frac{1}{6}, \frac{13}{24}\right], \left[\frac{13}{24}, \frac{5}{4}, \frac{9}{16}\right], \left[\frac{9}{16}, -\frac{1}{8}, -\frac{1}{16}\right], \\
&\quad \left[-\frac{1}{16}, 0, 0\right], \\
supp(\tilde{B}_3) \cap J = [0,2], \quad L(3) &= \left[0, 0, -\frac{1}{16}\right], \left[-\frac{1}{16}, -\frac{1}{8}, \frac{9}{16}\right], \left[\frac{9}{16}, \frac{5}{4}, \frac{9}{16}\right], \\
&\quad \left[\frac{9}{16}, -\frac{1}{8}, -\frac{1}{16}\right], \\
supp(\tilde{B}_4) \cap J = [1,4], \quad L(4) &= \left[0, 0, -\frac{1}{16}\right], \left[-\frac{1}{16}, -\frac{1}{8}, \frac{9}{16}\right], \left[\frac{9}{16}, \frac{5}{4}, \frac{9}{16}\right], \\
supp(\tilde{B}_5) \cap J = [2,4], \quad L(5) &= \left[0, 0, -\frac{1}{16}\right], \left[-\frac{1}{16}, -\frac{1}{8}, \frac{9}{16}\right], \\
supp(\tilde{B}_6) \cap J = [3,4], \quad L(6) &= \left[0, 0, -\frac{1}{16}\right],
\end{aligned}$$

Drawing Λ_2 reveals that the abscissa $\bar{x}$ of its maximum lies in the interval [0.6, 1]. In this interval, we obtain successively:

$$\Lambda_2(x) = -\tilde{B}_0(x) + \tilde{B}_1(x) + \tilde{B}_2(x) - \tilde{B}_3(x) = -(1-x)^2 + \frac{10}{3}x(1-x) + \frac{35}{24}x^2$$

whence $\Lambda_2'(x) = \frac{1}{12}(64 - 69x)$ and $\bar{x} = \frac{64}{69}$. This leads to

$$\|S_2\|_\infty = \|\Lambda_2\|_\infty = \Lambda_2(\bar{x}) = \frac{305}{207} \approx 1.4734.$$

$\square$

So, we have proved the following result:

THEOREM 3. *For uniform partitions of the interval I, the infinite norm of S_2 is equal to* $\frac{305}{207} \approx 1.4734$.

REMARK 1. Further results on various types of dQIs will be given in [21].

Now, we will give some bounds for the error $f - S_2 f$. Using the fact that the dQI S_2 is exact on the subspace $\mathbb{P}_2 \subset \mathcal{S}_2$ of quadratic polynomials and a classical result of approximation theory (see e.g. [17], chapter 5), we have for all partitions X of I in virtue of Theorem 4:

$$\|f - S_2 f\|_\infty \leq (1 + \|S_2\|_\infty)\, dist(f, \mathcal{S}_2)_\infty \leq 3.5\, dist(f, \mathcal{S}_2)_\infty$$

So, the approximation order is that of the best quadratic spline approximation. For example, from [17], we know that for any continuous function f

$$dist(f, \mathcal{S}_2)_\infty \leq 3\, \omega(f, h)_\infty$$

where $h = max\{h_i, 1 \leq i \leq n\}$, so we obtain

$$\|f - S_2 f\|_\infty \leq 10.5\, \omega(f, h)_\infty$$

But a direct study allows to decrease the constant in the right-hand side.

THEOREM 4. *For a continuous function f, there holds:*

$$\|f - S_2 f\|_\infty \leq 6\, \omega(f, h)_\infty$$

Proof. For any $x \in I$, we have

$$f(x) - S_2 f(x) = \sum_{i=0}^{n+1} [f(x) - f(\theta_i)] \tilde{B}_i(x)$$

Assuming $n \geq 5$ and $x \in I_p = [x_{p-1}, x_p]$, for some $3 \leq p \leq n-2$, this error can be written, since $supp(\tilde{B}_i) = [x_{i-3}, x_{i+2}]$:

$$f(x) - S_2 f(x) = \sum_{i=p-2}^{p+2} [f(x) - f(\theta_i)] \tilde{B}_i(x).$$

As $\theta_i = \frac{1}{2}(x_{i-1} + x_i)$, we have $|x - \theta_i| \leq r_i h$, with $r_i = |p - i| + 0.5$. Using a well known property of the modulus of continuity of f, $\omega(f, r_i h) \leq (1 + r_i)\omega(f, h)$, we deduce

$$|f(x) - S_2 f(x)| \leq \left[\sum_{i=p-2}^{p+2} (1 + r_i) \bar{B}_i(x) \right] \omega(f, h).$$

Without going into details, we use the local BB-coefficients of $\bar{B}_i$, $p - 2 \leq i \leq p + 2$ in the subinterval $[x_{p-1}, x_p]$, and we can prove that for all partitions of I, we have

$$\sum_{i=p-2}^{p+2} (1.5 + |p - i|) \bar{B}_i(x) \leq 6$$

so, we obtain finally a lower constant (but not the best one) in the right-hand side of the previous inequality:

$$\| f - S_2 f \|_\infty \leq 6 \ \omega(f, h)_\infty$$

□

Now, let us assume that $f \in C^3(I)$, then we have the following

THEOREM 5. *For all function $f \in C^3(I)$ and for all partitions X of I, the following error estimate holds, with $C_0 \leq 1$:*

$$\| f - S_2 f \|_\infty \leq C_0 h^3 \| f^{(3)} \|_\infty$$

Proof. Given $x \in I_p$ fixed and $t \in [x_{p-3}, x_{p+2}]$, we use the Taylor formula with integral remainder

$$f(t) = f(x) + (t - x) f'(x) + \frac{1}{2}(t - x)^2 f''(x) + \frac{1}{2} \int_x^t (t - s)^2 f^{(3)}(s) ds$$

As $p_1(t) = t - x$ and $p_2(t) = (t - x)^2$ are in $\mathbb{P}_2$, we have $S_2 p_1 = p_1$ and $S_2 p_2 = p_2$, which can be written explicitly as

$$S_2 p_1(t) = t - x = \sum_{i=0}^{n+1} (\theta_i - x) \tilde{B}_i(t), \quad S_2 p_2(t) = (t - x)^2 = \sum_{i=0}^{n+1} (\theta_i - x)^2 \tilde{B}_i(t)$$

and this proves that $S_2 p_1(x) = S_2 p_2(x) = 0$. Therefore it remains:

$$S_2 f(x) - f(x) = \frac{1}{2} \sum_{i=p-2}^{p+2} \left[\int_x^{\theta_i} (\theta_i - s)^2 f^{(3)}(s) ds \right] \tilde{B}_i(x)$$

As $|\int_x^{\theta_i}(\theta_i - s)^2 ds| \le \frac{1}{3}|x-\theta_i|^3$, we get the following upper bound:

$$\begin{aligned}|S_2 f(x) - f(x)| &\le \frac{1}{6}\|f^{(3)}\|_\infty \sum_{i=p-2}^{p+2} |x-\theta_i|^3 \bar{B}_i(x)\\ &\le \frac{h^3}{6}\|f^{(3)}\|_\infty \sum_{i=p-2}^{p+2} (|p-i| + \frac{1}{2})^3 \bar{B}_i(x)\end{aligned}$$

As in the proof of theorem above, and without going into details, one can prove that the last sum in the r.h.s. is uniformly bounded by 6 for any partition of I. So, we obtain finally:

$$|S_2 f(x) - f(x)| \le h^3 \|f^{(3)}\|_\infty$$

□

By using the same techniques, the results of theorem 5 can be improved when X is a *uniform partition* of I:

THEOREM 6. *(i) For $f \in C(I)$, there holds:*

$$|S_2 f(x) - f(x)| \le 2.75\ \omega(f, \frac{h}{2})_\infty$$

(ii) for $f \in C^3(I)$ and for all $x \in I$ there holds:

$$|S_2 f(x) - f(x)| \le \frac{h^3}{3}\|f^{(3)}\|_\infty$$

$$|(S_2 f)'(x) - f'(x)| \le 1.2\ h^2 \|f^{(3)}\|_\infty$$

and locally, in each subinterval of I:

$$|(S_2 f)''(x) - f''(x)| \le 2.4\ h \|f^{(3)}\|_\infty$$

3. Quadratic spline dQIs on a bounded rectangle

In this section, we study some C^1 quadratic spline dQIs on a nonuniform criss-cross triangulation of a rectangular domain. More specifically, let $\Omega = [a_1, b_1] \times [a_2, b_2]$ be a rectangle decomposed into mn subrectangles by the two partitions

$$X_m = \{x_i,\ \ 0 \le i \le m\},\ \ Y_n = \{y_j,\ \ 0 \le j \le n\}$$

respectively of the segments $I = [a_1, b_1] = [x_0, x_m]$ and $J = [a_2, b_2] = [y_0, y_n]$. For $1 \le i \le m$ and $1 \le j \le n$, we set $h_i = x_i - x_{i-1}$, $k_j = y_j - y_{j-1}$, $I_i = [x_{i-1}, x_i]$, $J_j = [y_{j-1}, y_j]$, $s_i = \frac{1}{2}(x_{i-1} + x_i)$ and $t_j = \frac{1}{2}(y_{j-1} + y_j)$. Moreover

$s_0 = x_0, s_{m+1} = x_m, t_0 = y_0, t_{n+1} = y_n$. In this section and the next one, we use the following notations:

$$\sigma_i = \frac{h_i}{h_{i-1}+h_i}, \quad \sigma_i' = \frac{h_{i-1}}{h_{i-1}+h_i} = 1-\sigma_i,$$

$$\tau_j = \frac{k_j}{k_{j-1}+k_j}, \quad \tau_j' = \frac{k_{j-1}}{k_{j-1}+k_j} = 1-\tau_j,$$

for $1 \le i \le m$ and $1 \le j \le n$, with the convention $h_0 = h_{m+1} = k_0 = k_{n+1} = 0$.

$$a_i = -\frac{\sigma_i^2\sigma_{i+1}'}{\sigma_i+\sigma_{i+1}'}, \quad b_i = 1+\sigma_i\sigma_{i+1}', \quad c_i = -\frac{\sigma_i(\sigma_{i+1}')^2}{\sigma_i+\sigma_{i+1}'},$$

$$\bar{a}_j = \frac{\tau_j^2\tau_{j+1}'}{\tau_j+\tau_{j+1}'}, \quad \bar{b}_j = 1+\tau_j\tau_{j+1}', \quad \bar{c}_j = -\frac{\tau_j(\tau_{j+1}')^2}{\tau_j+\tau_{j+1}'}.$$

for $0 \le i \le m+1$ and $0 \le j \le n+1$. Let $\mathcal{K}_{mn} = \{(i,j) : 0 \le i \le m+1, \quad 0 \le j \le n+1\}$, then the data sites are the mn intersection points of diagonals in subrectangles $\Omega_{ij} = I_i \times J_j$, the $2(m+n)$ midpoints of the subintervals on the four edges, and the four vertices of Ω, i.e. the $(m+2)(n+2)$ points of the following set

$$\mathcal{D}_{mn} := \{M_{ij} = (s_i, t_j), \quad (i,j) \in \mathcal{K}_{mn}\}.$$

As in Section 2, the simplest dQI is the bivariate Schoenberg-Marsden operator:

$$S_1 f = \sum_{(i,j)\in\mathcal{K}_{mn}} f(M_{ij}) B_{ij}$$

where

$$\mathcal{B}_{mn} := \{B_{ij}, 0 \le i \le m+1, \quad 0 \le j \le n+1\}$$

is the collection of $(m+2)(n+2)$ B-splines (or generalized box-splines) generating the space $\mathcal{S}_2(\mathcal{T}_{mn})$ of all C^1 piecewise quadratic functions on the criss-cross triangulation $\mathcal{T}_{mn}$ associated with the partition $X_m \times Y_n$ of the domain Ω (see e.g. [14], [13]). There are mn *inner B-splines* associated with the set of indices

$$\hat{\mathcal{K}}_{mn} = \{(i,j), 1 \le i \le m, 1 \le j \le n\}$$

whose restrictions to the boundary Γ of Ω are equal to zero. To the latter, we add $2m+2n+4$ *boundary B-splines* whose restrictions to Γ are univariate quadratic B-splines. Their set of indices is

$$\tilde{\mathcal{K}}_{mn} := \{(i,0), (i,n+1), 0 \le i \le m+1; \ (0,j), (m+1,j), \quad 0 \le j \le n+1\}$$

The BB-coefficients of inner B-splines whose indices are in $\{(i,j), 2 \le i \le m-1, 2 \le j \le n-1\}$ are given in [32]. The other ones can be found in the technical reports [37] (uniform partition) and [38](non-uniform partitions). The B-splines are positive and form a partition of unity (blending system). The boundary B-splines are *linearly independent* as the univariate ones. But the inner B-splines are *linearly dependent*, the dependence relationship being:

$$\sum_{(i,j)\in\hat{\mathcal{K}}_{mn}} (-1)^{i+j} h_i k_j B_{ij} = 0$$

It is well known that S_1 is exact on bilinear polynomials, i.e.

$$S_1 e_{rs} = e_{rs} \quad for \quad 0 \le r, s \le 1$$

In [36], we obtained the following dQI, which is exact on $\mathbb{P}_2$:

$$S_2 f = \sum_{(i,j)\in\mathcal{K}_{mn}} \mu_{ij}(f) B_{ij}$$

where the coefficient functionals are given by

$$\begin{aligned} \mu_{ij}(f) &= (b_i + \bar{b}_j - 1) f(M_{ij}) + a_i f(M_{i-1,j}) + c_i f(M_{i+1,j}) \\ &+ \bar{a}_j f(M_{i,j-1}) + \bar{c}_j f(M_{i,j+1}). \end{aligned}$$

As in Section 2, we introduce the fundamental functions:

$$\tilde{B}_{ij} = (b_i + \bar{b}_j - 1) B_{ij} + a_{i+1} B_{i+1,j} + c_{i-1} B_{i-1,j} + \bar{a}_{j+1} B_{i,j+1} + \bar{c}_{j-1} B_{i,j-1}.$$

We also proved the following theorems, by bounding above the Lebesgue function of S_2:

$$\Lambda_2 = \sum_{(i,j)\in\mathcal{K}_{mn}} |\tilde{B}_{ij}|$$

THEOREM 7. *The infinite norm of S_2 is uniformly bounded independently of the partition $\mathcal{T}_{mn}$ of the domain:*

$$\|S_2\|_\infty \le 5$$

THEOREM 8. *For uniform partitions, we have the following bound:*

$$\|S_2\|_\infty \le 2.4$$

These bounds are probably not optimal and can still be slightly reduced.

4. A biquadratic blending sum of univariate dQIs

In this section, we study a biquadratic dQI on a rectangular domain $\Omega = [a_1, b_1] \times [a_2, b_2]$ which is a blending sum of bivariate extensions of quadratic spline dQIs of Section 2. We use the same notations as in Section 2 for the domain Ω, the partitions of $I = [a_1, b_1]$, $J = [a_2, b_2]$ and data sites. The partition considered on Ω is the tensor product of partitions of I and J. We use the two sets of univariate B-splines

$$\{B_i(x), 0 \le i \le m+1\}, \quad \{B_j(y), 0 \le j \le n+1\}$$

and the two sets of univariate fundamental functions introduced in Section 2:

$$\{\tilde{B}_i(x), 0 \le i \le m+1\}, \quad \{\tilde{B}_j(y), 0 \le j \le n+1\}$$

The associated *extended bivariate* dQIs are respectively (see e.g. [14] for bivariate extensions of univariate operators)

$$P_1 f(x,y) := \sum_{i=0}^{m+1} f(s_i, y) B_i(x), \quad P_2 f(x,y) := \sum_{i=0}^{m+1} f(s_i, y)\tilde{B}_i(x)$$

$$Q_1 f(x,y) := \sum_{j=0}^{n+1} f(x, t_j) B_j(y), \quad Q_2 f(x,y) := \sum_{j=0}^{n+1} f(x, t_j)\tilde{B}_j(y)$$

The bivariate dQI considered in this section is now defined as the blending sum

$$R := P_1 Q_2 + P_2 Q_1 - P_1 Q_1$$

and it can be written in the following form

$$Rf(x,y) = \sum_{(i,j)\in\mathcal{K}_{mn}} f(M_{ij})\bar{B}_{ij}(x,y)$$

where the biquadratic fundamental functions are defined by

$$B_{ij}^{\flat}(x,y) := B_i(x)\tilde{B}_j(y) + \tilde{B}_i(x)B_j(y) - B_i(x)B_j(y)$$

In terms of tensor-product B-splines $B_{ij}(x,y) = B_i(x)B_j(y)$, we have:

$$Rf(x,y) = \sum_{(i,j)\in\mathcal{K}_{mn}} \mu_{ij}(f) B_{ij}(x,y),$$

where the coefficient functionals are given by

$$\begin{array}{rcl} \mu_{ij}(f) & := & a_i f(M_{i-1,j}) + c_i f(M_{i+1,j}) + \bar{a}_j f(M_{i,j-1}) \\ & + & \bar{c}_j f(M_{i,j+1}) + (b_i + \bar{b}_j - 1) f(M_{ij}) \end{array}$$

We have proved in [36] the following

THEOREM 9. *The operator R is exact on the 8-dimensional subspace* $(\mathbb{P}_{12}[x,y]) \oplus (\mathbb{P}_{21}[x,y])$ *of biquadratic polynomials. Moreover, its infinite norm is bounded above independently of the nonuniform partition $X_m \otimes Y_n$ of the domain Ω*

$$\|R\|_\infty \le 5$$

5. A trivariate blending sum of univariate and bivariate quadratic dQIs

In this section, we study a trivariate dQI on a parallelepiped $\Omega = [a_1, b_1] \times [a_2, b_2] \times [a_3, b_3]$ which is a blending sum of trivariate extensions of univariate and bivariate dQIs seen in Sections 2 and 3. We consider the three partitions

$$X_m := \{x_i, \ \ 0 \le i \le m\}, \ \ Y_n = \{y_j, \ \ 0 \le j \le n\}, \ \ Z_p := \{z_k, \ \ 0 \le k \le p\}$$

respectively of the segments $I = [a_1, b_1] = [x_0, x_m]$, $J = [a_2, b_2] = [y_0, y_n]$ and $K = [a_3, b_3] = [z_0, z_p]$. For the projection $\Omega' = [a_1, b_1] \times [a_2, b_2]$ of Ω on the $xy - plane$, the notations are those of Section 3. For the projection $\Omega'' = [a_3, b_3]$ of Ω on the $z - axis$, we use the following notations, for $1 \le k \le p$:

$$l_k = z_k - z_{k-1}, \ \ K_k = [z_{k-1}, z_k], \ \ u_k = \frac{1}{2}(z_{k-1} + z_k),$$

with $u_0 = z_0$ and $u_{p+1} = z_p$. For mesh ratios of subintervals, we set respectively

$$\omega_k = \frac{l_k}{l_{k-1} + l_k}, \ \ \omega'_k = \frac{l_{k-1}}{l_{k-1} + l_k} = 1 - \omega_k$$

for $1 \le k \le p$, with $l_0 = l_{p+1} = 0$ (all these ratios lie between 0 and 1), and

$$\hat{a}_k = -\frac{\omega_k^2 \omega'_{k+1}}{\omega_k + \omega'_{k+1}}, \ \ \hat{b}_k = 1 + \omega_k \omega'_{k+1}, \ \ \hat{c}_k = -\frac{\omega_k (\omega'_{k+1})^2}{\omega_k + \omega'_{k+1}}.$$

Let $\mathcal{K} = \mathcal{K}_{mnp} = \{(i, j, k), \ \ 0 \le i \le m+1, \ 0 \le j \le n+1, \ 0 \le k \le p+1\}$, then the set of data sites is

$$\mathcal{D} = \mathcal{D}_{mnp} = \{N_{ijk} = (x_i, y_j, z_k), (i, j, k) \in \mathcal{K}_{mnp}\},$$

The partition of Ω considered here is the tensor product of partitions on Ω' and Ω'', i.e. a partition into *vertical prisms with triangular horizontal sections*. Setting $\mathcal{K}'_{mn} = \{(i, j), 0 \le i \le m+1, \ 0 \le j \le n+1\}$, we consider the bivariate B-splines and fundamental splines on $\Omega' = [a_1, b_1] \times [a_2, b_2]$ defined in Section 3 above:

$$\{B_{ij}(x, y), \ \ (i, j) \in \mathcal{K}'_{mn}\}, and \ \ \{\tilde{B}_{ij}(x, y), \ \ (i, j) \in \mathcal{K}'_{mn}\}$$

and the univariate B-splines and fundamental splines on $[a_3, b_3]$ defined in Section 2:

$$\{B_k(z), \ \ 0 \le k \le p+1\} \ \ and \ \ \{\tilde{B}_k(z), \ \ 0 \le k \le p+1\}.$$

The extended trivariate dQIs that we need for the construction are the following

$$P_1 f(x, y, z) := \sum_{(i,j) \in \mathcal{K}'_{mn}} f(s_i, t_j, z) B_{ij}(x, y),$$

$$P_2 f(x, y, z) := \sum_{(i,j) \in \mathcal{K}'_{mn}} f(s_i, t_j, z) \tilde{B}_{ij}(x, y),$$

$$Q_1 f(x,y,z) := \sum_{k=0}^{p+1} f(x,y,u_k) B_k(z), \quad Q_2 f(x,y,z) := \sum_{k=0}^{p+1} f(x,y,u_k) \tilde{B}_k(z).$$

For the sake of clarity, we give the expressions of P_2 and Q_2 in terms of B-splines:

$$P_2 f(x,y,z) = \sum_{(i,j)\in \mathcal{K}'_{mn}} \mu_{ij}(f) B_{ij}(x,y)$$

$$\mu_{ij}(f) = a_i f(s_{i-1}, t_j, z) + c_i f(s_{i+1}, t_j, z)) + \bar{a}_j f(s_i, t_{j-1}, z) + \bar{c}_j f(s_i, t_{j+1}, z)$$
$$+(b_i + \bar{b}_j - 1) f(s_i, t_j, z)$$

$$Q_2 f(x,y,z) := \sum_{k=0}^{p+1} \{\hat{a}_k f(x,y,u_{k-1}) + \hat{b}_k f(x,y,u_k) + \hat{c}_k f(x,y,u_{k+1})\} B_k(z)$$

We now define the trivariate blending sum

$$R = P_1 Q_2 + P_2 Q_1 - P_1 Q_1$$

Setting

$$B^{\flat}_{ijk}(x,y,z) = B_{ij}(x,y)\tilde{B}_k(z) + \tilde{B}_{ij}(x,y) B_k(z) - B_{ij}(x,y) B_k(z)$$

we obtain

$$Rf = \sum_{(i,j,k)\in \mathcal{K}_{mnp}} f(N_{ijk}) B^{\flat}_{ijk}$$

In terms of tensor product B-splines $B_{ijk} = B_{ij} B_k$, one has

$$Rf = \sum_{(i,j,k)\in \mathcal{K}_{mnp}} \nu_{ijk}(f) B_{ijk}$$

where $\nu_{ijk}(f)$ is based on the 7 neighbours of N_{ijk} in $\mathbb{R}^3$:

$$\begin{aligned} \nu_{ijk}(f) \quad = \quad & \hat{a}_k f(N_{i,j,k-1}) + \hat{c}_k f(N_{i,j,k+1}) + a_i f(N_{i-1,j,k}) + c_i f(N_{i+1,j,k}) \\ & + \bar{a}_j f(N_{i,j-1,k}) + \bar{c}_j f(N_{i,j+1,k}) + (b_i + \bar{b}_j + \hat{c}_k - 1) f(N_{ijk}). \end{aligned}$$

In [36], we proved the following

THEOREM 10. *The operator R is exact on the 15-dimensional subspace* $(\mathbb{P}_1[x,y] \otimes \mathbb{P}_2[z]) \oplus (\mathbb{P}_2[x,y] \otimes \mathbb{P}_1[z])$ *of the 18-dimensional space* $\mathbb{P}_2[x,y] \otimes \mathbb{P}_2[z]$. *Moreover, its infinite norm is bounded above independently of the nonuniform partition of the domain* Ω

$$\|R\|_\infty \leq 8.$$

6. Some applications

We present some applications of the preceding sections. For sake of simplicity, we give results for uniform partitions only. Let Q be any of the previous dQIs.
1) *Approximate integration.* Approximating $\int_\Omega f$ by $\int_\Omega Qf$ gives rise to several interesting quadrature formulas (abbr. QF) in $\mathbb{R}^d$, mainly for $d = 2, 3$. For $d = 1$ and for a uniform partition of I with meshlength h, we obtain the QF:

$$QF_n(f) = \int_a^b S_2 f = h(\frac{1}{9}f_0 + \frac{7}{8}f_1 + \frac{73}{72}f_2 + \sum_{i=3}^{n-2} f_i + \frac{73}{72}f_{n-1} + \frac{7}{8}f_n + \frac{1}{9}f_{n+1}),$$

where $f_i = f(\theta_i)$ for $0 \le i \le n+1$. This formula is exact for $\mathbb{P}_3$, like composite Simpson's formula, i.e. $QF_n(f) = \int_a^b f$ for all $f \in \mathbb{P}_3$. Therefore $\int_a^b f - QF_n(f) = O(h^4)$ for functions $f \in C^4(I)$. Numerical experiments show that it is better than Simpson's formula based on $n+1$ points (n even). Moreover, the errors associated with the two QFs have often opposite signs, thus giving upper and lower values of the exact integral.
2) *Approximate differentiation: pseudo-spectral methods.* One can approximate the first (partial) derivatives of f by those of Qf at the data sites. We thus obtain differentiation matrices which can be also used for second derivatives and for pseudo-spectral methods. Let us give an example for $d = 1$ and for a uniform partition of meshlength h of the interval I. Denoting $g = S_2 f$, then we get:

$$\begin{aligned}
g'(\theta_0) &= \frac{1}{h}\left(-\frac{8}{3}f_0 + 3f_1 - \frac{1}{3}f_2\right),\\
g'(\theta_1) &= \frac{1}{h}\left(-\frac{7}{6}f_0 + \frac{11}{16}f_1 + \frac{13}{24}f_2 - \frac{1}{16}f_3\right),\\
g'(\theta_2) &= \frac{1}{h}\left(\frac{1}{6}f_0 - \frac{3}{4}f_1 + \frac{1}{48}f_2 + \frac{5}{8}f - 3 - \frac{1}{16}f_4\right),\\
g'(\theta_{n-1}) &= \frac{1}{h}\left(\frac{1}{16}f_{n-3} - \frac{5}{8}f_{n-2} - \frac{1}{48}f_{n-1} + \frac{3}{4}f_n - \frac{1}{6}f_{n+1}\right),\\
g'(\theta_n) &= \frac{1}{h}\left(\frac{1}{16}f_{n-2} - \frac{13}{24}f_{n-1} - \frac{11}{16}f_n + \frac{7}{6}f_{n+1}\right),\\
g'(\theta_{n+1}) &= \frac{1}{h}\left(\frac{1}{3}f_{n-1} - 3f_n + \frac{8}{3}f_{n+1}\right),
\end{aligned}$$

and for $3 \le i \le n-2$:

$$g'(\theta_i) = \frac{1}{h}\left(\frac{1}{16}f_{i-2} - \frac{5}{8}f_{i-1} + \frac{5}{8}f_{i+1} - \frac{1}{16}f_{i+2}\right).$$

3) *Approximation of zeros of polynomials.* We have tested the approximation of the *Legendre polynomial* $f(x) = P_8(x)$ and of its zeros in the interval $I = [-1, 1]$ by the dQI $S_2 f$ of Section 1 based on Chebyshev points with $n = 32$. We obtain

$\|f - S_2 f\|_\infty \approx 0.0034$. There is practically no difference between the approximation $S_2 f$ and the Marsden interpolant of f which needs the solution of a linear system of $n+2$ equations. We obtain also quite good approximations of the eight roots of f in the interval. For bivariate or trivariate functions, the advantage of using dQIs over interpolants is still bigger since one avoids the solution of large linear systems. Moreover, at least in the bivariate case, one can use the nice properties of piecewise quadratic surfaces (see e.g. the results given by M.J.D. Powell in [29]).
4) *Integral equations* The dQIs can be used for various types of approximation of the solution of Fredholm type integral equation with a regular or a weakly singular kernel. This work is still in progress (see e.g. [15]).

References

[1] BARRERA D., IBANẼZ M.J. AND SABLONNIÈRE P., *Near-best quasi-interpolants on uniform and nonuniform partitions in one and two dimensions*, in: "Curve and surface fitting, Saint Malo 2002" (Eds. Cohen A., Merrien J.L. and Schumaker L.L.), Nashboro Press, Brentwood 2003, 31–40.

[2] BASZENSKI G., *n-th order polynomial spline blending*, in: "Multivariate approximation theory III", (Eds. Schempp W. and Zeller K.), ISNM **75**, 35–46, Birkhäuser Verlag, Basel 1985.

[3] BEUTEL L. AND GONSKA H., *Simultaneous approximation by tensor product operators*, Technical report SM-DU-504, Universität Duisburg, 2001.

[4] BOJANOV B.D., HAKOPIAN H.A. AND SAHAKIAN A.A., *Spline functions and multivariate interpolation*, Kluwer, Dordrecht 1993.

[5] DE BOOR C., *A practical guide to splines*, Springer-Verlag, New-York 2001 (revised edition).

[6] DE BOOR C., *Splines as linear combinations of B-splines*, in: "Approximation Theory II", (Eds. Lorentz G.G. et al.), 1–47, Academic Press, New-York 1976.

[7] DE BOOR C., *Quasiinterpolants and approximation power of multivariate splines*, in: "Computation of curves and surfaces", (Eds. Dahmen W., Gasca M. and Micchelli C.A.), 313–345, Kluwer, Dordrecht 1990.

[8] DE BOOR C., HÖLLIG K. AND RIEMENSCHNEIDER S., *Box-splines*, Springer-Verlag, New-York 1993.

[9] DE BOOR C. AND FIX G., *Spline approximation by quasi-interpolants*, J. Approx. Theory **8** (1973), 19–45.

[10] CHEN G., CHUI C.K. AND LAI M.J., *Construction of real-time spline quasi-interpolation schemes*, Approx. Theory Appl. 4 (1988), 61–75.

[11] CHUI C.K. AND LAI M.J., *A multivariate analog of Marsden's identity and a quasi-interpolation scheme*, Constr. Approx. **3** (1987), 111–122.

[12] CHUI C.K., SCHUMAKER L.L. AND WANG R.H., *On spaces of piecewise polynomials with boundary conditions III. Type II triangulations*, in: "Canadian Mathematical Society Conference Proceedings" American Mathematical Society, New York 1983, 67–80.

[13] CHUI C.K. AND WANG R.H., *Concerning C^1 B-splines on triangulations of nonuniform rectangular partitions*, Approx. Theory Appl. **1** (1984), 11-18.

[14] CHUI C.K., *Multivariate splines*, CBMS-NSF Regional Conference Series in Applied Mathematics **54**, SIAM, Philadelphia 1988.

[15] DAGNINO C. AND LAMBERTI P., *Numerical integration of 2-D integrals based on local bivariate C^1 quasi-interpolating splines*, Adv. Comput. Math. **8** (1998), 19-31.

[16] DELVOS F.J. AND SCHEMPP W., *Boolean methods in interpolation and approximation*, Longman Scientific and Technical 1989.

[17] DEVORE R.A. AND LORENTZ G.G., *Constructive approximation*, Springer-Verlag, Berlin 1993.

[18] GONSKA H., *Degree of simultaneous approximation of bivariate functions by Gordon operators*, J. Approx. Theory **62** (1990), 170–191.

[19] GONSKA H., *Simultaneous approximation by generalized n-th order blending operators*, ISNM **90**, Birkhäuser Verlag, Basel 1989, 173–180.

[20] GORDON W.J., *Distributive lattices and approximation of multivariate functions*, in: "Approximation with special emphasis on spline functions", (Ed. I.J. Schoenberg), Academic Press, New-York 1969, 223–277.

[21] IBAÑEZ-PÉREZ M.J., *Cuasi-interpolantes spline discretos con norma casi minima : teoria y aplicaciones*, PhD Thesis, Universidad de Granada, Granada 2003.

[22] KAMMERER W.J., R£DIEN G.W. AND VARGA R.S., *Quadratic interpolatory splines*, Numer. math. **22** (1974), 241–259.

[23] LEE B.G., LYCHE T. AND MORKEN K., *Some examples of quasi-interpolants constructed from local spline projectors*, in: "Innovations in Applied Mathematics", (Eds. Lyche T. and Schumaker L.L.), Vanderbilt University Press, Nashville 2001, 243–252.

[24] LYCHE T. AND SCHUMAKER L.L., *Local spline approximation*, J. Approx. Theory **15** (1975), 294–325.

[25] MARSDEN J.M. AND SCHOENBERG I.J., *An identity for spline functions with applications to variation diminishing spline approximation*, J. Approx. Theory **3** (1970), 7–49.

[26] MARSDEN J.M., *Operator norm bounds and error bounds for quadratic spline interpolation*, in: "Approximation Theory", Banach Center Publications **4** 1979, 159–175.

[27] MICULA G. AND MICULA S., *Handbook of splines*, Kluwer, Dordrecht 1999.

[28] NÜRNBERGER G., *Approximation by spline functions*, Springer-Verlag, Berlin 1989.

[29] POWELL M.J.D., *Piecewise quadratic surface fitting in contour plotting*, in: "Software for numerical mathematics", (Ed. D.J. Evans), Academic Press, London 1974, 253–271.

[30] RONG-QING JIA, *Approximation by multivariate splines: an application of boolean methods*, in: "Numerical methods of approximation theory" **9**, (Eds. Braess D. and Schumaker L.L.), ISNM **105**, Birkhäuser Verlag, Basel 1992, 117–134.

[31] SABLONNIÈRE P., *Bases de Bernstein et approximants spline*, Thèse de doctorat, Université de Lille, Lille 1982.

[32] SABLONNIÈRE P., *Bernstein-Bézier methods for the construction of bivariate spline approximants*, Comput. Aided Geom. Design **2** (1985), 29–36.

[33] SABLONNIÈRE P., *Quasi-interpolants associated with H-splines on a three-direction mesh*, J. Comput. and Appl. Math. **66** (1996), 433–442.

[34] SABLONNIÈRE P., *Quasi-interpolantes sobre particiones uniformes*, in: "Meeting in Approximation Theory", Ubeda, Spain, July 2000. Prépublication IRMAR 00-38, 2000.

[35] SABLONNIÈRE P., *H-splines and quasi-interpolants on a three directional mesh*, in "Advanced problems in constructive approximation", (Eds. Buhmann M. and Mache D.H.), ISNM **142**, Birkhäuser-Verlag, Basel 2003, 187–201.

[36] SABLONNIÈRE P., *On some multivariate quadratic spline quasi-interpolants on bounded domains*, in: "Modern developments in multivariate approximation", (Eds. Hausmann W., Jetter K., Reimer M. and Stöckler J.), ISNM **145**, Birkhäuser-Verlag, Basel 2003, 263–278.

[37] SABLONNIÈRE P., *BB-coefficients of basic bivariate quadratic splines on rectangular domains with uniform criss-cross triangulations*, in: "Modern developments in multivariate approximation", (Eds. Hausmann W., Jetter K., Reimer M. and Stöckler J.), ISNM **145**, Birkhäuser-Verlag, Basel 2003, 263–278.

[38] SABLONNIÈRE P., *BB-coefficients of bivariate quadratic B-splines on rectangular domains with non-uniform criss-cross triangulations*. Prépublication IRMAR 03-14, 2003.

[39] SABLONNIÈRE P. AND JEEAWOCK-ZEDEK F., *Hermite and Lagrange interpolation by quadratic splines on nonuniform criss-cross triangulations*, in: "Curves and Surfaces", (Eds. Laurent P.J., Le Méhauté A. and Schumaker L.L.), 445–4525, A.K. Peters, Wellesley 1991.

[40] SCHOENBERG I.J., *Cardinal spline interpolation*, CBMS-NSF Regional Conference Series in Applied Mathematics **12**, SIAM, Philadelphia 1973.

[41] SCHOENBERG I.J., *Selected papers*, Vol. 1-2, (Ed. C. de Boor) Birkhäuser-Verlag, Boston 1988.

[42] SCHUMAKER L.L., *Spline functions: basic theory*, John Wiley & Sons, New-York 1981.

[43] WARD J., *Polynomial reproducing formulas and the commutator of a locally supported spline*, in: "Topics in multivariate approximation", (Eds. Chui C.K., Schumaker L.L. and Utreras F.), Academic Press, Boston 1987, 255–263.

AMS Subject Classification: 41A15, 41A63, 65D07, 65D17.

Paul SABLONNIÈRE
INSA de Rennes
20 av. des Buttes de Coësmes
35043 Rennes Cedex, FRANCE
e-mail: `Paul.Sablonniere@insa-rennes.fr`

Rend. Sem. Mat. Univ. Pol. Torino
Vol. 61, 3 (2003)
Splines and Radial Functions

A. Iske*

RADIAL BASIS FUNCTIONS: BASICS, ADVANCED TOPICS AND MESHFREE METHODS FOR TRANSPORT PROBLEMS

Abstract. This invited contribution first reviews basic features of multivariate interpolation by radial basis functions, before selected of its advanced topics are addressed, including recent results concerning local polyharmonic spline interpolation. The latter is one main ingredient of a novel class of adaptive meshfree semi-Lagrangian methods for transport problems. The construction of these particle-based advection schemes is discussed in detail. Numerical examples concerning meshfree flow simulation are provided.

1. Introduction

Radial basis functions are well-known as traditional and powerful tools for multivariate interpolation from scattered data, see [5, 12, 13, 38, 42] for some different surveys, and [24] for a recent tutorial with accompanying exercises and supplementary software, `www.ma.tum.de/primus2001/radial/`.

Just very recently, radial basis functions have gained enormous popularity in meshfree methods for partial differential equations (PDEs). The theory includes meshfree Galerkin methods [51], collocation methods [16, 17], and multilevel schemes [15]. First applications of radial basis functions in computational fluid dynamics are dating back to Kansa [26, 27]. There is nowadays a vast amount of literature on the subject, see e.g. the rich bibliography in [15, 44]. For a couple of more recent contributions concerning radial basis functions for solving PDEs, we refer to the special issue [56].

This paper first reviews basic features of multivariate interpolation by radial basis functions in the following Section 2, before recent results concerning local polyharmonic spline interpolation are discussed in Section 3. The latter have provided recent advances in the numerical simulation of transport processes by meshfree particle methods [1, 2, 3]. Details on these are explained in Section 4, and numerical examples concerning meshfree flow simulation are finally presented in Section 5.

*This paper is based on an invited lecture which I gave at the workshop *Spline Functions and Radial Functions: Applications to Integral and Differential Problems* of the GNCS, held at the University of Turin in February 2003. I wish to thank the organizers of the meeting for their generous support and their kind hospitality. Moreover, the assistance of Martin Käser with the preparation of the numerical examples is gratefully appreciated.

2. Radial basis function interpolation

2.1. Interpolation scheme

In order to explain multivariate scattered data interpolation by radial basis functions, suppose a data vector $u\big|_\Xi = (u(\xi_1), \ldots, u(\xi_n))^T \in \mathbb{R}^n$ of function values, sampled from an unknown function $u : \mathbb{R}^d \to \mathbb{R}$ at a *scattered* finite point set $\Xi = \{\xi_1, \ldots, \xi_n\} \subset \mathbb{R}^d$, $d \geq 1$, is given. Scattered data interpolation requirescomputing a *suitable* interpolant $s : \mathbb{R}^d \to \mathbb{R}$ satisfying $s\big|_\Xi = u\big|_\Xi$, i.e.,

$$s(\xi_j) = u(\xi_j), \qquad \text{for all } 1 \leq j \leq n. \tag{1}$$

To this end, the radial basis function interpolation scheme works with a fixed *radial* function $\phi : [0, \infty) \to \mathbb{R}$, and the interpolant s in (1) is assumed to have the form

$$s(x) = \sum_{j=1}^{n} c_j \phi(\|x - \xi_j\|) + p(x), \qquad p \in \mathcal{P}_m^d, \tag{2}$$

where $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^d$. Moreover, $\mathcal{P}_m^d$ denotes the linear space containing all real-valued polynomials in d variables of degree at most $m - 1$, where $m \equiv m(\phi)$ is said to be the *order* of the basis function ϕ. We come back to the dependence between m and ϕ later in Subsection 2.4. But let us first give some examples for ϕ.

Classical choices for radial basis functions ϕ, along with their order m, are shown in Table 1, where for any $x \in \mathbb{R}$, the symbol $\lceil x \rceil$ denotes as usual the smallest integer greater than or equal to x. Later in this text, $\lfloor x \rfloor$ denotes the largest integer less than or equal to x.

Among the most popular radial basis functions are the *polyharmonic splines*, which are discussed more detailed in Section 3. This class of radial basis functions includes the *thin plate splines*, where $\phi(r) = r^2 \log(r)$ and $m = 2$, which are particularly suited for interpolation from planar scattered data. Further commonly used radial basis functions are given by the *Gaussians*, $\phi(r) = \exp(-r^2)$, the *multiquadrics*, $\phi(r) = (1 + r^2)^{1/2}$ of order $m = 1$, and the *inverse multiquadrics*, $\phi(r) = (1 + r^2)^{-1/2}$, where $m = 0$. Table 1 gives a more general form for the (inverse) multiquadrics and their corresponding order m.

2.2. Compactly supported radial basis functions

More recent developments [50, 53] have provided a whole family of *compactly supported* radial basis functions. In this case, we have $m = 0$ for their order, and so the polynomial part in (2) is omitted. While the radial basis functions in Table 1 can be used in arbitrary space dimension d, the selection of one *suitable* compactly supported ϕ depends on d, see Table 2. Since the dimension d is known beforehand, this is no severe restriction, as shall be established below.

Table 1: Radial basis functions.

Radial Basis Function	$\phi(r) =$	**Parameters**	**Order**
Polyharmonic Splines	r^{ν}	$\nu > 0, \nu \notin 2\mathbb{N}$	$m = \lceil \nu/2 \rceil$
	$r^{2k}\log(r)$	$k \in \mathbb{N}$	$m = k + 1$
Gaussians	$\exp(-r^2)$		$m = 0$
Multiquadrics	$\left(1 + r^2\right)^{\nu}$	$\nu > 0, \nu \notin \mathbb{N}$	$m = \lceil \nu \rceil$
Inverse Multiquadrics	$\left(1 + r^2\right)^{\nu}$	$\nu < 0$	$m = 0$

To this end, let us further discuss some basics about compactly supported radial basis functions. As to Wendland's functions [50], these are of the form

$$\phi_{d,k}(r) = \begin{cases} p_{d,k}, & \text{for } 0 \le r \le 1, \\ 0, & \text{for } r > 1, \end{cases} \tag{3}$$

where $p_{d,k}$ is a specific univariate polynomial of degree $\lfloor d/2 \rfloor + 3k + 1$, and so the support $\text{supp}(\phi_{d,k})$ of $\phi_{d,k} : [0, \infty) \to \mathbb{R}$ is normalized to the unit interval $[0, 1]$. Moreover, due to Wendland's construction in [50], the basis function $\phi_{d,k}$ has derivatives up to order $2k$, i.e., $\phi_{d,k} \in C^{2k}(\mathbb{R}^d)$. Possible choices for $\phi_{d,k}$ are listed in the following Table 2, where the symbol $\doteq$ denotes equality up to a positive factor, and the *truncated power function* $(\cdot)_+ : \mathbb{R} \to [0, \infty)$ is given by $(x)_+ = x$, for $x > 0$, and $(x)_+ = 0$, for $x \le 0$.

By their construction, Wendland's radial basis functions $\phi_{d,k}$ are *positive definite* on $\mathbb{R}^d$.

DEFINITION 1. *A continuous radial function* $\phi : [0, \infty) \to \mathbb{R}$ *is said to be* positive definite *on* $\mathbb{R}^d$, $\phi \in \mathbf{PD}_d$, *iff for any finite set* $\Xi = \{\xi_1, \dots, \xi_n\}$, $\Xi \subset \mathbb{R}^d$, *of pairwise distinct points the matrix*

$$\Phi_{\phi,\Xi} = \left(\phi(\|\xi_j - \xi_k\|)\right)_{1 \le j,k \le n} \in \mathbb{R}^{n \times n}$$

is positive definite.

Due to the construction in [53], there exists, for any space dimension d, a positive definite and compactly supported $\phi \in \mathbf{PD}_d$ of the form (3). Remarkably enough,

Table 2: Wendland's compactly supported radial basis functions [50].

Dimension d	**Radial Basis Function**	**Smoothness** $2k$
$d = 1$	$\phi_{1,0} = (1-r)_+$ $\phi_{1,1} \doteq (1-r)_+^3(3r+1)$ $\phi_{1,2} \doteq (1-r)_+^5(8r^2+5r+1)$	C^0 C^2 C^4
$d \le 3$	$\phi_{3,0} = (1-r)_+^2$ $\phi_{3,1} \doteq (1-r)_+^4(4r+1)$ $\phi_{3,2} \doteq (1-r)_+^6(35r^2+18r+3)$ $\phi_{3,3} \doteq (1-r)_+^8(32r^3+25r^2+8r+1)$	C^0 C^2 C^4 C^6
$d \le 5$	$\phi_{5,0} = (1-r)_+^3$ $\phi_{5,1} \doteq (1-r)_+^5(5r+1)$ $\phi_{5,2} \doteq (1-r)_+^7(16r^2+7r+1)$	C^0 C^2 C^4

Wendland showed that any basis function $\phi_{d,k}$, constructed in [50] (such as any in Table 2), has minimal degree among all positive definite functions $\phi \in \mathbf{PD}_d \cap C^{2k}(\mathbb{R}^d)$ of the form (3). Moreover, by these properties, $\phi_{d,k}$ in (3) is unique up to a positive constant.

2.3. Well-posedness of the interpolation problem

Now let us turn to the well-posedness of the interpolation problem (1). To this end, we distinguish the case, where $m = 0$ from the one where $m > 0$.

First suppose $m = 0$ for the order of the basis function ϕ, such as for the *Gaussians*, the *inverse multiquadrics* (in Table 1) and *Wendland's functions* (in Table 2). In this case, the interpolant s in (2) has the form

$$s(x) = \sum_{j=1}^{n} c_j \phi(\|x - \xi_j\|). \tag{4}$$

By requiring the n interpolation conditions in (1), the computation of the unknown coefficients $c = (c_1, \ldots, c_n)^T \in \mathbb{R}^n$ of s in (4) amounts to solving the linear equation system

$$\Phi_{\phi,\Xi} \cdot c = u\big|_{\Xi}. \tag{5}$$

Recall that according to Definition 1, the matrix $\Phi_{\phi,\Xi}$ in (5) is guaranteed to be positive definite, provided that $\phi \in \mathbf{PD}_d$. In this case, the system (5) has a unique solution. This in turn implies the well-posedness of the given interpolation problem already.

THEOREM 1. *For $\phi \in \mathbf{PD}_d$, the interpolation problem (1) has a unique solution s of the form (4).*

Now let us turn to the case, where $m > 0$ for the order of ϕ. In this case, the interpolant s in (2) contains a nontrivial polynomial part, yielding q additional degrees of freedom, where $q = \binom{m-1+d}{d}$ is the dimension of the polynomial space $\mathcal{P}_m^d$. These additional degrees of freedom are usually eliminated by requiring the q *moment conditions*

$$\sum_{j=1}^{n} c_j p(\xi_j) = 0, \qquad \text{for all } p \in \mathcal{P}_m^d. \tag{6}$$

Altogether, this amounts to solving the linear system

$$\begin{bmatrix} \Phi_{\phi,\Xi} & \Pi_\Xi \\ \Pi_\Xi^T & 0 \end{bmatrix} \cdot \begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} u\big|_\Xi \\ 0 \end{bmatrix}, \tag{7}$$

where we let $\Pi_\Xi = \left((\xi_j)^\alpha\right)_{1 \le j \le n; |\alpha| < m} \in \mathbb{R}^{n \times q}$, and $d = (d_\alpha)_{|\alpha|<m} \in \mathbb{R}^q$ for the coefficients of the polynomial part in (2). Moreover, for any point $x = (x_1, \ldots, x_d)^T \in \mathbb{R}^d$, and multi-index $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}_0^d$ we let $x^\alpha = x_1^{\alpha_1} \cdot \cdots \cdot x_d^{\alpha_d}$, and $|\alpha| = \alpha_1 + \ldots + \alpha_d$.

In order to analyze the existence and uniqueness of a solution of (7), we first consider its corresponding *homogeneous* system

$$\Phi_{\phi,\Xi} \cdot c + \Pi_\Xi \cdot d = 0, \tag{8}$$

$$\Pi_\Xi^T \cdot c = 0, \tag{9}$$

here split into its interpolation conditions (8) and moment conditions (9). If we multiply the equation (8) from left with c^T, and by using the moment conditions (9), we immediately obtain the identity

$$c^T \cdot \Phi_{\phi,\Xi} \cdot c = 0. \tag{10}$$

Now in order to guarantee the existence of a solution to (8),(9) we require that the matrix $\Phi_{\phi,\Xi} \in \mathbb{R}^{n \times n}$ is, for any set Ξ of interpolation points, *positive definite* on the linear subspace of $\mathbb{R}^d$ containing all vectors $c \in \mathbb{R}^n$ satisfying (9), i.e.,

$$c^T \cdot \Phi_{\phi,\Xi} \cdot c > 0, \qquad \text{for all } c \in \mathbb{R}^n \setminus \{0\} \text{ with } \Pi_\Xi^T c = 0. \tag{11}$$

In this case, the basis function ϕ is said to be *conditionally positive definite*, which deserves the following definition.

DEFINITION 2. *A continuous radial function $\phi : [0, \infty) \to \mathbb{R}$ is said to be* conditionally positive definite *of order m on $\mathbb{R}^d$, $\phi \in \mathbf{CPD}_d(m)$, iff (11) holds for all possible choices of finite point sets $\Xi \subset \mathbb{R}^d$.*

As shall be established in the following Subsection 2.4, we remark that for every radial basis function ϕ in Table 1, we either have $\phi \in \mathbf{CPD}_d(m)$ or $-\phi \in \mathbf{CPD}_d(m)$, with the corresponding order m given in the last column of Table 1. In either case, we say that m is the *order* of the radial basis function ϕ. Note that every positive definite ϕ, such as for instance any of Wendland's functions in Table 2, is conditionally positive definite of order $m = 0$, and therefore $\mathbf{PD}_d = \mathbf{CPD}_d(0)$.

Now let us return to the above discussion concerning the solvability of the linear system (8),(9). With assuming $\phi \in \mathbf{CPD}_d(m)$ (or $-\phi \in \mathbf{CPD}_d(m)$), we conclude $c = 0$ directly from (10), and so (8) becomes $\Pi_\Xi \cdot d = 0$. Therefore, in order to guarantee a unique solution of (8),(9), it remains to require the *injectivity* of the matrix Π_Ξ. But this property depends on the geometry of the interpolation points in Ξ. Indeed, note that the matrix Π_Ξ is injective, iff for $p \in \mathcal{P}_m^d$ the implication

$$
p(\xi_j) = 0 \quad \text{for } 1 \le j \le n \qquad \Longrightarrow \qquad p \equiv 0 \tag{12}
$$

holds. In this case, any polynomial in $\mathcal{P}_m^d$ can uniquely be reconstructed from its function values sampled at the points in Ξ. The point set Ξ is then said to be $\mathcal{P}_m^d$*-unisolvent*. Note that the requirement (12) for the points in Ξ is rather weak. Indeed, when $m = 0$, the condition is empty, for $m = 1$ it is trivial, and for $m = 2$ the points in Ξ must not lie on a straight line.

We summarize the discussion of this subsection as follows.

THEOREM 2. *For $\phi \in \mathbf{CPD}_d(m)$, the interpolation problem (1) has under constraints (6) a unique solution s of the form (2), provided that the interpolation points in Ξ are $\mathcal{P}_m^d$-unisolvent by satisfying (12).*

2.4. Conditionally positive definite functions

By the discussion in the previous subsection, radial basis function interpolation essentially relies on the conditional positive definiteness the chosen basis function ϕ. Indeed, this is one of the key properties of the interpolation scheme. In this subsection, we discuss two alternative ways for the construction and characterization of conditionally positive definite functions.

One technique, dating back to Micchelli [35], works with *completely monotone functions*. The other alternative relies on *generalized Fourier transforms* [23]. We do not intend to discuss these two different techniques in all details. Instead of this we briefly review relevant results. For a more comprehensive discussion concerning conditionally positive definite functions, we refer to the recent survey [45].

Completely monotone functions

DEFINITION 3. *A function* $\psi \in C^{\infty}(0,\infty)$ *is said to be* completely monotone *on* $(0,\infty)$*, iff*

$$(-1)^{\ell}\psi^{\ell}(r) \geq 0, \qquad \ell = 0, 1, 2, \ldots,$$

holds for all $r \in (0,\infty)$.

Micchelli provides in [35] a sufficient criterion for $\phi \in \mathbf{CPD}_d(m)$, which generalizes an earlier result by Schoenberg [47, 48] for positive definite radial functions. Micchelli also conjectured the necessity of this criterion. This was finally shown by Guo, Hu and Sun in [20]. We summarize the relevant results from [20, 35, 47, 48] by

THEOREM 3. *Let* $\phi : [0,\infty) \to \mathbb{R}$ *be a continuous radial function. Moreover, let* $\phi_{\sqrt{}} \equiv \phi(\sqrt{\cdot})$*. Suppose* $\phi_m \equiv (-1)^m \phi_{\sqrt{}}^{(m)}$ *is well-defined and* ϕ_m *is not constant. Then, the following two statements are equivalent.*

(a) $\phi \in \mathbf{CPD}_d(m)$ *for all* $d \geq 1$*;*

(b) ϕ_m *is completely monotone on* $(0,\infty)$.

Now, by using Theorem 3, it is easy to show for any ϕ in Table 1 that either ϕ or $-\phi$ is conditionally positive definite of order m, with m given in the last column of Table 1. Note, however, that the characterization in Theorem 3 applies to *radial* functions only. Moreover, it excludes the construction of *compactly supported* radial basis functions. The latter is due to the Bernstein-Widder theorem [4] (see also [52]) which says that any function $\psi : [0,\infty) \to \mathbb{R}$ is completely monotone on $(0,\infty)$, if and only if it has a Laplace-Stieltjes-type representation of the form

$$\psi(r) = \int_0^{\infty} \exp(-rs)\, d\mu(s),$$

where μ is monotonically increasing with $\int_0^{\infty} d\mu(s) < \infty$. Hence, in this case ψ has no zero, and so any $\psi = \phi_m$ in (b) of Theorem 3 cannot be compactly supported.

Generalized Fourier transforms

A different technique for the characterization and construction of (not necessarily radial) functions $\phi \in \mathbf{CPD}_d(m)$, including compactly supported ones, is using (generalized) Fourier transforms, see the recent survey [45, Section 4] (which basically relies on the results in [23]). We do not explain generalized Fourier transforms here, but rather refer to the textbooks [18, 19], where a comprehensive treatment of the relevant technical background is provided.

For the purposes in this subsection, it is sufficient to say that every radial basis function ϕ in Table 1 has a *radial* (generalized) Fourier transform $\hat{\phi} \in C(0,\infty)$ satisfying the following two properties.

- $\hat{\phi}(\|\cdot\|)$ is L_1-integrable around infinity, i.e.,

$$\int_{\mathbb{R}^d \setminus B_1(0)} \left|\hat{\phi}(\|\omega\|)\right| d\omega < \infty, \tag{13}$$

- $\hat{\phi}(\|\cdot\|)$ has at most an algebraic singularity of *order* $s_0 \in \mathbb{N}_0$ at the origin, such that

$$\int_{B_1(0)} \|\omega\|^{s_0} \hat{\phi}(\|\omega\|)\, d\omega < \infty, \tag{14}$$

  holds, with $s_0 \in \mathbb{N}_0$ being minimal in (14).

Table 3 shows the (generalized) Fourier transforms of the radial basis functions in Table 1, along with their order s_0, where $\doteq$ means equality up to a constant factor, and where K_δ denotes the modified Bessel function.

We remark that if ϕ has a Fourier transform $\hat{\phi} \in L_1(\mathbb{R}^d)$,

$$\hat{\phi}(\|\omega\|) = \int_{\mathbb{R}^d} \phi(\|x\|) \exp(-i x^T \omega)\, dx,$$

in the classical sense, then this *classical* Fourier transform $\hat{\phi}$ coincides with the generalized Fourier transform of ϕ. Examples are given by the Gaussians, the inverse multiquadrics, and Wendland's compactly supported radial basis functions. In this case, we have $s_0 = 0$ for the order of $\hat{\phi}$.

Now let us turn straight to the characterization of conditionally positive definite functions by generalized Fourier transforms. This particular characterization relies on the identity

$$\sum_{j,k=1}^{n} c_j c_k \phi(\|\xi_j - \xi_k\|) = (2\pi)^{-d} \int_{\mathbb{R}^d} \hat{\phi}(\|\omega\|) \left| \sum_{j=1}^{n} c_j \exp(-i\xi_j^T \omega) \right|^2 d\omega, \tag{15}$$

which can be established [23] for any $\hat{\phi}$ satisfying (13) and (14), provided that the *symbol function*

$$\sigma_{c,\Xi}(\omega) = \sum_{j=1}^{n} c_j \exp(-i\xi_j^T \omega) \tag{16}$$

has a zero at the origin of order at least $m = \lceil s_0/2 \rceil$. Note that the latter can be guaranteed by requiring the moment conditions (6) with $m = \lceil s_0/2 \rceil$.

THEOREM 4. *A continuous radial function $\phi : [0, \infty) \to \mathbb{R}$ is conditionally positive definite on $\mathbb{R}^d$, if ϕ has a continuous nonnegative generalized Fourier transform $\hat{\phi} \not\equiv 0$ satisfying (13) and (14). In this case, we have $m = \lceil s_0/2 \rceil$ for the order of $\phi \in \mathbf{CPD}_d(m)$.*

Proof. Let $\hat{\phi}$ satisfy (13) and (14), and suppose (6) with $m = \lceil s_0/2 \rceil$, so that the identity (15) holds. By the nonnegativity of $\hat{\phi}$, the quadratic form

$$c^T \Phi_{\phi,\Xi} c = \sum_{j,k=1}^{n} c_j c_k \phi(\|\xi_j - \xi_k\|)$$

appearing in the left hand side of (15), is nonnegative. Hence it remains to show that $c^T \Phi_{\phi,\Xi} c$ vanishes, if and only if $c = 0$. In order to see this, suppose that $c^T \Phi_{\phi,\Xi} c$, and thus the right hand side in (15), vanishes. In this case, the symbol function $\sigma_{c,\Xi}$ in (16) must vanish on an open subset of $\mathbb{R}^d$ with nonempty interior. But then, due to the analyticity of $\sigma_{c,\Xi}$, this implies that the symbol function vanishes identically on $\mathbb{R}^d$, i.e., $\sigma_{c,\Xi} \equiv 0$. Since the points in Ξ are pairwise distinct, and so the exponentials $\exp(-i\xi_j^T \omega)$ are linearly independent, the latter is true, if and only if $c = 0$.

□

Table 3: Generalized Fourier transforms of radial basis functions.

Radial Basis Function	$\phi(r) =$	$\hat{\phi}(s) \doteq$	**Order** s_0
Polyharmonic Splines	r^ν	$s^{-d-\nu}$	$\lfloor \nu \rfloor + 1$
	$r^{2k}\log(r)$	s^{-d-2k}	$2k+1$
Gaussians	$\exp(-r^2)$	$\exp(-s^2/4)$	0
Multiquadrics	$\left(1+r^2\right)^\nu$	$K_{d/2+\nu}(s) \cdot s^{-(d/2+\nu)}$	$\lfloor 2\nu \rfloor + 1$
Inverse Multiquadrics	$\left(1+r^2\right)^\nu$	$K_{d/2+\nu}(s) \cdot s^{-(d/2+\nu)}$	0

2.5. Error estimates in native function spaces

This subsection is devoted to available bounds on the error $\|u - s_{u,\Xi}\|_{L_\infty(\Omega)}$, where $\Omega \subset \mathbb{R}^d$ is a bounded and open domain comprising Ξ, i.e., $\Xi \subset \Omega$. Moreover, it is assumed that $\Omega \subset \mathbb{R}^d$ satisfies an *interior cone condition*, and u lies in the *native function space* $\mathcal{F}_\phi$ associated with the radial basis function $\phi \in \mathbf{CPD}_d(m)$.

In order to explain the native function space $\mathcal{F}_\phi$ just very briefly, let

$$\mathcal{R}_\phi = \left\{ s_{c,p,\Xi} = \sum_{\xi\in\Xi} c_\xi \phi(\|\cdot - \xi\|) + p : \Xi \subset \mathbb{R}^d \text{ finite }, \Pi_\Xi^T c = 0, p \in \mathcal{P}_m^d \right\}$$

denote the *recovery space* of $\phi \in \mathbf{CPD}_d(m)$ containing all possible interpolants of the form (2). Due to its conditional positive definiteness, ϕ provides by

$$(s_{c,p,\Xi}, s_{d,q,\Upsilon})_\phi = \sum_{\xi\in\Xi, \upsilon\in\Upsilon} c_\xi d_\upsilon \phi(\|\xi - \upsilon\|), \qquad \text{for } s_{c,p,\Xi}, s_{d,q,\Upsilon} \in \mathcal{R}_\phi,$$

a semi-inner product $(\cdot,\cdot)_\phi$, and semi-norm $|\cdot|_\phi = (\cdot,\cdot)_\phi^{1/2}$, whose kernel are the polynomials in $\mathcal{P}_m^d$. The topological closure of the linear space $(\mathcal{R}_\phi, |\cdot|_\phi)$ is the native function space $\mathcal{F}_\phi$, i.e., $\overline{\mathcal{R}_\phi} = \mathcal{F}_\phi$.

One key feature of the radial basis function interpolation scheme is its *optimal recovery*, which can be explained as follows. For $u \in \mathcal{F}_\phi$ and any finite point set $\Xi \subset \mathbb{R}^d$, the interpolant $s_{u,\Xi}$ satisfying $s_{u,\Xi}\big|_\Xi = u\big|_\Xi$ is the orthogonal projection of u onto the recovery space $\mathcal{R}_\phi \subset \mathcal{F}_\phi$, so that the *Pythagoras theorem*

$$|s_{u,\Xi}|_\phi^2 + |u - s_{u,\Xi}|_\phi^2 = |u|_\phi^2, \qquad \text{for } u \in \mathcal{F}_\phi,$$

holds. Hence, by

$$|s_{u,\Xi}|_\phi^2 \le |u|_\phi^2, \qquad \text{for } u \in \mathcal{F}_\phi,$$

the interpolation process is *optimal* w.r.t. the *optimal recovery space* $\mathcal{F}_\phi$. For more details on this, we refer to the variational theory in the seminal papers by Madych & Nelson [29, 30, 31].

Now let us turn to error estimates. For the radial basis functions in Table 1, available bounds on the pointwise error $\epsilon_x = u(x) - s(x)$, $x \in \Omega$, are due to [30, 31, 54] of the form

$$|u(x) - s_{u,\Xi}(x)| \le C \cdot |u|_\phi \cdot F_\phi^{1/2}(h_{\varrho,\Xi}(x)), \qquad \text{for } u \in \mathcal{F}_\phi, \tag{17}$$

where, for some specific radius $\varrho > 0$, the *local fill distance*

$$h_{\varrho,\Xi}(x) = \max_{y\in B_\varrho(x)} \min_{\xi\in\Xi} \|y - \xi\|$$

reflects the local density of Ξ around x, where $B_\varrho(x) = \{y : \|y - x\| \le \varrho\}$. Moreover, $F_\phi : [0,\infty) \to [0,\infty)$ is a monotonically increasing function with $F_\phi(0) = 0$, which depends merely on ϕ. For the radial basis functions ϕ in Table 1, its corresponding F_ϕ is listed in Table 4, see also [42].

It can be shown that the given pointwise error bounds carry over to uniform bounds in the domain Ω, yielding error estimates depending on the *fill distance*

$$h_{\Xi,\Omega} = \max_{y\in\Omega} \min_{\xi\in\Xi} \|y - \xi\| \tag{18}$$

Table 4: Radial basis functions: Convergence Rates (see [42] for details).

Radial Basis Function	$\phi(r) =$	$F_\phi(h) \doteq$
Polyharmonic Splines	r^ν $r^{2k}\log(r)$	h^ν h^{2k}
Gaussians	$\exp(-r^2)$	$\exp(-\alpha/h)$
(Inverse) Multiquadrics	$\left(1+r^2\right)^\nu$	$\exp(-\alpha/h)$

of Ξ in Ω, i.e.,

$$\|u - s_{u,\Xi}\|_{L_\infty(\Omega)} \leq C \cdot |u|_\phi \cdot F_\phi^{1/2}(h_{\Xi,\Omega}), \qquad \text{for } u \in \mathcal{F}_\phi. \tag{19}$$

For further details, we refer to [42, 46].

2.6. Lagrange representation of the interpolant

In the following discussion of this paper, especially in the following Section 3, it is convenient to work with the *Lagrange representation*

$$s_{u,\Xi}(x) = \sum_{j=1}^{n} \lambda_j(x) u(\xi_j) \tag{20}$$

of the interpolant $s \equiv s_{u,\Xi}$ in (2), where the *Lagrange basis functions* $\lambda_1(x), \ldots, \lambda_n(x)$ satisfy

$$\lambda_j(\xi_k) = \begin{cases} 1, & \text{for } j = k, \\ 0, & \text{for } j \neq k \end{cases} \qquad 1 \leq j, k \leq n, \tag{21}$$

and so $s\big|_\Xi = u\big|_\Xi$.

For a fixed $x \in \mathbb{R}^d$, the vectors

$$\lambda(x) = (\lambda_1(x), \ldots, \lambda_n(x))^T \in \mathbb{R}^n \quad \text{and} \quad \mu(x) = (\mu_1(x), \ldots, \mu_q(x))^T \in \mathbb{R}^q$$

are the unique solution of the linear system

$$\begin{bmatrix} \Phi_{\phi,\Xi} & \Pi_\Xi \\ \Pi_\Xi^T & 0 \end{bmatrix} \cdot \begin{bmatrix} \lambda(x) \\ \mu(x) \end{bmatrix} = \begin{bmatrix} \varphi(x) \\ \pi(x) \end{bmatrix}, \tag{22}$$

where $\varphi(x) = (\phi(\|x - \xi_j\|))_{1 \le j \le n} \in \mathbb{R}^n$ and $\pi(x) = (x^\alpha)_{|\alpha|<m} \in \mathbb{R}^q$. We abbreviate this linear system as

$$A \cdot \nu(x) = \beta(x)$$

by letting

$$A = \begin{bmatrix} \Phi_{\phi,\Xi} & \Pi_\Xi \\ \Pi_\Xi^T & 0 \end{bmatrix}, \quad \nu(x) = \begin{bmatrix} \lambda(x) \\ \mu(x) \end{bmatrix}, \quad \beta(x) = \begin{bmatrix} \varphi(x) \\ \pi(x) \end{bmatrix}.$$

This allows us to combine the two alternative representations for s in (20) and (2) by

$$\begin{aligned} s(x) &= < \lambda(x), u\big|_\Xi > \\ &= < \nu(x), u_\Xi > \\ &= < A^{-1} \cdot \beta(x), u_\Xi > \\ &= < \beta(x), A^{-1} \cdot u_\Xi > \\ &= < \beta(x), b >, \end{aligned} \tag{23}$$

where $< \cdot, \cdot >$ denotes the inner product of the Euclidean space $\mathbb{R}^d$, and where we let

$$u_\Xi = \begin{bmatrix} u\big|_\Xi \\ 0 \end{bmatrix} \in \mathbb{R}^{n+q} \quad \text{and} \quad b = \begin{bmatrix} c \\ d \end{bmatrix} \in \mathbb{R}^{n+q}$$

for the right hand side and the solution of the linear system (7).

3. Polyharmonic spline interpolation

In this section, details on the interpolation by *polyharmonic splines*, often also referred to as *surface splines*, are explained. The utility of polyharmonic splines for multivariate interpolation was established by Duchon [8, 9, 10]. In order to discuss the particular setting of Duchon, let us be more specific about the choice of the basis function ϕ. According to [8, 9, 10], we assume from now the form

$$\phi_{d,k}(r) = \begin{cases} r^{2k-d} \log(r), & \text{for } d \text{ even}, \\ r^{2k-d}, & \text{for } d \text{ odd}, \end{cases}$$

for the polyharmonic splines, where k is required to satisfy $2k > d$. According to Table 1 (last column), the order of $\phi_{d,k}$ is given by $m = k - \lceil d/2 \rceil + 1$.

Now note that the inclusion $\mathbf{CPD}_d(m_1) \subset \mathbf{CPD}_d(m_2)$, for $m_1 \le m_2$, allows us to also work with any order greater than m. In order to comply with Duchon's setting, we replace the *minimal* choice $m = k - \lceil d/2 \rceil + 1$ by $k \ge m$. Therefore, we let from now $m = k$ for the order of $\phi_{d,k} \in \mathbf{CPD}_d(m)$. We come back with an explanation concerning this particular choice for m later in Subsection 3.1.

With using $m = k$, the resulting interpolant in (2) has the form

$$s(x) = \sum_{i=1}^{n} c_i \phi_{d,k}(\|x - \xi_i\|) + \sum_{|\alpha|<k} d_\alpha x^\alpha. \tag{24}$$

We remark that the polyharmonic spline $\phi_{d,k}$ is the fundamental solution of the k-th iterated Laplacian, i.e.,

$$\Delta^k \phi_{d,k}(\|x\|) = c\delta_x.$$

For instance, for $d = k = 2$, the thin plate spline $\phi_{2,2}(r) = r^2 \log(r)$ solves the *biharmonic equation*

$$\Delta\Delta\phi_{2,2}(\|x\|) = c\delta_x.$$

In this case, the interpolant s in (24) has the form

$$s(x) = \sum_{i=1}^{n} c_i \|x - \xi_i\|^2 \log(\|x - \xi_i\|) + d_1 + d_2 x_1 + d_3 x_2, \tag{25}$$

where we let $x = (x_1, x_2)^T \in \mathbb{R}^2$. Finally, we remark that for the univariate case, where $d = 1$, the polyharmonic spline $\phi_{1,k} = r^{2k-1}$, $k \geq 1$, coincides with the *natural spline* of order $2k$.

3.1. Optimal recovery in Beppo-Levi spaces

Recall the discussion in Subsection 2.5 concerning optimal recovery of radial basis function interpolation in native function spaces. In this subsection, we wish to discuss the native function space of polyharmonic splines.

Due to fundamental results in the seminal papers [8, 9, 10] of Duchon and [32, 33, 34] of Meinguet, for a fixed finite point set $\Xi \subset \mathbb{R}^d$, an interpolant s in (24) minimizes the energy

$$|u|^2_{\mathrm{BL}^k(\mathbb{R}^d)} = \int_{\mathbb{R}^d} \sum_{|\alpha|=k} \begin{pmatrix} k \\ \alpha \end{pmatrix} \left(D^\alpha u\right)^2 dx, \qquad \begin{pmatrix} k \\ \alpha \end{pmatrix} = \frac{k!}{\alpha_1! \cdot \cdots \cdot \alpha_d!}, \tag{26}$$

among all functions u of the *Beppo-Levi space*

$$\mathrm{BL}^k(\mathbb{R}^d) = \left\{ u \in C(\mathbb{R}^d) : D^\alpha u \in L^2(\mathbb{R}^d) \text{ for all } |\alpha| = k \right\} \subset C(\mathbb{R}^d)$$

satisfying $u\big|_\Xi = s\big|_\Xi$. So the Beppo-Levi space $\mathrm{BL}^k(\mathbb{R}^d)$ is equipped with the semi-norm $|\cdot|_{\mathrm{BL}^k(\mathbb{R}^d)}$, whose kernel is the polynomial space $\mathcal{P}_k^d$. The latter explains why we use order $m = k$ rather than the *minimal* choice $m = k - \lceil d/2 \rceil + 1$. In this case, the Beppo-Levi space $\mathrm{BL}^k(\mathbb{R}^d)$ is the *optimal recovery space* $\mathcal{F}_\phi$ for the polyharmonic splines $\phi_{d,k}$. Note that $\mathrm{BL}^k(\mathbb{R}^d)$ is the Sobolev space $H^k(\mathbb{R}^d)$.

When working with thin plate splines, $\phi_{2,2}(r) = r^2 \log(r)$, in two dimensions we have

$$|u|^2_{\mathrm{BL}^2(\mathbb{R}^2)} = \int_{\mathbb{R}^2} \left(u^2_{x_1x_1} + 2u^2_{x_1x_2} + u^2_{x_2x_2} \right) dx_1\, dx_2, \quad \text{for } u \in \mathrm{BL}^2(\mathbb{R}^2).$$

In this case, the semi-norm $|\cdot|_{\mathrm{BL}^2(\mathbb{R}^2)}$ is the *bending energy* of a thin plate of infinite extent, and this explains the naming of *thin plate splines*.

3.2. Approximation order

The following result concerning the convergence rate of polyharmonic spline interpolation is dating back to Wu & Schaback [54] (cf. Subsection 2.5).

THEOREM 5. *Let Ω be a bounded and open domain satisfying an interior cone condition. Then, there exist constants h_0, C, such that for any finite point set $\Xi \subset \Omega$ satisfying $h_{\Xi,\Omega} \le h_0$ and any function $u \in \mathrm{BL}^k(\mathbb{R}^d)$ the error bound*

$$\|u - s\|_{L_\infty(\Omega)} \le C \cdot |u|_{\mathrm{BL}^k(\mathbb{R}^d)} h^{k-d/2}_{\Xi,\Omega}$$

holds, where s is the polyharmonic spline interpolant in (24), using $\phi_{d,k}$, satisfying $s\big|_\Xi = u\big|_\Xi$.

Hence, in this sense, the *global* approximation order of the polyharmonic spline interpolation scheme, using $\phi_{d,k}$, is $p = k - d/2$ with respect to the Beppo-Levi space $\mathrm{BL}^k(\mathbb{R}^d)$.

In the following discussion of this subsection, we analyze the approximation order of *local* polyharmonic spline interpolation. We remark that this analysis in combination with the subsequent discussion concerning the stability of local polyharmonic spline interpolation is relevant for the application in the following Section 4.

As regards the local approximation order, we consider solving, for some fixed point $\xi_0 \in \mathbb{R}^d$ and any $h > 0$, the interpolation problem

$$u(\xi_0 + h\xi_j) = s^h(\xi_0 + h\xi_j), \quad 1 \le j \le n, \tag{27}$$

where $\Xi = \{\xi_1, \ldots, \xi_n\} \subset \mathbb{R}^d$ is a $\mathcal{P}^d_k$-unisolvent point set of *moderate* size, i.e., n is small. Moreover, s^h denotes the unique polyharmonic spline interpolant of the form

$$s^h(hx) = \sum_{j=1}^n c^h_j \phi_{d,k}(\|hx - h\xi_j\|) + \sum_{|\alpha|<k} d^h_\alpha (hx)^\alpha \tag{28}$$

satisfying (27). The discussion in this subsection is dominated by the following definition.

DEFINITION 4. *Let s^h denote the polyharmonic spline interpolant, using $\phi_{d,k}$, satisfying (27). We say that the* approximation order *of local polyharmonic spline interpolation at $\xi_0 \in \mathbb{R}^d$ and with respect to the function space $\mathcal{F}$ is p, iff for any $u \in \mathcal{F}$ the* *asymptotic bound*

$$|u(\xi_0 + hx) - s^h(\xi_0 + hx)| = \mathcal{O}(h^p), \quad h \to 0,$$

holds for any $x \in \mathbb{R}^d$*, and any finite* $\mathcal{P}_k^d$*-unisolvent point set* $\Xi \subset \mathbb{R}^d$.

For the sake of notational simplicity, we let from now $\xi_0 = 0$, which is, due to the shift-invariance of the interpolation scheme, without loss of generality.

Note that the coefficients $c^h = (c_1^h, \ldots, c_n^h)^T \in \mathbb{R}^n$, $d^h = (d_\alpha^h)_{|\alpha|<k} \in \mathbb{R}^q$ of (28) are solving the linear system

$$\begin{bmatrix} \Phi_h & \Pi_h \\ \Pi_h^T & 0 \end{bmatrix} \cdot \begin{bmatrix} c^h \\ d^h \end{bmatrix} = \begin{bmatrix} u\big|_{h\Xi} \\ 0 \end{bmatrix}, \tag{29}$$

where we let

$$\begin{aligned} \Phi_h &= \big(\phi_{d,k}(\|h\xi_i - h\xi_j\|\big)_{1\le i,j\le n} \in \mathbb{R}^{n\times n}, \\ \Pi_h &= \big((h\xi_i)^\alpha\big)_{1\le i\le n; |\alpha|<k} \in \mathbb{R}^{n\times q}, \\ u\big|_{h\Xi} &= (u(h\xi_i))_{1\le i\le n} \in \mathbb{R}^n. \end{aligned}$$

We abbreviate the above linear system (29) as

$$A_h \cdot b^h = u_h, \tag{30}$$

i.e., for notational brevity, we let

$$A_h = \begin{bmatrix} \Phi_h & \Pi_h \\ \Pi_h^T & 0 \end{bmatrix}, \quad b^h = \begin{bmatrix} c^h \\ d^h \end{bmatrix}, \quad \text{and} \quad u_h = \begin{bmatrix} u\big|_{h\Xi} \\ 0 \end{bmatrix}.$$

Recall from the discussion in Subsection 2.6 that any interpolant s^h satisfying (27) has a Lagrange-type representation of the form

$$s^h(hx) = \sum_{i=1}^n \lambda_i^h(hx)u(h\xi_i), \tag{31}$$

corresponding to the one in (20), where moreover

$$\sum_{i=1}^n \lambda_i^h(hx)p(h\xi_i) = p(hx), \quad \text{for all } p \in \mathcal{P}_k^d, \tag{32}$$

due to the reconstruction of polynomials in $\mathcal{P}_k^d$.

Moreover, for $x \in \mathbb{R}^d$, the vector $\lambda^h(hx) = (\lambda_1^h(hx), \ldots, \lambda_n^h(hx))^T \in \mathbb{R}^n$ is, together with $\mu^h(hx) = (\mu_\alpha^h(hx))_{|\alpha|<k} \in \mathbb{R}^q$, the unique solution of the linear system

$$\begin{bmatrix} \Phi_h & \Pi_h \\ \Pi_h^T & 0 \end{bmatrix} \cdot \begin{bmatrix} \lambda^h(hx) \\ \mu^h(hx) \end{bmatrix} = \begin{bmatrix} \varphi_h(hx) \\ \pi_h(hx) \end{bmatrix}, \tag{33}$$

where

$$\begin{aligned} \Phi_h &= \left(\phi_{d,k}(\|h\xi_i - h\xi_j\|)\right)_{1\le i,j\le n} \in \mathbb{R}^{n\times n}, \\ \Pi_h &= \left((h\xi_i)^\alpha\right)_{1\le i\le n; |\alpha|<k} \in \mathbb{R}^{n\times q}, \\ \varphi_h(hx) &= \left(\phi_{d,k}(\|hx - h\xi_j\|)\right)_{1\le j\le n} \in \mathbb{R}^n, \\ \pi_h(hx) &= \left((hx)^\alpha\right)_{|\alpha|<k} \in \mathbb{R}^q. \end{aligned}$$

It is convenient to abbreviate the system (33) as $A_h \cdot \nu^h(hx) = \beta_h(hx)$, i.e., we let

$$A_h = \begin{bmatrix} \Phi_h & \Pi_h \\ \Pi_h^T & 0 \end{bmatrix}, \quad \nu^h(hx) = \begin{bmatrix} \lambda^h(hx) \\ \mu^h(hx) \end{bmatrix}, \quad \beta_h(hx) = \begin{bmatrix} \varphi_h(hx) \\ \pi_h(hx) \end{bmatrix}.$$

Starting with the Lagrange representation of s^h in (31), we obtain

$$\begin{aligned} s^h(hx) &= \, < \lambda^h(hx), u\big|_{h\Xi} > \\ &= \, < \nu^h(hx), u_h > \\ &= \, < A_h^{-1} \cdot \beta_h(hx), u_h > \\ &= \, < \beta_h(hx), A_h^{-1} \cdot u_h > \\ &= \, < \beta_h(hx), b_h >, \end{aligned} \tag{34}$$

see the identity (23). This in particular combines the two alternative representations for s^h in (31) and (28).

The following lemma, proven in [25], plays a key role in the following discussion. It states that the Lagrange basis of the polyharmonic spline interpolation scheme is invariant under uniform scalings. As established in the recap of the proof from [25] below, this result mainly relies on the (generalized) homogeneity of $\phi_{d,k}$.

LEMMA 1. *For any $h > 0$, let $\lambda^h(hx)$ be the solution in (33). Then,*

$$\lambda^h(hx) = \lambda^1(x), \qquad \textit{for every } x \in \mathbb{R}^d.$$

Proof. For fixed $\Xi = \{\xi_1, \ldots, \xi_n\} \subset \mathbb{R}^d$, and any $h > 0$, let

$$\mathcal{R}^h_{\phi,\Xi} = \left\{ \sum_{i=1}^n c_i \phi_{d,k}(\| \cdot - h\xi_i\|) + p : p \in \mathcal{P}_k^d, \sum_{i=1}^n c_i q(\xi_i) = 0 \text{ for all } q \in \mathcal{P}_k^d \right\}$$

denote the space of all possible polyharmonic spline interpolants of the form (28) satisfying (27). In what follows, we show that $\mathcal{R}^h_{\phi,\Xi}$ is a scaled version of $\mathcal{R}^1_{\phi,\Xi}$, so that $\mathcal{R}^h_{\phi,\Xi} = \left\{\sigma_h(s) : s \in \mathcal{R}^1_{\phi,\Xi}\right\}$, where the dilatation operator σ_h is given by $\sigma_h(s) = s(\cdot/h)$. This then implies that, due to the unicity of the interpolation in either space, $\mathcal{R}^h_{\phi,\Xi}$ or $\mathcal{R}^1_{\phi,\Xi}$, their Lagrange basis functions must coincide by satisfying $\lambda^h = \sigma_h(\lambda^1)$, as stated above.

In order to show that $\mathcal{R}^h_{\phi,\Xi} = \sigma_h(\mathcal{R}^1_{\phi,\Xi})$, we distinguish the special case where d is even from the one where d is odd. If the space dimension d is odd, then $\mathcal{R}^h_{\phi,\Xi} = \sigma_h(\mathcal{R}^1_{\phi,\Xi})$ follows immediately from the homogeneity of $\phi_{d,k}$, where $\phi_{d,k}(hr) = h^{2k-d}\phi_{d,k}(r)$.

Now suppose that d is even. In this case we have

$$\phi_{d,k}(hr) = h^{2k-d}\left(\phi_{d,k}(r) + r^{2k-d}\log(h)\right).$$

Therefore, any function $s^h \in \mathcal{R}^h_{\phi,\Xi}$ has, for some $p \in \mathcal{P}^d_k$, the form

$$s^h(hx) = h^{2k-d}\left(\sum_{i=1}^n c_i^h \phi_{d,k}(\|x-\xi_i\|) + \log(h)q(x)\right) + p(x),$$

where we let

$$q(x) = \sum_{i=1}^n c_i^h \|x-\xi_i\|^{2k-d}.$$

In order to see that s^h is contained in $\sigma_h(\mathcal{R}^1_{\phi,\Xi})$, it remains to show that the degree of the polynomial q is at most $k-1$. To this end, we rewrite q as

$$q(x) = \sum_{i=1}^n c_i^h \sum_{|\alpha|+|\beta|=2k-d} c_{\alpha,\beta} \cdot x^\alpha (\xi_i)^\beta = \sum_{|\alpha|+|\beta|=2k-d} c_{\alpha,\beta} \cdot x^\alpha \sum_{i=1}^n c_i^h (\xi_i)^\beta,$$

for some coefficients $c_{\alpha,\beta} \in \mathbb{R}$ with $|\alpha| + |\beta| = 2k-d$. Due to the vanishing moment conditions

$$\sum_{i=1}^n c_i^h p(h\xi_i) = 0, \qquad \text{for all } p \in \mathcal{P}^d_k,$$

for the coefficients $c_1^h, \ldots, c_n^h$, this implies that the degree of q is at most $2k-d-k = k-d < k$. Therefore, $s^h \in \sigma_h(\mathcal{R}^1_{\phi,\Xi})$, and so $\mathcal{R}^h_{\phi,\Xi} \subset \sigma_h(\mathcal{R}^1_{\phi,\Xi})$. The inclusion $\mathcal{R}^1_{\phi,\Xi} \subset \sigma_h^{-1}(\mathcal{R}^h_{\phi,\Xi})$ can be proven accordingly.

Altogether, we find that $\mathcal{R}^h_{\phi,\Xi} = \sigma_h(\mathcal{R}^1_{\phi,\Xi})$ for any d, which completes our proof. □

Now let us draw important conclusions on the approximation order of local polyharmonic spline interpolation with respect to C^k. To this end, regard for $u \in C^k$, any

$x \in \mathbb{R}^d$ and $h > 0$, the k-th order Taylor polynomial

$$p^h(y) = \sum_{|\alpha|<k} \frac{1}{\alpha!} D^\alpha u(hx)(y - hx)^\alpha. \tag{35}$$

By using

$$u(hx) = p^h(h\xi_i) - \sum_{0<|\alpha|<k} \frac{1}{\alpha!} D^\alpha u(hx)(h\xi_i - hx)^\alpha, \quad \text{for all } 1 \le i \le n,$$

in combination with (31) and (32), we obtain the identity

$$u(hx) - s^h(hx) = \sum_{i=1}^n \lambda_i^h(hx) \left[p^h(h\xi_i) - u(h\xi_i) \right].$$

Now due to Lemma 1, the *Lebesgue constant*

$$\Lambda = \sup_{h>0} \sum_{i=1}^n \left| \lambda_i^h(hx) \right| = \sum_{i=1}^n \left| \lambda_i^1(x) \right|$$

is bounded, locally around the origin $\xi_0 = 0$, and therefore we can conclude

$$|u(hx) - s^h(hx)| = \mathcal{O}(h^k), \quad h \to 0.$$

Altogether, this yields the following result.

THEOREM 6. *The approximation order of local polyharmonic spline interpolation, using $\phi_{d,k}$, with respect to C^k is $p = k$.*

We remark that the above Theorem 6 generalizes a previous result in [21] concerning the local approximation order of thin plate spline interpolation in the plane.

COROLLARY 1. *The approximation order of local thin plate spline interpolation, using $\phi_{2,2} = r^2 \log(r)$, with respect to C^2 is $p = 2$.*

3.3. Numerical stability

This section is devoted to the construction of a numerically stable algorithm for the evaluation of polyharmonic spline interpolants. Recall that the stability of an algorithm always depends on the conditioning of the given problem. For a more general discussion on the relevant principles and concepts from error analysis, especially the *condition number* of a given *problem* versus the *stability* of a *numerical algorithm*, we recommend the textbook [22].

In order to briefly explain the conditioning of polyharmonic spline interpolation, let $\Omega \subset \mathbb{R}^d$ denote a compact domain comprising $\Xi = \{\xi_1, \ldots, \xi_n\}$, i.e., $\Xi \subset \Omega$, the

$\mathcal{P}_k^d$-unisolvent set of interpolation points. Now recall that the condition number of an interpolation operator $\mathcal{I} : C(\Omega) \to C(\Omega)$, $\Omega \subset \mathbb{R}^d$, w.r.t. the L_∞-norm $\|\cdot\|_{L_\infty(\Omega)}$, is the smallest number κ_∞ satisfying

$$\|\mathcal{I}u\|_{L_\infty(\Omega)} \le \kappa_\infty \cdot \|u\|_{L_\infty(\Omega)} \qquad \text{for all } u \in C(\Omega).$$

Thus, κ_∞ is the operator norm of $\mathcal{I}$ w.r.t. the norm $\|\cdot\|_{L_\infty(\Omega)}$. In the situation of polyharmonic spline interpolation, the interpolation operator $\mathcal{I}_{d,k} : C(\Omega) \to C(\Omega)$, returns, for any given argument $u \in C(\Omega)$ the polyharmonic spline interpolant $\mathcal{I}_{d,k}(u) = s_u \in C(\Omega)$ of the form (24) satisfying $s_u\big|_\Xi = u\big|_\Xi$. The following result is useful for the subsequent discussion on the stability of local interpolation by polyharmonic splines.

THEOREM 7. *The condition number κ_∞ of interpolation by polyharmonic splines is given by the Lebesgue constant*

$$\Lambda(\Omega, \Xi) = \max_{x\in\Omega} \sum_{i=1}^{n} |\lambda_i(x)|. \tag{36}$$

Proof. For $u \in C(\Omega)$, let $u\big|_\Xi$ be given, and let $s_u = \mathcal{I}_{d,k}(u) \in C(\Omega)$ denote the interpolant of the form (24) satisfying $u\big|_\Xi = s_u\big|_\Xi$. Using the Lagrange-type representation

$$s_u(x) = \sum_{i=1}^{n} \lambda_i(x) u(\xi_i)$$

of s_u, we obtain

$$\|\mathcal{I}_{d,k}u\|_{L_\infty(\Omega)} = \|s_u\|_{L_\infty(\Omega)} \le \max_{x\in\Omega} \sum_{i=1}^{n} |\lambda_i(x)| \cdot |u(\xi_i)| \le \Lambda(\Omega, \Xi) \cdot \|u\|_{L_\infty(\Omega)}$$

for all $u \in C(\Omega)$, and therefore $\kappa_\infty \le \Lambda(\Omega, \Xi)$.

In order to see that $\kappa_\infty \ge \Lambda(\Omega, \Xi)$, suppose that the maximum of $\Lambda(\Omega, \Xi)$ in (36) is attained at $x^* \in \Omega$. Moreover, let $g \in C(\Omega)$ denote any function satisfying $g(\xi_i) = \text{sign}(\lambda_i(x^*))$, for all $1 \le i \le n$, and $\|g\|_{L_\infty(\Omega)} = 1$. Then, we obtain

$$\|\mathcal{I}_{d,k}g\|_{L_\infty(\Omega)} \ge \left(\mathcal{I}_{d,k}g\right)(x^*) = \sum_{i=1}^{n} \lambda_i(x^*) g(\xi_i) = \sum_{i=1}^{n} |\lambda_i(x^*)| = \Lambda(\Omega, \Xi)$$

and thus $\|\mathcal{I}_{d,k}g\|_{L_\infty(\Omega)} \ge \Lambda(\Omega, \Xi)\|g\|_{L_\infty(\Omega)}$. But this implies $\Lambda(\Omega, \Xi) \le \kappa_\infty$. Altogether, $\kappa_\infty = \Lambda(\Omega, \Xi)$, which completes our proof.

□

The above Lemma 1 immediately yields the following important result concerning the stability of interpolation by polyharmonic splines.

THEOREM 8. *The absolute condition number of polyharmonic spline interpolation is invariant under rotations, translations and uniform scalings.*

Proof. Interpolation by polyharmonic splines is invariant under rotations and translations. It is easy to see that this property carries over to the absolute condition number. In order to see that $\kappa_\infty \equiv \kappa_\infty(\Omega, \Xi)$ is also invariant under uniform scalings, let $\Omega^h = \{hx : x \in \Omega\}$ and $\Xi^h = \{h\xi : \xi \in \Xi\}$. Then, we obtain

$$\Lambda(\Omega^h, \Xi^h) = \max_{hx \in \Omega^h} \sum_{i=1}^{n} \lambda_i^h(hx) = \max_{x \in \Omega} \sum_{i=1}^{n} \lambda_i(x) = \Lambda(\Omega, \Xi)$$

which shows that $\kappa_\infty(\Omega^h, \Xi^h) = \kappa_\infty(\Omega, \Xi)$.

□

Now let us turn to the construction of a numerically stable algorithm for evaluating the polyharmonic spline interpolant s^h satisfying (27). To this end, we require that the given interpolation problem (27) is well-conditioned. Note that according to Theorem 8, this requirement depends on the geometry of the interpolation points Ξ w.r.t. the center ξ_0, but not on the scale h.

However, the spectral condition number of the matrix A_h depends on h. The following rescaling can be viewed as a simple way of preconditioning the matrix A_h for very small h. To this end, in order to evaluate the polyharmonic spline interpolant s^h satisfying (27), we prefer to work with the representation

$$s^h(hx) = < \beta_1(x), A_1^{-1} \cdot u_h >, \tag{37}$$

which immediately follows from the identity (34) and the scale-invariance of the Lagrange basis, Lemma 1. Due to (37) we can evaluate s^h at hx by solving the linear system

$$A_1 \cdot \mathbf{b} = u_h. \tag{38}$$

The solution $\mathbf{b} \in \mathbb{R}^{n+q}$ in (38) then yields the coefficients of $s^h(hx)$ w.r.t. the basis functions in $\beta_1(x)$.

By working with the representation (37) for s^h instead of the one in (28), we can avoid solving the linear system (30). This is useful insofar as the linear system (30) is ill-conditioned for very small h, but well-conditioned for sufficiently large h. The latter relies on earlier results due to Narcowich and Ward [37], where it is shown that the spectral norm of the matrix Φ_h^{-1} is bounded above by a monotonically decreasing function of the minimal Euclidean distance between the points in $h\Xi$. This in turns implies that one should, for the sake of numerical stability, avoid solving the system (30) directly for very small h. For further details on this, see [37] and the more general discussion provided by the recent paper [43] of Schaback.

4. Meshfree methods for transport problems

4.1. Transport equations

Numerical methods in flow simulation are concerned with time-dependent *hyperbolic conservation laws* of the form

$$\frac{\partial u}{\partial t} + \nabla f(u) = 0, \tag{39}$$

where for some domain $\Omega \subset \mathbb{R}^d$, $d \geq 1$, and a compact time interval $I = [0, T]$, $T > 0$, the solution $u : I \times \Omega \to \mathbb{R}$ of (39) is sought.

Moreover, $f(u) = (f_1(u), \dots, f_d(u))^T$ denotes a given *flux tensor*, and we assume that *initial conditions*

$$u(0, x) = u_0(x), \qquad \text{for } x \in \Omega, \tag{40}$$

at time $t = 0$ are given.

One special case for (39) is *passive advection*, where the flux f is linear, i.e.,

$$f(u) = v \cdot u,$$

and thus (39) becomes

$$\frac{\partial u}{\partial t} + v \cdot \nabla u = 0, \tag{41}$$

provided that the given *velocity field*

$$v = v(t, x) \in \mathbb{R}^d, \qquad t \in I,\ x \in \Omega,$$

is *divergence-free*, i.e.,

$$\operatorname{div} v = \sum_{j=1}^{d} \frac{\partial v_j}{\partial x_j} \equiv 0.$$

For a comprehensive introduction to hyperbolic conservation laws, we recommend the textbook [28].

4.2. Semi-lagrangian advection

For the special case of passive advection, the resulting Cauchy problem (41), (40) is well-posed. In this case, the solution u is constant along the *streamlines* of fluid particles, and the shapes of these streamlines are entirely determined by the given velocity field v.

This suggests to work with a *semi-Lagrangian method* (SLM) in order to solve the Cauchy problem for passive advection. Loosely speaking, a SLM is one which follows the flow of a discrete set of particles along their streamline trajectories, and moreover the particle set is subject to dynamic changes during the simulation. Therefore, any SLM may be regarded as a special instance of the classical *method of characteristics* (MOC). Indeed, this is because the streamlines of the flow particles are the *characteristic curves* of the equation (41) [28].

In order to be more precise about the SLM, let $\Xi \subset \Omega$ denote a current finite set of nodes, at time $t \in I$, each of whose elements $\xi \in \Xi$ corresponds to a fluid particle located at ξ. Now for a fixed time step size $\tau > 0$, the advection in the SLM at time step $t \to t + \tau$ is accomplished as follows. For any node $\xi \in \Xi$, an approximation to its *upstream point* $x^- \equiv x^-(\xi)$ is computed. The upstream point x^- of ξ is the spatial location of that particle at time t, which by traversing along its corresponding streamline arrives at the node ξ at time $t + \tau$. Figure 1 shows the corresponding upstream point of a node ξ, along with its streamline trajectory.

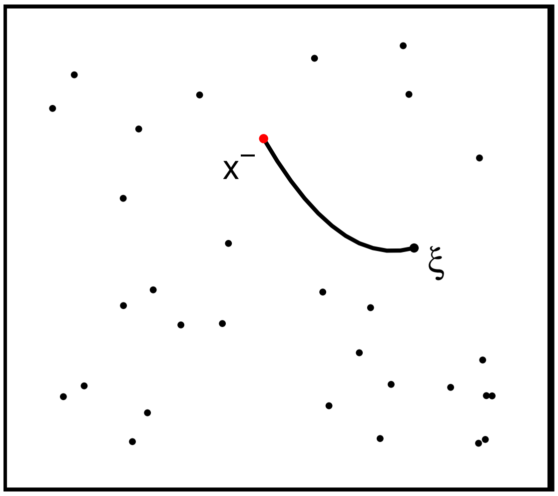


Figure 1: The point x^- is the upstream point of the node ξ.

We remark that computing the upstream point x^- of any node $\xi \in \Xi$ amounts to solving the *ordinary differential equation* (ODE)

$$\dot{x} = \frac{dx}{dt} = v(t, x) \tag{42}$$

with initial condition $x(t + \tau) = \xi$, and so $x(t) = x^-$.

Adopting some standard notation from dynamic systems, we can express the upstream point x^- of ξ as

$$x^- = \Phi^{t,t+\tau}\xi, \tag{43}$$

where $\Phi^{t,t+\tau} : \Omega \to \Omega$ denotes the *continuous evolution* of the (backward) flow of (42). An equivalent formulation for (43) is given by $\xi = \Phi^{t+\tau,t}x^-$, since $\Phi^{t+\tau,t}$ is the inverse of $\Phi^{t,t+\tau}$.

Now since the solution u of (41) is constant along the trajectories of the flow particles, we have $u(t, x^-) = u(t + \tau, \xi)$, and so the desired values $\{u(t + \tau, \xi)\}$, $\xi \in \Xi$, may immediately be obtained from the upstream point values $u(t, x^-)$. But in general, neither the *exact* location of x^-, nor the value $u(t, x^-)$ is known.

Therefore, during the performance of the flow simulation, this requires first computing an approximation $\tilde{x}$ of the upstream point $x^- = \Phi^{t,t+\tau}\xi$ for each $\xi \in \Xi$. It is convenient to express the approximation $\tilde{x}$ of x^- as

$$\tilde{x} = \Psi^{t,t+\tau}\xi,$$

where $\Psi^{t,t+\tau} : \Omega \to \Omega$ is the *discrete evolution* of the flow, corresponding to the continuous evolution $\Phi^{t,t+\tau}$ in (43) [7]. The operator $\Psi^{t,t+\tau}$ is given by any specific numerical method for solving the above ODE (42).

Having computed $\tilde{x}$, the value $u(t, \tilde{x})$ is then determined from the current values $\{u(t, \xi)\}_{\xi \in \Xi}$ by local interpolation. Altogether, the above discussion leads us to the following algorithm concerning the advection step $t \to t + \tau$ of the semi-Lagrangian method.

ALGORITHM 1. **(Semi-lagrangian advection).**

INPUT: Time step size $\tau > 0$, node set $\Xi \subset \Omega$, and values $\{u(t, \xi)\}_{\xi \in \Xi}$.

FOR each $\xi \in \Xi$ **DO**

(a) Compute the upstream point approximation $\tilde{x} = \Psi^{t,t+\tau}\xi$;

(b) Determine the value $u(t, \tilde{x})$ by local interpolation;

(c) Advect by letting $u(t + \tau, \xi) = u(t, \tilde{x})$.

OUTPUT: The values $u(t + \tau, \xi)$, for all $\xi \in \Xi$, at time $t + \tau$.

The local interpolation in step **(b)** of the above algorithm needs some comments. First note that $\tilde{x}$, the approximation of the upstream point of ξ, is not necessarily contained in the node set Ξ. Therefore, the desired value $u(t, \tilde{x})$ is to be computed from the given values $\{u(t, \xi)\}_{\xi \in \Xi}$ of u at the nodes in Ξ. This is done by local interpolation. To this end, a set $\mathcal{N} \equiv \mathcal{N}(\tilde{x}) \subset \Xi$ of neighbouring nodes of $\tilde{x}$ is determined. In order to make one concrete example, $\mathcal{N}$ could, for some suitable number n, be the set of n nearest neighbours of $\tilde{x}$ in Ξ. The given function values of $u(t, \cdot)$ at the neighbouring nodes are then used in order to solve the interpolation problem

$$u(t, \nu) = s(\nu), \qquad \text{for all } \nu \in \mathcal{N}, \tag{44}$$

by a *suitable* scattered data interpolation scheme, which outputs an interpolant $s : \Omega \to \mathbb{R}$ satisfying (44). For this purpose, we prefer to work with polyharmonic spline interpolation, so that s in (44) is required to have the form (24). The desired approximation of $u(t, \tilde{x})$ is obtained by the evaluation of s at $\tilde{x}$, so we let $u(t, \tilde{x}) = s(\tilde{x})$.

We remark that semi-Lagrangian advection schemes of the above form are unconditionally stable. This is in contrast to Eulerian schemes, which, for the sake of stability, typically work with very small time steps [28]. For a concise analysis concerning the convergence and stability of semi-Lagrangian methods, we refer to the paper [14] by Falcone and Ferretti. A more general discussion on semi-Lagrangian methods is provided in the textbooks [11, 36]; for applications of the SLM in atmospheric problems, see the review [49] by Staniforth and Côté and the seminal papers [40, 41] of Robert.

4.3. Method of backward characteristics

Now let us return to the general case of (39) where the flux function f is, unlike in (41), *nonlinear*. We remark that nonlinear cases are much more complicated than the linear one of passive advection. Therefore, the construction of a generalization to the above semi-Lagrangian method in Algorithm 1 requires particular care. Indeed, in contrast to the linear case, a nonlinear flux function f usually leads to *discontinuities* in the solution u, *shocks*, as observed in many relevant applications, such as fluid flow and gas dynamics. In such situations, the classical method of characteristics becomes unwieldy or impossible, as the evolution of the flow along the characteristic curves is typically very complicated, or characteristic curves may even be undefined (see [11, Subsection 6.3.1] for a discussion on these phenomena).

Now in order to be able to model the behaviour of the solution with respect to shock formation and shock propagation we work with a *vanishing viscosity* approach, yielding the modified *advection-diffusion equation*

$$\frac{\partial u}{\partial t} + \nabla f(u) = \epsilon \cdot \Delta u, \tag{45}$$

where the parameter $\epsilon > 0$ is referred to as the *diffusion coefficient*. In this way, the solution u of the *hyperbolic* equation (39) is approximated arbitrarily well by the solution of the modified *parabolic* equation (45), provided that the parameter ϵ is sufficiently small. This modification is a standard stabilization technique for nonlinear equations, dating back to Burgers [6], who utilized a flux function of the form

$$f(u) = \frac{1}{2}u^2 \cdot r, \tag{46}$$

with some flow direction $r \in \mathbb{R}^d$, for modelling free turbulences in fluid dynamics. The resulting *Burgers equation* is nowadays a popular standard test case for nonlinear transport equations. We come back to this test case in Subsection 5.2.

Now let us propose a meshfree advection scheme for solving the above nonlinear equation (45). Starting point for this modified approach is the discretization

$$\frac{u(t+\tau,\xi) - u(t,x^-)}{\tau} = \epsilon \cdot \Delta u(t,x^-) \tag{47}$$

of the Lagrangian form

$$\frac{du}{dt} = \epsilon \cdot \Delta u,$$

of (45), where

$$\frac{du}{dt} = \frac{\partial u}{\partial t} + \nabla f(u)$$

is the *material derivative*.

Note that the discretization in (47) allows us to work with a similar advection scheme as in the linear case, given by Algorithm 1. Indeed, having computed for any $\xi \in \Xi$ an approximation $\tilde{x} = \Psi^{t,t+\tau}\xi$ to its upstream point $x^- = \Phi^{t,t+\tau}\xi$, the desired approximation of $u(t+\tau,\xi)$ is then given by

$$u(t+\tau,\xi) = u(t,\tilde{x}) + \tau \cdot \epsilon \cdot \Delta u(t,\tilde{x}), \qquad \text{for } \xi \in \Xi.$$

However, in contrast to plain passive advection, the characteristic curves of the equation (45) depend also on u. In particular, the advection velocity $v = \frac{\partial f(u)}{\partial u}$ depends on u. This amounts to applying a more sophisticated integration scheme (compared with the one of the previous subsection) in order to compute for any $\xi \in \Xi$ its corresponding upstream point approximation $\tilde{x} = \Psi^{t,t+\tau}\xi$. For the sake of brevity, we prefer to omit these lengthy technical details, which are immaterial for the purposes of this chapter. Instead, we refer to the discussion in [3].

The following algorithm reflects the advection step $t \to t+\tau$ of the suggested method of (backward) characteristics.

ALGORITHM 2. **(Method of characteristics).**

INPUT: Time step τ, nodes Ξ, values $\{u(t,\xi)\}_{\xi\in\Xi}$, diffusion coefficient ϵ.

FOR each $\xi \in \Xi$ **DO**

- **(a)** Compute the upstream point approximation $\tilde{x} = \Psi^{t,t+\tau}\xi$;
- **(b)** Determine the values $u(t,\tilde{x})$ and $\Delta u(t,\tilde{x})$ by local interpolation;
- **(c)** Advect by letting $u(t+\tau,\xi) = u(t,\tilde{x}) + \tau \cdot \epsilon \cdot \Delta u(t,\tilde{x})$.

OUTPUT: The values $u(t+\tau,\xi)$, for all $\xi \in \Xi$, at time $t+\tau$.

Step **(b)** of Algorithm 2 deserves a comment concerning the interpolation of the value $\Delta u(t,\tilde{x})$. Similar as in Algorithm 1 we work with local interpolation by polyharmonic splines, but with a *smoother* basis function, such that the Laplacian Δs of the interpolant s satisfying (44) is everywhere well-defined. The desired approximation of $\Delta u(t,\tilde{x})$ is then obtained by $\Delta s(t,\tilde{x})$.

4.4. Adaption rules

In this section, the adaptive modification of the node set Ξ is explained. This is done after each time step $t \to t+\tau$ of the semi-Lagrangian method (Algorithm 1) in case of passive advection (41), or of the method of characteristics (Algorithm 2), when solving nonlinear advection-diffusion equations of the form (45). In either case, the *current* node values $u(t, \xi)$, $\xi \in \Xi$, are used in order to adaptively modify Ξ. Immediately before the first time step $0 \to \tau$, the nodes are first randomly chosen in Ω, before the adaption rules, to be explained below, are applied. This then yields the initial node set $\Xi \equiv \Xi(0)$.

The modification of the current node set $\Xi \equiv \Xi(t)$ (at time t) is accomplished by the removal (coarsening), and the insertion (refinement) of selected nodes, so that a modified node set $\Xi \equiv \Xi(t+\tau)$ (at time $t+\tau$) is obtained. The adaptive modification of the nodes in Ξ relies on a customized a posteriori error indicator, to be explained in the following subsection.

Error indication.

An effective strategy for the adaptive modification of the nodes requires well-motivated refinement and coarsening rules as well as a customized error indicator. We understand the error indicator $\eta : \Xi \to [0, \infty)$ as a function of the current node set $\Xi \equiv \Xi(t)$ (at time t) which assigns a *significance* value $\eta(\xi)$ to each node $\xi \in \Xi$. The value $\eta(\xi)$ is required to reflect the local approximation quality of the interpolation around $\xi \in \Xi$. The significances $\eta(\xi)$, $\xi \in \Xi$, are then used in order to flag single nodes $\xi \in \Xi$ as "to be refined" or "to be coarsened" according to the following criteria.

DEFINITION 5. *Let* $\eta^* = \max_{\xi \in \Xi} \eta(\xi)$, *and let* $\theta_{\text{crs}}, \theta_{\text{ref}}$ *be two tolerance values satisfying* $0 < \theta_{\text{crs}} < \theta_{\text{ref}} < 1$. *We say that a node* $\xi \in \Xi$ *is* to be refined, *iff* $\eta(\xi) > \theta_{\text{ref}} \cdot \eta^*$, *and* ξ *is* to be coarsened, *iff* $\eta(\xi) < \theta_{\text{crs}} \cdot \eta^*$.

In our numerical examples, typical choices for the relative tolerance values are $\theta_{\text{crs}} = 0.001$ and $\theta_{\text{ref}} = 0.1$. Note that a node ξ cannot be refined and be coarsened at the same time; in fact, it may neither be refined nor be coarsened.

Now let us turn to the definition of the error indicator η. To this end, we follow along the lines of [21], where a scheme for the detection of discontinuities of a surface, *fault lines*, from scattered data was developed. We let

$$\eta(\xi) = |u(\xi) - s(\xi)|,$$

where $s \equiv s_{\mathcal{N}}$ denotes the polyharmonic spline interpolant, which matches the values of $u \equiv u(t, \cdot)$ at a neighbouring set $\mathcal{N} \equiv \mathcal{N}(\xi) \subset \Xi \setminus \xi$ of current nodes, i.e., $s(\nu) = u(\nu)$ for all $\nu \in \mathcal{N}$. In our numerical examples for bivariate data, where $d = 2$, we work with local thin plate spline interpolation. Recall that this particular interpolation scheme reconstructs linear polynomials. In this case, the value $\eta(\xi)$ vanishes whenever u is linear around ξ. Moreover, the indicator $\eta(\xi)$ is small whenever the *local* reproduction quality of the interpolant s is good. In contrast to this, a high value of $\eta(\xi)$ typically indicates that u is subject to strong variation locally around ξ.

Coarsening and refinement

In order to obtain good approximation quality at small computational costs, we insert new nodes into regions where the value of η is high (refinement), whereas nodes from

Ξ are removed in regions where the value of η is small (coarsening).

To avoid additional computational overhead and complicated data structures, effective adaption rules are required to be as simple as possible. In particular, these rules ought to be given by *local* operations on the current node set Ξ. The following coarsening rule is in fact very easy and, in combination with the refinement, it turned out to be very effective as well.

Coarsening. A node $\xi \in \Xi$ is *coarsened* by its removal from the current node set Ξ, i.e., Ξ is modified by replacing Ξ with $\Xi \setminus \xi$.

As to the refinement rules, these are constructed on the basis of the local error estimate (17) for polyharmonic spline interpolation. The refinement of any node $\xi \in \Xi$ should aim at the reduction of the local error (17) around ξ. We accomplish this by reducing the distance function

$$d_{\mathcal{N}} = \min_{\nu \in \mathcal{N}} \| \cdot - \nu \|$$

in a local neighbourhood of ξ. In order to explain this, we need to recall some ingredients from computational geometry, in particular Voronoi diagrams [39].

For any node $\xi \in \Xi$, its corresponding *Voronoi tile*

$$V_\Xi(\xi) = \left\{ y \in \mathbb{R}^d \, : \, d_\Xi(y) = \|y - \xi\| \right\} \subset \mathbb{R}^d$$

w.r.t. the point set Ξ is a convex polyhedron containing all points in $\mathbb{R}^d$ which are at least as close to ξ as to any other point in Ξ. The boundary vertices of $V_\Xi(\xi)$, called *Voronoi points*, form a finite point set $\mathcal{V}_\xi$ in the neighbourhood of ξ. Figure 2 shows the Voronoi tile $V_\Xi(\xi)$ of a node ξ along with the set $\mathcal{V}_\xi$ of its Voronoi points.

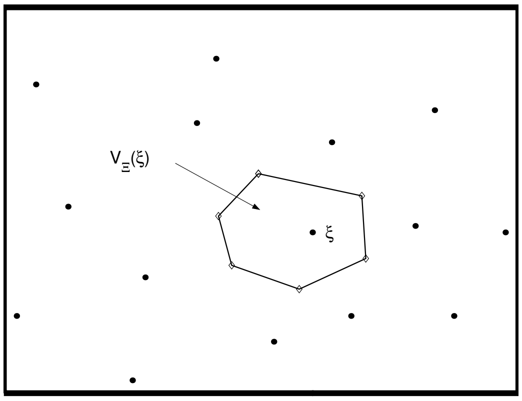


Figure 2: Refinement of the node ξ. The Voronoi points ($\diamond$) are inserted.

Now observe that for any $\xi \in \mathcal{N}$, the distance function $d_{\mathcal{N}}$ is convex on $V_\Xi(\xi)$. Moreover, the function $d_{\mathcal{N}}$ has local maxima at the Voronoi points in $\mathcal{V}_\xi$. Altogether, this gives rise to define the local refinement of nodes as follows.

Refinement. A node $\xi \in \Xi$ is *refined* by the insertion of its Voronoi points into the current node set Ξ, i.e., Ξ is modified by replacing Ξ with $\Xi \cup \mathcal{V}_\xi$.

5. Meshfree fluid flow simulation

In this section, the good performance of the proposed meshfree advection scheme is shown, where the utility of the adaptive sampling strategy is demonstrated. To this end, we work with the following two popular test case scenarios from flow simulation.

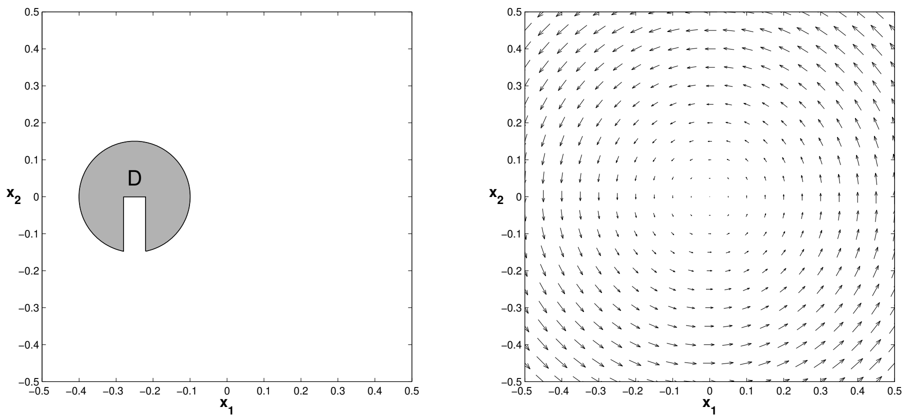
The *slotted cylinder*, subject of Subsection 5.1, is concerning passive advection. In this case, the semi-Lagrangian method (Algorithm 1) in combination with the adaption rules of Subsection 4.4 is used. The subsequent discussion in Subsection 5.2 is then devoted to the aforementioned nonlinear *Burgers equation*. In this test case, we work with the method of characteristics (Algorithm 2) in combination with the adaption rules of Subsection 4.4

5.1. The slotted cylinder: a test case for passive advection

The *slotted cylinder*, suggested by Zalesak [55], is a popular test case scenario for flow simulation concerning passive advection. In this test case, the domain $\Omega = [-0.5, 0.5] \times [-0.5, 0.5] \subset \mathbb{R}^2$ is the shifted unit square, and the initial conditions in (40) are given by

$$u_0(x) = \begin{cases} 1 & \text{for } x \in D, \\ 0 & \text{otherwise,} \end{cases}$$

where $D \subset \Omega$ is the slotted disc of radius $r = 0.15$ centered at $(-0.25, 0)$ with slot width 0.06 and length 0.22, see Figure 3 **(a)**.



(a) **(b)**

Figure 3: The slotted cylinder. **(a)** Initial condition and **(b)** velocity field.

The slotted cylinder is rotated counter-clockwise by the steady flow field $v(x) = (-x_2, x_1)$, see Figure 3 **(b)**. Throughout the simulation, we work with a constant time step size $\tau = 0.1$, so that one revolution of the slotted cylinder around the origin requires 63 time steps. In our numerical experiment, we have recorded ten revolutions of the slotted cylinder, i.e., we let $I = [0, 629\tau]$.

Figure 6 shows the 3D view on the solution u and the corresponding node distribution during the first revolution of the slotted cylinder, at times $t_{16} = 16\tau$, $t_{32} = 32\tau$, and $t_{47} = 47\tau$. Observe that the distribution of the nodes is well-adapted to the edges of the slotted cylinder. In fact, the dense sampling along the edges leads to a high resolution of the model near the discontinuities of the solution u. On the other hand, the sparsity of the sampling in *flat* regions, where u is constant, serves to reduce the data size of the model, and thus the required computational costs. In conclusion, due to the adaptive sampling strategy, the two (conflicting) requirements of good approximation quality and computational efficiency are well-balanced.

As to the long-term behaviour of the simulation, Figure 7 shows in comparison the 3D view and the node distribution for the initial condition u_0, along with the numerical solution u obtained after five (at time $t_{315} = 315\tau$) and ten (at time $t_{629} = 629\tau$) full revolutions of the slotted cylinder. Observe that the shape of the slotted cylinder is maintained remarkably well. Moreover, numerical diffusion is widely avoided. This robust behaviour is due to the adaptive node sampling, which continues to resolve the edges of the slotted cylinder very well.

5.2. Burgers equation: a nonlinear standard test

The equation

$$\frac{\partial u}{\partial t} + u\nabla u \cdot r = \epsilon \cdot \Delta u, \tag{48}$$

was introduced in 1940 by Burgers [6] as a mathematical model of free turbulence in fluid dynamics. Burgers equation (48) is nowadays a popular standard test case for the simulation of nonlinear flow processes, and for the modelling of *shock waves*.

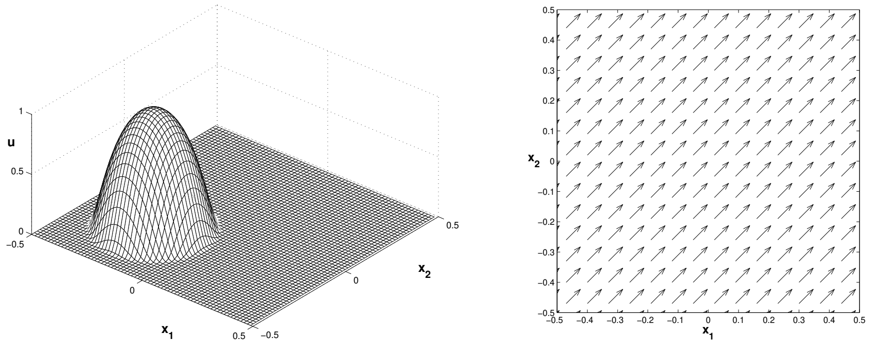
The nonlinear flux tensor (46) leads, in the hyperbolic equation (39), to *shocks*. As soon as the shock front occurs, there is no classical solution of the equation (39), and its weak solution becomes discontinuous. However, the modified parabolic equation (48) has for all $t > 0$ a *smooth* solution u_ϵ which approximates (for sufficiently small ϵ) the occuring shock front propagation arbitrarily well.

We use Burgers equation (48) in order to demonstrate the utility of adaptive sampling, in combination with the meshfree method of characteristics (Algorithm 2), for the modelling of shock fronts.

In the considered test case, we let

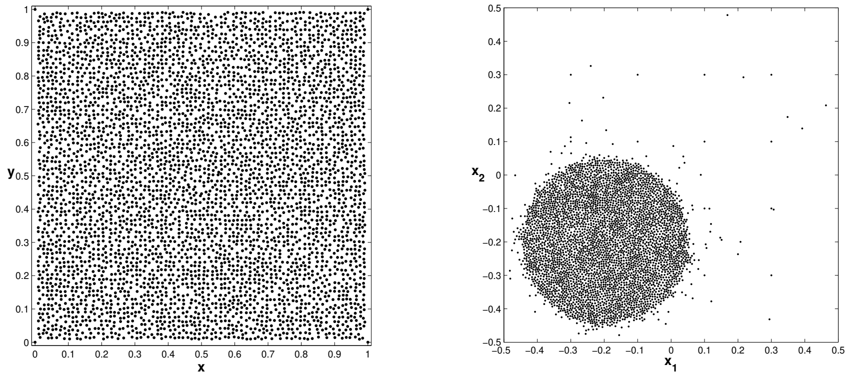
$$u_0(x) = \begin{cases} 0 & \text{for } \|x - c\| \geq R, \\ \exp\left(\frac{\|x-c\|^2}{\|x-c\|^2 - R^2}\right) & \text{otherwise,} \end{cases}$$

for the initial condition in (40), where $R = 0.25$, $c = (0.3, 0.3)$, and we let the unit square $\Omega = [0, 1]^2$ be the computational domain. Figure 4 shows the initial condition and the flow field $r = (1, 1)$, being aligned along the diagonal in Ω.



(a) **(b)**

Figure 4: Burgers equation. **(a)** Initial condition u_0 and **(b)** flow field.

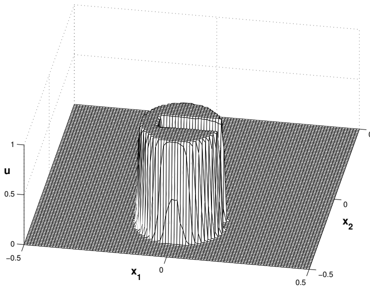


(a) **(b)**

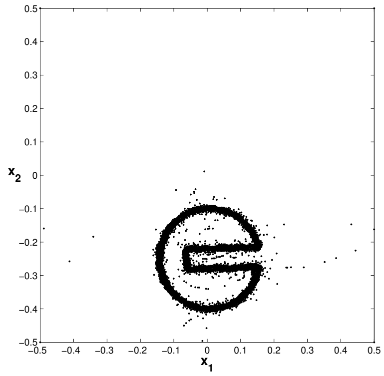
Figure 5: Burgers equation. **(a)** Randomly chosen node set of size $|\Xi| = 4446$ and **(b)** initial node distribution comprising $|\Xi| = 1484$ nodes.

The adaptive distribution of the initial node set is shown in Figure 5 **(b)**. Recall that the construction of this node set is done by applying the adaption rules of Subsection 4.4 on a randomly chosen node set in Ω. To this end, we first selected the node set Ξ displayed in Figure 5 **(a)**, of size $|\Xi| = 4446$, by random, before the significances $\eta(\xi)$ at the nodes in Ξ are used in order to compute the initial node set $\Xi \equiv \Xi(0)$ of the simulation, shown in Figure 5 **(b)**. Observe that the adaptive distribution of the nodes in Figure 5 **(b)** manages to localize the support of the initial condition u_0 very well.

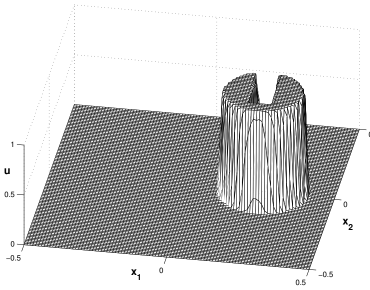
Moreover, in this simulation, a constant time step size $\tau = 0.004$ is selected, and we let $I = [0, 330\tau]$. A plot of the numerical solution u at the three time steps $t_{110} = 110\tau$, $t_{220} = 220\tau$, and $t_{330} = 330\tau$ is shown in Figure 8, along with the corresponding distribution of the nodes. Observe that the adaptive node distribution continues to localize the support of the solution u very well. This helps, on the one hand, to reduce the resulting computational costs of the simulation. On the other hand, the shock front propagation is well-resolved by the high density of the nodes around the shock, see Figure 8. Altogether, the adaptive node distribution manages to capture the evolution of the flow very effectively. This confirms the utility of the customized adaption rules yet once more.



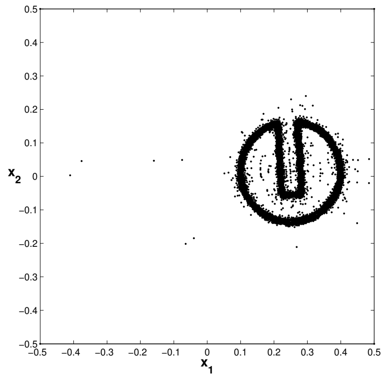
$t = t_{16}$



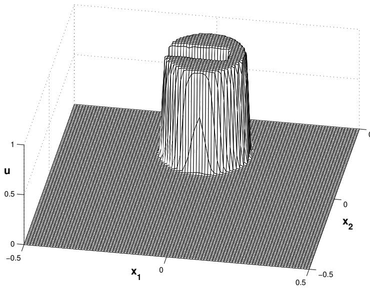
$t = t_{16}$



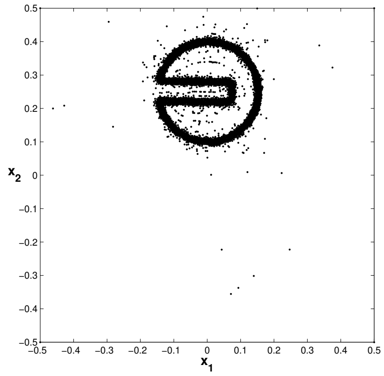
$t = t_{32}$



$t = t_{32}$



$t = t_{47}$



$t = t_{47}$

Figure 6: The slotted cylinder. 3D view on the solution u (left column) and the corresponding node distribution (right column) at time $t_{16} = 16\tau$, $t_{32} = 32\tau$, and $t_{47} = 47\tau$.

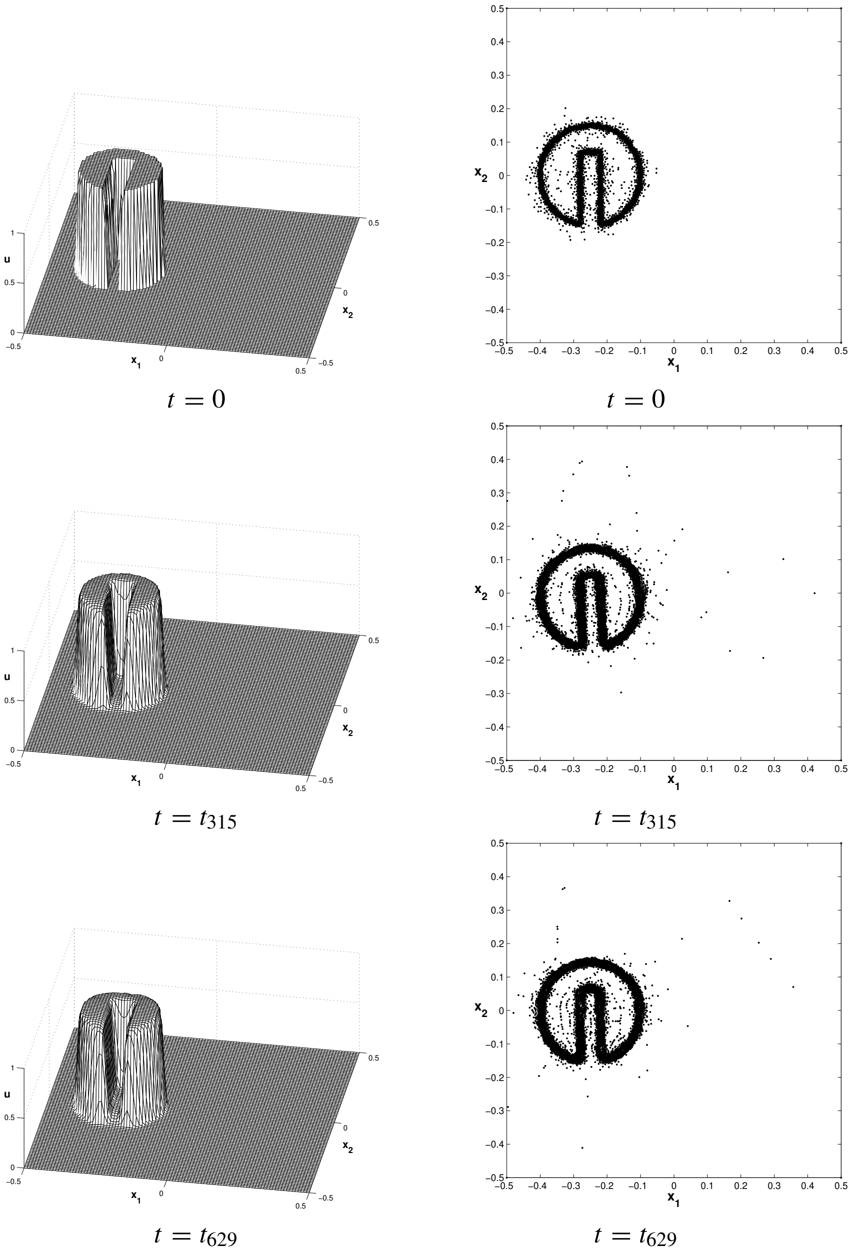
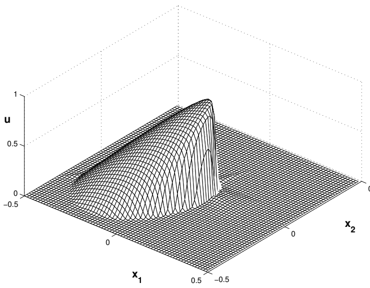
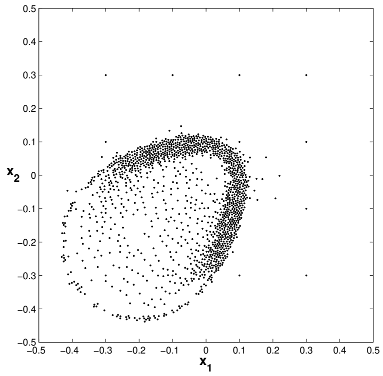


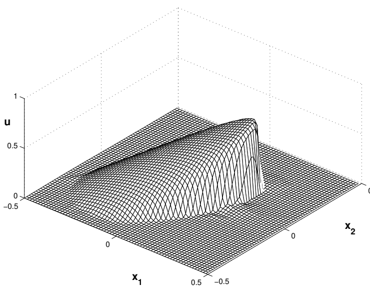
Figure 7: The slotted cylinder. 3D view and node distribution of the initial condition (top row), where $t = 0$, after five revolutions (middle row), at time $t_{315} = 315\tau$, and after ten revolutions (bottom row), at time $t_{629} = 629\tau$.



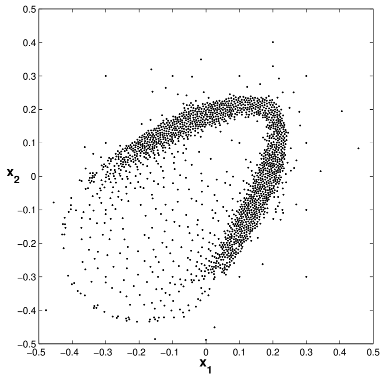
$t = t_{110}$



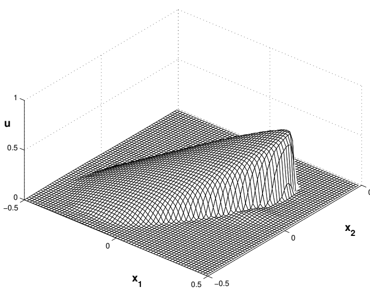
$t = t_{110}$



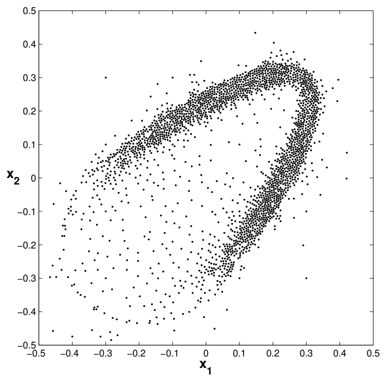
$t = t_{220}$



$t = t_{220}$



$t = t_{330}$



$t = t_{330}$

Figure 8: Burgers equation. Evolution of the solution u at three different time steps, $t_{110} = 110\tau$, $t_{220} = 220\tau$, and $t_{330} = 330\tau$ (left column), and the corresponding node distribution (right column).

References

[1] BEHRENS J. AND ISKE A., *Grid-free adaptive semi-Lagrangian advection using radial basis functions*, Computers and Mathematics with Applications **43** 3-5 (2002), 319–327.

[2] BEHRENS J., ISKE A. AND PÖHN S., *Effective node adaption for grid-free semi-Lagrangian advection*, in: "Discrete Modelling and Discrete Algorithms in Continuum Mechanics", (Eds. Sonar Th. and Thomas I.), Logos Verlag, Berlin 2001, 110–119.

[3] BEHRENS J., ISKE A. AND KÄSER M., *Adaptive meshfree method of backward characteristics for nonlinear transport equations*, in: "Meshfree Methods for Partial Differential Equations", (Eds. Griebel M. and Schweitzer M.A.), Springer-Verlag, Heidelberg 2002, 21–36.

[4] BERNSTEIN S., *Sur les fonctions absolument monotones*, Acta Mathematica **51** (1928), 1–66.

[5] BUHMANN M.D., *Radial basis functions*, Acta Numerica (2000), 1–38.

[6] BURGERS J.M., *Application of a model system to illustrate some points of the statistical theory of free turbulence*, Proc. Acad. Sci. Amsterdam **43** (1940), 2–12.

[7] DEUFLHARD P. AND BORNEMANN F., *Scientific computing with ordinary differential equations*, Springer, New York 2002.

[8] DUCHON J., *Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces*, R.A.I.R.O. Analyse Numeriques **10** (1976), 5–12.

[9] DUCHON J., *Splines minimizing rotation-invariant semi-norms in Sobolev spaces*, in: "Constructive Theory of Functions of Several Variables", (Eds. Schempp W. and Zeller K.), Springer, Berlin 1977, 85–100.

[10] DUCHON J., *Sur l'erreur d'interpolation des fonctions de plusieurs variables par les D^m-splines*, R.A.I.R.O. Analyse Numeriques **12** (1978), 325–334.

[11] DURRAN D.R., *Numerical methods for wave equations in geophysical fluid dynamics*, Springer, New York 1999.

[12] DYN N., *Interpolation of scattered data by radial functions*, in: "Topics in Multivariate Approximation", (Eds. Chui C.K., Schumaker L.L. and Utreras F.I.), Academic Press 1987, 47–61.

[13] DYN N., *Interpolation and approximation by radial and related functions*, (Eds. Chui C.K., Schumaker L.L. and Ward J.D.), Academic Press, New York 1989, 211–234.

[14] FALCONE M. AND FERRETTI R., *Convergence analysis for a class of high-order semi-Lagrangian advection schemes*, SIAM J. Numer. Anal. **35** 3 (1998), 909–940.

[15] FASSHAUER G.E., *Solving differential equations with radial basis functions: multilevel methods and smoothing*, Advances in Comp. Math. **11** (1999), 139–159.

[16] FRANKE C. AND SCHABACK R., *Convergence orders of meshless collocation methods using radial basis functions*, Advances in Comp. Math. **8** (1998), 381–399.

[17] FRANKE C. AND SCHABACK R., *Solving partial differential equations by collocation using radial basis functions*, Appl. Math. Comp. **93** (1998), 73–82.

[18] GELFAND I.M. AND SHILOV G.E., *Generalized functions, Vol. 1: properties and operations*, Academic Press, New York 1964.

[19] GELFAND I.M. AND VILENKIN N.Y., *Generalized functions, Vol. 4: applications of harmonic analysis*, Academic Press, New York 1964.

[20] GUO K., HU S. AND SUN X., *Conditionally positive definite functions and Laplace-Stieltjes integrals*, J. Approx. Theory **74** (1993), 249–265.

[21] GUTZMER T. AND ISKE A., *Detection of discontinuities in scattered data approximation*, Numerical Algorithms **16** 2 (1997), 155–170.

[22] HIGHAM N.J., *Accuracy and stability of numerical algorithms*, SIAM, Philadelphia 1996.

[23] ISKE A., *Charakterisierung bedingt positiv definiter Funktionen für multivariate Interpolationsmethoden mit radialen Basisfunktionen*, Dissertation, Universität Göttingen 1994.

[24] ISKE A., *Scattered data modelling using radial basis functions*, in: "Tutorials on Multiresolution in Geometric Modelling", (Eds. Iske A., Quak E. and Floater M.S.), Springer-Verlag, Heidelberg 2002, 205–242.

[25] ISKE A., *On the approximation order and numerical stability of local Lagrange interpolation by polyharmonic splines*, to appear in: "Proceedings of the *5th International Conference on Multivariate Approximation*", Bommerholz 2002.

[26] KANSA E.J., *Multiquadrics - a scattered data approximation scheme with applications to computational fluid dynamics - I: surface approximations and partial derivative estimates*, Comput. Math. Appl. **19** (1990), 127–145.

[27] KANSA E.J., *Multiquadrics - a scattered data approximation scheme with applications to computational fluid dynamics - I: solution to parabolic, hyperbolic, and elliptic partial differential equations*, Comput. Math. Appl. **19** (1990), 147–161.

[28] LEVEQUE R.L., *Numerical methods for conservation laws*, Birkhäuser, Basel 1992.

[29] MADYCH W.R. AND NELSON S.A., *Multivariate interpolation: a variational theory*, manuscript, 1983.

[30] MADYCH W.R. AND NELSON S.A., *Multivariate interpolation and conditionally positive definite functions I*, Approx. Theory Appl. **4** (1988), 77–89.

[31] MADYCH W.R. AND NELSON S.A., *Multivariate interpolation and conditionally positive definite functions II*, Math. Comp. **54** (1990), 211–230.

[32] MEINGUET J., *Multivariate interpolation at arbitrary points made simple*, Z. Angew. Math. Phys. **30** (1979), 292–304.

[33] MEINGUET J., *An intrinsic approach to multivariate spline interpolation at arbitrary points*, in: "Polynomial and spline approximations", (Ed. Sahney N.B.), Reidel, Dordrecht 1979, 163–190.

[34] MEINGUET J., *Surface spline interpolation: basic theory and computational aspects*, in: "Approximation Theory and Spline Functions", (Eds. Singh S.P., Bury J.H. and Watson B.), Reidel, Dordrecht 1984, 127–142.

[35] MICCHELLI C.A., *Interpolation of scattered data: distance matrices and conditionally positive definite functions*, Constr. Approx. **2** (1986), 11–22.

[36] MORTON K.W., *Numerical solution of convection-diffusion problems*, Chapman & Hall, London 1996.

[37] NARCOWICH F.J. AND WARD J.D., *Norm estimates for the inverses of a general class of scattered-data radial-function interpolation matrices*, J. Approx. Theory **69** (1992), 84–109.

[38] POWELL M.J.D., *The theory of radial basis function approximation in 1990*, in: "Advances in numerical analysis II: wavelets, subdivision and radial basis functions", (Ed. Light W.A.), Clarendon Press, Oxford 1992, 105–210.

[39] PREPARATA F.P. AND SHAMOS M.I., *Computational geometry*, Springer, New York 1988.

[40] ROBERT A., *A stable numerical integration scheme for the primitive meteorological equations*, Atmos. Ocean **19** (1981), 35–46.

[41] ROBERT A., *A semi-Lagrangian and semi-implicit numerical integration scheme for the primitive meteorological equations*, J. Meteor. Soc. Japan **60** (1982), 319–324.

[42] SCHABACK R., *Multivariate interpolation and approximation by translates of a basis function*, in: "Approximation theory VIII, Vol. 1: approximation and interpolation", (Eds. Chui C.K. and Schumaker L.L.), World Scientific, Singapore 1995, 491–514.

[43] SCHABACK R., *Stability of radial basis function interpolants*, in: "Approximation theory X: wavelets, splines and applications", (Eds. Chui C.K., Schumaker L.L. and Stöckler J.) Vanderbilt Univ. Press, Nashville 2002, 433–440.

[44] SCHABACK R. AND WENDLAND H., *Using compactly supported radial basis functions to solve partial differential equations*, in: "Boundary Element Technology XIII", (Eds. Chen C.S., Brebbia C.A. and Pepper D.W.), WitPress, Southampton, Boston 1999, 311–324.

[45] SCHABACK R. AND WENDLAND H., *Characterization and construction of radial basis functions*, in: "Multivariate Approximation and Applications", (Eds. Dyn N., Leviatan D., Levin D. and Pinkus A.), Cambridge University Press, Cambridge 2001, 1–24.

[46] SCHABACK R. AND WENDLAND H., *Inverse and saturation theorems for radial basis function interpolation*, Math. Comp. **71** (2002), 669–681.

[47] SCHOENBERG I.J., *Metric spaces and positive definite functions*, Trans. Amer. Math. Soc. **44** (1938), 522–536.

[48] SCHOENBERG I.J., *Metric spaces and completely monotone functions*, Annals of Mathematics **39** (1938), 811–841.

[49] STANIFORTH A. AND CÔTÉ J., *Semi-Lagrangian integration schemes for atmospheric models – a review*, Mon. Wea. Rev. **119** (1991), 2206–2223.

[50] WENDLAND H., *Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree*, Advances in Comp. Math. **4** (1995), 389–396.

[51] WENDLAND H., *Meshless Galerkin methods using radial basis functions*, Math. Comp. **68** (1999), 1521–1531.

[52] WIDDER D.V., *An introduction to transform theory*, Academic Press, New York 1971.

[53] WU Z., *Multivariate compactly supported positive definite radial functions*, Advances in Comp. Math. **4** (1995), 283–292.

[54] WU Z. AND SCHABACK R., *Local error estimates for radial basis function interpolation of scattered data*, IMA J. Numer. Anal. **13** (1993), 13–27.

[55] ZALESAK S.T., *Fully multidimensional flux-corrected transport algorithms for fluids*, J. Comput. Phys. **31** (1979), 335–362.

[56] *Special issue on radial basis functions and partial differential equations*, (Eds. Kansa E.J. and Hon Y.C.), Comput. Math. Appl. **43** (2002).

AMS Subject Classification: 65D15, 65M25.

Armin ISKE
Zentrum Mathematik
Technische Universität München
D-85747 Garching, GERMANY
e-mail: `iske@ma.tum.de`

Rend. Sem. Mat. Univ. Pol. Torino
Vol. 61, 3 (2003)
Splines and Radial Functions

F. Caliò - E. Marchetti - R. Pavani

ABOUT THE DEFICIENT SPLINE COLLOCATION METHOD FOR PARTICULAR DIFFERENTIAL AND INTEGRAL EQUATIONS WITH DELAY

Abstract. The aim of this paper is to present the application of a particular collocation method (recently developed by the authors) to numerically solve some differential and Volterra integral equations with constant delay. The unknown function is approximated by using deficient spline functions. The existence and uniqueness of the numerical solution are studied; some aspects of the problem related to the estimation of the errors as well as the convergence properties are presented. Numerical examples are provided.

1. Introduction

In recent years a great deal of dynamical processes has been described and investigated by differential and integral equations with deviating arguments. It is well known that the versatility of such equations in modelling processes in various applications, especially in physics, engineering, biomathematics, medical sciences, economics, etc., provides the best, and sometimes the only, realistic simulation of observed phenomena.

Since solutions of such equations in general are not found explicitly, methods for their approximate solutions reveal very useful.

Recently we have proposed a deficient spline collocation method to approximate the solution of the first and second order delay differential equations (DDEs) [2] also in the neutral case (NDDEs) [3], [4], [5] and the solution of Volterra integral equations with delay (VDIEs) [6].

More precisely, we deal with the numerical solutions by combining two classic Numerical Analysis methodologies: approximation through the spline functional class and determination of the approximating function by a collocation method. In literature the two techniques are frequently used separately, but they are rarely combined to solve delay differential and integral equations. For instance in [1] they are applied in the numerical solution of first order delay differential equation, in [8] they are extended in the numerical solution of second order differential equations with delay, in [10] they are proposed in the case of Volterra integral equations. In all those works some advantages of that technique are outlined.

In any case, from those works one can draw the conclusion that spline methods are characterized by a large application spectrum, thanks to their weak convergence requirements, but they are affected by serious stability problems when their order increases. This explains why spline collocation techniques are not so often used.

In our works [2], [3], [4], [5], [6] taking into account that the phenomena described by the delay equations are very irregular, we proposed the following ideas:

i) the use of low order splines, in order to guarantee stability;

ii) the weakness of the continuity requirements at connecting points, so that lowly regular functions can be satisfactorily dealt with.

Therefore we propose the collocation using deficient splines (as defined in the next section), namely splines pertaining to class $\mathcal{C}^{m-2}$ (deficiency 1), where $m \in \mathbb{N}$, $m \geq 2$, is the spline degree.

Consequently we can use the advantages of the two (collocation and deficient spline) aspects.

The collocation methods provide the global spline expression, therefore they are selected:

i) in the case of DDE and NDDE, to eliminate the problems due to high-order interpolation, in the continuous extension

ii) in the case of VDIE, to use the expression of the spline in the evaluation of integrals in intervals preceding the current one

iii) to allow the use of variable intervals and spline degrees

iv) to state numerical models such that existence and unicity of the solution can be proved

v) to implement a simple and efficient algorithm.

About the deficient spline of polynomial degree $m \geq 2$:

i) we choose a classical convenient expression of the spline

ii) we choose low polynomial degree spline to maintain stability of the method and to deal with weakly regular solution

iii) weakening the spline regularity in the linking points, we can adapt the continuity class of the spline approximating to solutions at very low regularity.

As the equations with delay argument concerning the modelling processes are very often linear and with constant delay, in this paper we study the application of the numerical method proposed to these cases. We refer to works [2] to [6] for non linear cases.

In the second section we study the numerical model both for differential and Volterra integral equations. In the third section we give some numerical examples.

2. The description of spline collocation method

In this section we study the application of the numerical method to some linear DDE, NDDE and VDIE with constant delay.

2.1. The case of delay differential equations

We consider the following second order delay differential equation (DDE):

$$(1)\qquad \begin{array}{l} y''(x) = k_1 y(x) + k_2 y'(x) + f(x, y(g(x)), y'(g(x))), \quad x \in [a, b] \\ y(x) = \varphi(x), y'(x) = \varphi'(x), \quad x \in [\alpha, a], \quad \alpha \leq a, \quad \alpha = \underset{x\in[a,b]}{Inf}(g(x)) \end{array}$$

$\alpha \leq g(x) \leq x, \quad x \in [\alpha, b], \quad \varphi \in \mathcal{C}^{m-2}[\alpha, a], m > 2, \; k_1, k_2 \in \mathbb{R}$
$f : [a, b] \times \mathcal{C}^1[\alpha, b] \times \mathcal{C}[\alpha, b] \to \mathbb{R}.$

We suppose verified the hypotheses so that the problem (1) has a unique solution $y \in \mathcal{C}^2[a, b] \cap \mathcal{C}^1[\alpha, b]$ (see [7]).

As it is known (see [7]) jump discontinuities can occur in various higher order derivatives of the solution y even if f, g, φ are analytic in their arguments. Such jump discontinuities are caused by the delay function g and propagate from the point a, moving ahead with the increasing order of derivatives.

If we denote the jump discontinuities by $\{\xi_j\}$, it is also known that ξ_j are the roots of equation $g(\xi_j) = \xi_{j-1}$ [7]; $\xi_0 = a$ is a jump discontinuity of φ (or of its derivatives). Since in (1) the delay function g does not depend on y (no state depending argument), we can consider the jump discontinuities for sufficiently high order derivatives to be such that $\xi_0 < \xi_1 < ... < \xi_{k-1} < \xi_k < ... < \xi_M$.

In the following we will consider $g(x) := x - \tau$ ($\tau \in \mathbb{R}$, $\tau > 0$) so that $\xi_j = a + j\tau$ ($j = 0, 1, ..., M$) and $\alpha = a - \tau$.

We shall construct for the problem (1) a deficient polynomial spline approximating function of degree $m \geq 3$, denoted by $s : [a, b] \to \mathbb{R}$, $s \in \mathcal{S}_m$, $s \in \mathcal{C}^{m-2}$, which will be defined on each interval $[\xi_j, \xi_{j+1}]$ ($j = 0, 1, ..., M - 1$). For this construction we shall use successively the collocation methods as in [8]. Let us consider the first interval $[\xi_0, \xi_1]$ which is $[a, \xi_1]$. Let us define a uniform partition $\xi_0 = t_0 < t_1 < ... < t_{k-1} < t_k < ... < t_N = \xi_1$ where $t_j := t_0 + jh$ ($j = 0, 1, ..., N$), $h = (\xi_1 - \xi_0)/N$. On the first interval $[t_0, t_1]$ the spline component is defined by

$$(2)\qquad \begin{aligned} s_0(t) \;\; : \;\; &= \varphi(t_0) + \varphi'(t_0)(t - t_0) + \varphi''(t_0)(t - t_0)^2/2 + ... + \\ &+ \varphi^{(m-2)}(t_0)(t - t_0)^{m-2}/(m - 2)! + \\ &+ a_0/(m - 1)!(t - t_0)^{m-1} + b_0/m!(t - t_0)^m \end{aligned}$$

with a_0, b_0 to be determined by the following system of collocation conditions:

$$\begin{cases} s_0''(t_0 + h/2) = k_1 s_0(t_0 + h/2) + k_2 s_0'(t_0 + h/2) + \\ \qquad + f(t_0 + h/2, \varphi(t_0 + h/2 - \tau), \varphi'(t_0 + h/2 - \tau)) \\ s_0''(t_1) = k_1 s_0(t_1) + k_2 s_0'(t_1) + f(t_1, \varphi(t_1 - \tau), \varphi'(t_1 - \tau)) \end{cases}$$

Once determined the polynomial (2), on the next interval $[t_1, t_2]$, we define

$$(3) \quad s_1(t) := \sum_{j=0}^{m-2} s_0^{(j)}(t_1)(t-t_1)^j/j! + a_1/(m-1)!(t-t_1)^{m-1} + b_1/m!(t-t_1)^m$$

where $s_0^{(j)}(t_1),\ 0 \le j \le m-2$, are left-hand limits of derivative as $t \to t_1$ of the segment of s defined on $[t_0, t_1]$ and a_1, b_1 are determined from the following collocation conditions:

$$\begin{cases} s_1''(t_1+h/2) = k_1 s_1(t_1+h/2) + k_2 s_1'(t_1+h/2) + \\ \qquad + f(t_1+h/2, \varphi(t_1+h/2-\tau), \varphi'(t_1+h/2-\tau)) \\ s_1''(t_2) = k_1 s_1(t_2) + k_2 s_1'(t_2) + f(t_2, \varphi(t_2-\tau), \varphi'(t_2-\tau)) \end{cases}$$

We remark that the peculiarity of these collocation conditions is the fact that they take into account the historical behaviour of the approximating spline, which is relevant for the delay nature of the considered equation.

Analogously for $t \in [t_k, t_{k+1}]$ we have

$$(4) \quad s_k(t) := \sum_{j=0}^{m-2} s_{k-1}^{(j)}(t_k)(t-t_k)^j/j! + a_k/(m-1)!(t-t_k)^{m-1} + b_k/m!(t-t_k)^m$$

where $s_{k-1}^{(j)}(t_k) = \lim_{t \to t_k} s_{k-1}^{(j)}(t)$, $t \in [t_{k-1}, t_k]$ and a_k, b_k are determined from

$$(5) \quad \begin{cases} s_k''(t_k+\frac{h}{2}) = k_1 s_k(t_k+h/2) + k_2 s_k'(t_k+h/2) + \\ \qquad + f(t_k+\frac{h}{2}, \varphi(t_k+\frac{h}{2}-\tau), \varphi'(t_k+\frac{h}{2}-\tau)) \\ s_k''(t_{k+1}) = k_1 s_k(t_{k+1}) + k_2 s_k'(t_{k+1}) + \\ \qquad + f(t_{k+1}, \varphi(t_{k+1}-\tau), \varphi'(t_{k+1}-\tau)) \end{cases}$$

In general the spline function $s : [a, b] \to \mathbb{R}$, $(s \in \mathcal{S}_m,\ s \in \mathcal{C}^{m-2})$ approximating the solution of (1) on the interval $I_i := [\xi_i, \xi_{i+1}]$ $(i = 0, 1, ..., M-1)$ is defined in $[t_k, t_{k+1}]$ where $t_k := t_0 + kh,\ k = 0, 1, ..., N-1;\ t_0 := \xi_i,\ t_N = \xi_{i+1},\ h := \frac{\xi_{i+1}-\xi_i}{N}$ as:

$$(6) \quad s_{k/I_i}(t) := \sum_{j=0}^{m-2} s_{k-1/I_i}^{(j)}(t_k)(t-t_k)^j/j! + \frac{a_k}{(m-1)!}(t-t_k)^{m-1} + \frac{b_k}{m!}(t-t_k)^m$$

with $a_k,\ b_k$ determined, as in (5) by

$$(7) \quad \begin{cases} s_{k/I_i}''(t_k+\frac{h}{2}) = k_1 s_{k/I_i}(t_k+h/2) + k_2 s_{k/I_i}'(t_k+h/2) + \\ \qquad + f(t_k+\frac{h}{2}, s_{I_{i-1}}(t_k+\frac{h}{2}-\tau), s_{I_{i-1}}'(t_k+\frac{h}{2}-\tau)) \\ s_{k/I_i}''(t_{k+1}) = k_1 s_{k/I_i}(t_{k+1}) + k_2 s_{k/I_i}'(t_{k+1}) + \\ \qquad + f(t_{k+1}, s_{I_{i-1}}(t_{k+1}-\tau), s_{I_{i-1}}'(t_{k+1}-\tau)) \end{cases}$$

where $s_{I_{i-1}} \in \mathcal{S}_m,\ s_{I_{i-1}} \in \mathcal{C}^{m-2}$ is the spline approximating the solution of (1) on the interval I_{i-1}.

In the following, to simplify the notations, the theoretical results will be related only to (5); their generalization to (7) is immediate.

If we set

$$A_k(t) = \sum_{j=0}^{m-2} s_{k-1}^{(j)}(t_k)(t-t_k)^j/j! \tag{8}$$

then (5) becomes:

$$\begin{cases} \frac{a_k}{(m-3)!}\left(1-\frac{h}{2(m-2)}\left(k_1\frac{h}{2(m-1)}+k_2\right)\right)\left(\frac{h}{2}\right)^{m-3}+ \\ \quad +\frac{b_k}{(m-2)!}\left(1-\frac{h}{2(m-1)}\left(k_1\frac{h}{2m}+k_2\right)\right)\left(\frac{h}{2}\right)^{m-2}= \\ \quad -A_k''(t_k+\frac{h}{2})+k_1A_k(t_k+\frac{h}{2})+k_2A_k'(t_k+\frac{h}{2})+ \\ \quad +f(t_k+\frac{h}{2},\varphi(t_k+\frac{h}{2}-\tau),\varphi'(t_k+\frac{h}{2}-\tau)) \\ \\ \frac{a_k}{(m-3)!}\left(1-\frac{h}{m-2}\left(k_1\frac{h}{m-1}+k_2\right)\right)h^{m-3}+ \\ \quad +\frac{b_k}{(m-2)!}\left(1-\frac{h}{m-1}\left(k_1\frac{h}{m}+k_2\right)\right)h^{m-2}= \\ \quad -A_k''(t_{k+1})+k_1A_k(t_{k+1})+k_2A_k'(t_{k+1})+ \\ \quad +f(t_{k+1},\varphi(t_{k+1}-\tau),\varphi'(t_{k+1}-\tau)) \end{cases} \tag{9}$$

It remains to find under what conditions on h, the parameters $a_k, b_k,\ 0 \le k \le N-1$ can be uniquely determined from (9).

It is easy to prove the following:

THEOREM 1. *Let us consider the delay differential problems in (1). Under the hypotheses of existence and uniqueness of the analytic solution, there exists a unique spline approximation solution $s:[a,b]\to\mathbb{R}$, ($s\in\mathcal{S}_m,\ s\in\mathcal{C}^{m-2}$) of (1) given by the above construction for $h\neq 0$ if and only if the following condition is satisfied:*

$$\begin{vmatrix} 1-\frac{h}{2(m-2)}\left(k_1\frac{h}{2(m-1)}+k_2\right) & \frac{1}{2}\left(1-\frac{h}{2(m-1)}\left(k_1\frac{h}{2m}+k_2\right)\right) \\ 1-\frac{h}{m-2}\left(k_1\frac{h}{m-1}+k_2\right) & 1-\frac{h}{m-1}\left(k_1\frac{h}{m}+k_2\right) \end{vmatrix} \neq 0$$

COROLLARY 1. *If $k_1=k_2=0$ and $m\ge 3$ the condition is satisfied* $\forall h$ $(h\neq 0)$.

COROLLARY 2. *If $k_1=0$, $k_2\neq 0$ and $3\le m<10$ the condition is satisfied* $\forall h$ $(h\neq 0)$.

COROLLARY 3. *If $k_1\neq 0$, $k_2=0$ and $3\le m<10$ the condition is satisfied* $\forall h$ $(h\neq 0)$.

We can tackle by the same method also the following neutral delay differential equation (NDDE):

$$\begin{gathered} y'(x)=k_1y(x)+f(x,y(g(x)),y'(g(x))),\quad x\in[a,b] \\ y(x)=\varphi(x),\quad x\in[\alpha,a],\quad \alpha\le a,\quad \alpha=\underset{x\in[a,b]}{Inf}(g(x)) \\ \alpha\le g(x)\le x\ ,\quad x\in[\alpha,b] \end{gathered} \tag{10}$$

Let us assume that: $f : [a, b] \times \mathcal{C}^1[\alpha, b] \times \mathcal{C}[\alpha, b] \to \mathbb{R}$, $g \in \mathcal{C}[\alpha, b]$, $\alpha \leq g(x) \leq x$, $x \in [\alpha, b]$, $\varphi \in \mathcal{C}^{m-1}[\alpha, a]$, $m \geq 1$, $m \in N$, $k_1 \in \mathbb{R}$.

We suppose verified the hypotheses so that the problem (10) has a unique solution $y \in \mathcal{C}^1[a, b] \cap \mathcal{C}[\alpha, b]$ (see [7]).

Analogously to (1), we consider $g(x) := x - \tau$ ($\tau \in \mathbb{R}, \tau > 0$) and the jump discontinuities $\xi_j = a + j\tau$ $(j = 0, 1, ..., M)$, $\alpha = a - \tau$. In each interval $I_i = [\xi_i, \xi_{i+1}]$ $(i = 0, 1, ..., M - 1)$ we shall construct for the problem (10) a polynomial spline approximating function (6) of degree $m \geq 2$ and deficiency 1 and we determine the coefficients a_k, b_k through the following collocation system:

$$\begin{cases} s'_{k/I_i}(t_k + \frac{h}{2}) = k_1 s_{k/I_i}(t_k + h/2) + \\ \qquad + f(t_k + \frac{h}{2}, s_{I_{i-1}}(t_k + \frac{h}{2} - \tau), s'_{I_{i-1}}(t_k + \frac{h}{2} - \tau)) \\ s'_{k/I_i}(t_{k+1}) = k_1 s_{k/I_i}(t_{k+1}) + \\ \qquad + f(t_{k+1}, s_{I_{i-1}}(t_{k+1} - \tau), s'_{I_{i-1}}(t_{k+1} - \tau)) \end{cases}$$

It follows that in the first interval $[\xi_0, \xi_1]$ (the generalization to I_i, $i = 1, ..., M - 1$ is immediate) assuming $A_k(t)$ as in (8):

$$\begin{cases} \frac{a_k}{(m-2)!}\left(1 - k_1 \frac{h}{2(m-1)}\right)\left(\frac{h}{2}\right)^{m-2} + \frac{b_k}{(m-1)!}\left(1 - k_1 \frac{h}{2m}\right)\left(\frac{h}{2}\right)^{m-1} = \\ \qquad -A'_k(t_k + \frac{h}{2}) + k_1 A_k(t_k + \frac{h}{2}) + \\ \qquad + f(t_k + \frac{h}{2}, \varphi(t_k + \frac{h}{2} - \tau), \varphi'(t_k + \frac{h}{2} - \tau)) \\ \\ \frac{a_k}{(m-2)!}\left(1 - k_1 \frac{h}{m-1}\right) h^{m-2} + \frac{b_k}{(m-1)!}\left(1 - k_1 \frac{h}{m}\right) h^{m-1} = \\ \qquad -A'_k(t_{k+1}) + k_1 A_k(t_{k+1}) + \\ \qquad + f(t_{k+1}, \varphi(t_{k+1} - \tau), \varphi'(t_{k+1} - \tau)) \end{cases} \tag{11}$$

It is easy to prove the following:

THEOREM 2. *Let us consider the delay neutral differential problems in (10). Under the hypotheses of existence and uniqueness of the analytic solution, there exists a unique spline approximation solution* $s : [a, b] \to \mathbb{R}$, *(*$s \in \mathcal{S}_m$, $s \in \mathcal{C}^{m-2}$*) of (10) given by the above construction for* $h \neq 0$, *if and only if the following condition is satisfied:*

$$\begin{vmatrix} 1 - k_1 \frac{h}{2(m-1)} & \frac{1}{2}\left(1 - k_1 \frac{h}{2m}\right) \\ 1 - k_1 \frac{h}{m-1} & 1 - k_1 \frac{h}{m} \end{vmatrix} \neq 0$$

COROLLARY 4. *If* $k_1 = 0$ *and* $m \geq 2$ *the condition is satisfied* $\forall h$ $(h \neq 0)$.

COROLLARY 5. *If* $k_1 \neq 0$ *and* $2 \leq m < 9$ *the condition is satisfied* $\forall h$ $(h \neq 0)$.

REMARK 1. As the condition of the Theorem provides the non singularity of the coefficient matrix of system (11), its extension to a linear system of n delay differential equations of first order is immediate.

REMARK 2. For the consistency and convergence of the numerical solutions of (1) and (10) we can take into account the results obtained in more general cases. In [2] , [5] it is shown that the spline collocation method appears as (m–1)step-method. Consequently for $m = 3$ and $m = 4$ the cubic and quartic approximating splines yield the same values of the solution of (1) in the knots as discrete 2-step and 3-step method respectively. Analogously for $m = 2$ and $m = 3$ the trapezoidal and the Simpson's rule give the same discrete solutions of (10) as quadratic and cubic spline respectively. Consequently it is possible to prove consistency and convergence of the method. *The numerical stability of the method is not guaranteed (see [2], [5]) when $m > 4$ for (1) and when $m > 3$ for (10).*

2.2. The case of Volterra integral equations

Let us use the same method for the following Volterra integral equation with positive and constant delay (VDIE):

$$y(x) = \int_0^x k_1\, y(t)dt + \int_0^{x-\tau} K_2(x,t,y(t))dt + g(x),\ x \in J = [0,T] \tag{12}$$

with $k_1 \in \mathbb{R}$, the delay $\tau \in \mathbb{R},\ \tau > 0,\ y(x) = \phi(x)$ for $x \in [-\tau, 0)$.

We assume that the given functions $\phi : [-\tau, 0] \to \mathbb{R},\ g : J \to \mathbb{R},\quad K_2 : \Omega_\tau \times \mathbb{R} \to \mathbb{R}$ ($\Omega_\tau := J \times [-\tau, T-\tau]$) are at least continuous on their domains such that (12) possesses a unique solution $y \in \mathcal{C}(J)$.

If $K_2 = 0$ equation (12) reduces to Volterra integral equation (VIE).

We suppose that $T = M\tau$ for some $M \in \mathbb{N}$. For $N \in \mathbb{N}$ (which satisfies $N/M \in \mathbb{N}$), let $h = T/N$ and $r = \tau/h \in \mathbb{N}$.

Chosen $t_i = ih$ ($i = -r, ..., 0, 1, ..., N;\ t_{-r} = -\tau, t_N = T$), the coefficients $a_k,\ b_k$ of $s_k(t)$ defined in $[t_k, t_{k+1}]$ ($k = 0, ..., N-1$) with $\tau \le t_k < T$ are determined through the following collocation system:

$$\begin{cases} s_k(t_k + \frac{h}{2}) = \sum_{j=0}^{k-1} \int_{jh}^{(j+1)h} k_1\, s_j(t)dt + \int_{kh}^{kh+\frac{h}{2}} k_1 s_k(t)dt + \\ \qquad + \sum_{j=0}^{k-1-r} \int_{jh}^{(j+1)h} K_2(t_k + \frac{h}{2}, t, s_j(t))dt + \\ \qquad + \int_{(k-r)h}^{(k-r)h+\frac{h}{2}} K_2(t_k + \frac{h}{2}, t, s_{k-r}(t))dt + g(t_k + \frac{h}{2}) \\ s_k(t_{k+1}) = \sum_{j=0}^{k} \int_{jh}^{(j+1)h} k_1\, s_j(t)dt + \\ \qquad + \sum_{j=0}^{k-r} \int_{jh}^{(j+1)h} K_2(t_{k+1}, t, s_j(t))dt + g(t_{k+1}) \end{cases} \tag{13}$$

and if $0 \le t_k < \tau$ from

$$\begin{cases} s_k(t_k + \frac{h}{2}) = \sum_{j=0}^{k-1} \int_{jh}^{(j+1)h} k_1\, s_j(t)dt + \\ \qquad + \int_{kh}^{kh+\frac{h}{2}} k_1 s_k(t)dt + \sum_{j=k-r}^{-1} \int_{jh}^{(j+1)h} K_2(t_k + \frac{h}{2}, t, s_j(t))dt + \\ \qquad - \int_{(k-r)h}^{(k-r)h+\frac{h}{2}} K_2(t_k + \frac{h}{2}, t, s_{k-r}(t))dt + g(t_k + \frac{h}{2}) \\ s_k(t_{k+1}) = \sum_{j=0}^{k} \int_{jh}^{(j+1)h} k_1\, s_j(t)dt + \\ \qquad + \sum_{j=k-r+1}^{-1} \int_{jh}^{(j+1)h} K_2(t_{k+1}, t, s_j(t))dt + g(t_{k+1}) \end{cases} \tag{14}$$

provided that $s_k(t) = \phi(t)$ in $[t_k, t_{k+1}]$ $(k = -r, ..., -1)$.

Consequently (13), with $A_k(t)$ as in (8), becomes:

$$\begin{cases} \frac{a_k}{(m-1)!}(\frac{h}{2})^{m-1}\left(1-k_1\frac{h}{2m}\right)+\frac{b_k}{m!}(\frac{h}{2})^m\left(1-k_1\frac{h}{2(m+1)}\right)= \\ \quad -A_k(t_k+\frac{h}{2})+\int_{kh}^{kh+\frac{h}{2}}k_1A_k(t)dt+\sum_{j=0}^{k-1}\int_{jh}^{(j+1)h}k_1(A_j(t) \\ \quad +\frac{a_j}{(m-1)!}(t-t_j)^{m-1}+\frac{b_j}{m!}(t-t_j)^m)dt+ \\ \quad +\sum_{j=0}^{k-1-r}\int_{jh}^{(j+1)h}K_2(t_k+\frac{h}{2},t,A_j(t)+\frac{a_j}{(m-1)!}(t-t_j)^{m-1}+ \\ \quad +\frac{b_j}{m!}(t-t_j)^m)dt+ \\ \quad +\int_{(k-r)h}^{(k-r)h+\frac{h}{2}}K_2(t_k+\frac{h}{2},t,A_{k-r}(t)+\frac{a_{k-r}}{(m-1)!}(t-t_{k-r})^{m-1}+ \\ \quad +\frac{b_{k-r}}{m!}(t-t_{k-r})^m)dt+g(t_k+\frac{h}{2}) \\ \\ \frac{a_k}{(m-1)!}h^{m-1}\left(1-k_1\frac{h}{m}\right)+\frac{b_k}{m!}h^m\left(1-k_1\frac{h}{m+1}\right)= \\ \quad -A_k(t_{k+1})+\int_{kh}^{(k+1)h}k_1A_k(t)dt+\sum_{j=0}^{k-1}\int_{jh}^{(j+1)h}k_1(A_j(t)+ \\ \quad +\frac{a_j}{(m-1)!}(t-t_j)^{m-1}+\frac{b_j}{m!}(t-t_j)^m)dt+ \\ \quad +\sum_{j=0}^{k-r}\int_{jh}^{(j+1)h}K_2(t_{k+1},t,A_j(t)+\frac{a_j}{(m-1)!}(t-t_j)^{m-1}+ \\ \quad +\frac{b_j}{m!}(t-t_j)^m)dt+g(t_{k+1}) \end{cases}$$

Analogously (14) becomes:

$$\begin{cases} \frac{a_k}{(m-1)!}(\frac{h}{2})^{m-1}\left(1-k_1\frac{h}{2m}\right)+\frac{b_k}{m!}(\frac{h}{2})^m\left(1-k_1\frac{h}{2(m+1)}\right)= \\ \quad -A_k(t_k+\frac{h}{2})+\int_{kh}^{kh+\frac{h}{2}}k_1A_k(t)dt+ \\ \quad +\sum_{j=0}^{k-1}\int_{jh}^{(j+1)h}k_1(A_j(t)+\frac{a_j}{(m-1)!}(t-t_j)^{m-1}+\frac{b_j}{m!}(t-t_j)^m)dt+ \\ \quad +\sum_{j=k-r}^{-1}\int_{jh}^{(j+1)h}K_2(t_k+\frac{h}{2},t,A_j(t)+\frac{a_j}{(m-1)!}(t-t_j)^{m-1}+ \\ \quad +\frac{b_j}{m!}(t-t_j)^m)dt+ \\ \quad -\int_{(k-r)h}^{(k-r)h+\frac{h}{2}}K_2(t_k+\frac{h}{2},t,A_{k-r}(t)+\frac{a_{k-r}}{(m-1)!}(t-t_{k-r})^{m-1}+ \\ \quad +\frac{b_{k-r}}{m!}(t-t_{k-r})^m)dt+g(t_k+\frac{h}{2}) \\ \\ \frac{a_k}{(m-1)!}h^{m-1}\left(1-k_1\frac{h}{m}\right)+\frac{b_k}{m!}h^m\left(1-k_1\frac{h}{m+1}\right)= \\ \quad -A_k(t_{k+1})+\int_{kh}^{(k+1)h}k_1A_k(t)dt+\sum_{j=0}^{k-1}\int_{jh}^{(j+1)h}k_1(A_j(t)+ \\ \quad +\frac{a_j}{(m-1)!}(t-t_j)^{m-1}+\frac{b_j}{m!}(t-t_j)^m)dt+ \\ \quad +\sum_{j=k-r+1}^{-1}\int_{jh}^{(j+1)h}K_2(t_{k+1},t,A_j(t)+\frac{a_j}{(m-1)!}(t-t_j)^{m-1}+ \\ \quad +\frac{b_j}{m!}(t-t_j)^m)dt+g(t_{k+1}) \end{cases}$$

It is easy to prove the following:

THEOREM 3. *Let us consider equation (12). Under the hypotheses of existence and uniqueness of the analytic solution, there exists a unique spline approximation solution* $s : [0, T] \to \mathbb{R}$, *(*$s \in \mathcal{S}_m$, $s \in \mathcal{C}^{m-2}$*) of (12) given by the above construction* *for h $\neq 0$ if and only if the following condition is satisfied:*

$$\begin{vmatrix} 1-k_1\frac{h}{2m} & \frac{1}{2}(1-k_1\frac{h}{2(m+1)}) \\ 1-k_1\frac{h}{m} & 1-k_1\frac{h}{m+1} \end{vmatrix} \neq 0$$

COROLLARY 6. *If* $k_1 = 0$ *and* $m \geq 2$ *the condition is satisfied* $\forall h$ $(h \neq 0)$.

COROLLARY 7. *If* $k_1 \neq 0$ *and* $2 \leq m < 8$ *the condition is satisfied* $\forall h$ $(h \neq 0)$.

REMARK 3. About the convergence and the numerical stability of the method applied to (12) we refer to [9].

3. Numerical examples

In the following we present some numerical results to enlighten the features of the presented numerical method. We emphasize that we will show examples just for cases with exact solutions belonging to a low regularity class, because our method is dedicated just to these cases. In all the examples the existence and uniqueness of the numerical solution is guaranteed for any value of the integration step h.

Our computer programs are written in MATLAB5.3, which has a machine precision $\varepsilon \simeq 10^{-16}$.

Our first example refers to the following second order DDE

$$y''(t) = \left|t - \frac{1}{2}\right| + y'(t-1)$$

which is to be solved on [0, 1] with history $y(t) = 1$ for $t \leq 0$.
The analytical solution is $y(t) = -\frac{1}{6}t^3 + \frac{1}{4}t^2 + 1$ in $[0, 1/2]$, and $y(t) = \frac{1}{6}t^3 - \frac{1}{4}t^2 + \frac{1}{4}t + \frac{23}{24}$ in $[1/2, 1]$; therefore the solution $y(t) \in \mathcal{C}^2$, so using $m = 4$, our approximating deficient spline function belongs exactly to the same class of regularity of the analytical solution. We remark that this problem is smooth, as at $t = 1/2$ the left third derivative slightly differs from the right third derivative. We chose a quite large value $h = 0.1$ and we get a comparison with an analogous collocation method using classical splines. The following Table 1 reports the errors es_d and es_c we obtained respectively using deficient spline $s_d \in \mathcal{S}_4$, $s_d \in \mathcal{C}^2$ and classical spline $s_c \in \mathcal{S}_4$,

$s_c \in \mathcal{C}^3$.

t	es_d	es_c
0.1	1.0E-4	1.4E-4
0.2	3.1E-5	9.4E-5
0.3	2.7E-5	2.9E-4
0.4	6.8E-5	1.0E-3
0.5	9.1E-5	2.0E-3
0.6	1.0E-4	3.3E-3
0.7	1.4E-4	4.3E-3
0.8	1.9E-4	5.0E-3
0.9	2.5E-4	5.4E-3
1.0	3.3E-4	5.5E-3

Table 1

It is clear that deficient spline behaves better than classical spline, as it exhibits the same class of regularity as the analytical solution, even when large integration steps are used.

As a second example, we consider the following NDDE:

$$y'(t) = -500\frac{y(t-1)}{y'(t-1)}$$

which is to be solved on $[0, 2]$ with history $y(t) = e^{-t}$ for $t \leq 0$.
The analytical solution is $y(t) = 500t + 1$ in $[0, 1]$, and $y(t) = -250t^2 + 499t + 252$ in $[1, 2]$. Therefore the solution $y(t) \in \mathcal{C}^0[0, 2]$; so using $m = 2$, our approximating deficient spline function belongs exactly to the same class of regularity as the analytical solution. We emphasize that this problem is really rough, as at $t = 1$ the left first derivative and the right first derivative differ significantly: indeed $y'(1)_- = 500$ whereas $y'(1)_+ = -1$. Therefore we could expect some numerical troubles. On the contrary our method deals very well with this kind of problems, as already pointed out. We chose a quite large value $h = 0.25$; at $t = 1$ we obtain numerically the exact value and at $t = 2$ the final absolute error is $2.6E - 4$. This suffices to show how our method is accurate and efficient and cheap. Figure 1 reports the behavior of the analytical solution (solid line) together with the numerical solution (rectangles) for the case $h = 0.25$.
As a third example we consider the following system of first order DDE's suggested as Example 1 in [11] . The equations

$$\begin{aligned} y_1'(t) &= y_1(t-1) \\ y_2'(t) &= y_1(t-1) + y_2(t-0.2) \\ y_3'(t) &= y_2(t-1) \end{aligned}$$

are to be solved on $[0, 1]$ with history $y_1(t) = 1,\ y_2(t) = 1,\ y_3(t) = 1$ for $t \leq 0$.

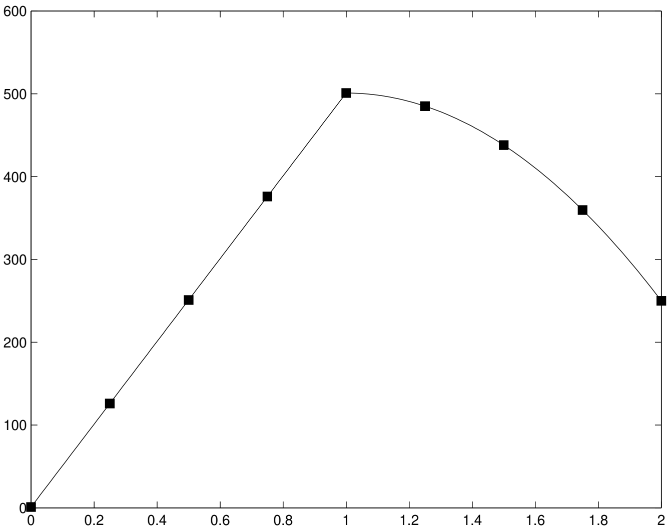


Figure 1: h=0.25

A comparison between the solutions computed by means of **dde23** (see [11]) and by our method (with $h = 0.01$) show that the three solution curves coincide and the number of required flops has the same order of magnitude; in this case using our method no advantages occur, because the solutions are very regular. However this example is interesting in order to show that *our method works efficiently also for systems of equations and moreover that different delays are allowed and can be conveniently handled.* We remark that in this case the linear system to be solved has M equations and M unknowns, where $M = 2n$ with n equal to the number of given first order equations; in this case $M = 6$.

About the integral equations, at first we consider the following Volterra integral equation without delay argument. This example is reported just to show that even in this case our method works really well, when solution exhibits low regularity.

$$
\begin{aligned}
y(x) &= g(x) + \int_0^x y(s)ds \\
g(x) &= \begin{cases} \frac{x^3}{3} - x^2 + \frac{1-x}{4} & 0 \le x \le \frac{1}{2} \\ -\left(\frac{x^3}{3} - x^2 + \frac{1-x}{4}\right) - \frac{1}{6} & \frac{1}{2} \le x \le 1 \end{cases}
\end{aligned}
$$

The exact solution is:

$$y(x) = \left| x^2 - \frac{1}{4} \right|$$

We computed our solution in $x = 1$. Using $m = 2$, we built splines $s \in S_2$, $s \in$

$\mathcal{C}^0[0,1]$, that is of the same class of regularity of the analytical solution. Even in this case, we obtain very good numerical results; in particular at $t = 1$ our error is comparable with the machine precision, even when large integration steps are used.

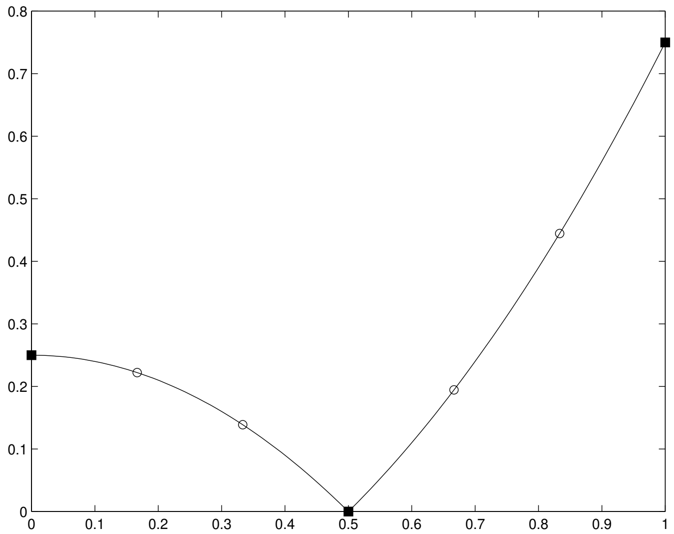


Figure 2: h=0.5

Figure 2 refers just to the case $h = 0.5$; there solid line shows the exact solution in [0, 1] ; rectangles show the integration points and circles show intermediate points of our numerical solution computed by means of spline analytical expression relating to each integration interval. It is evident that even when a large integration step is used, our numerical solution coincides with the analytical one.

At last we consider the following integral equation with delay arguments:

$$\begin{aligned} y(x) &= g(x) + \int_0^x y(s)ds - \int_0^{x-\tau} y(s)ds \\ \tau &= 1, \qquad y(x) = 0 \quad \text{for} \quad x \in [-1, 0] \\ g(x) &= \begin{cases} 100x - 50x^2 \quad \text{for} \ x \in [0, 1/2] \\ -400(x-1)^3 + 100(x-1)^4 - 75/4 \quad \text{for} \ x \in [1/2, 1] \end{cases} \end{aligned}$$

The exact solution is:

$$y(x) = \begin{cases} 100x \quad \text{for} \ x \in [0, 1/2] \\ -400(x-1)^3 \quad \text{for} \ x \in [1/2, 1] \end{cases}$$

Even in this case the solution $y(x)$ to be approximated belongs to class $\mathcal{C}^0[0,1]$. We used a large integration step $h_1 = 0.5$ in $[0, 1/2]$ and a shorter step h_2 in $[1/2, 1]$, where the solution is not linear.

Figure 3 refers to the case $h_1 = 0.5$ and $h_2 = 0.125$; there solid line shows the exact solution in [0, 1] together with the history in $[-1, 0]$; rectangles show the numerical solution in the integration points and circles show the numerical solution in the intermediate points (computed by means of the analytical expression of spline).

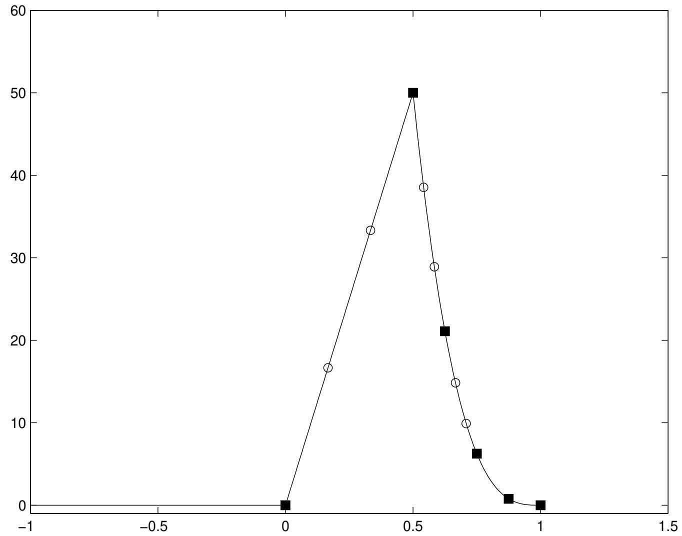


Figure 3: $h_1 = 0.5$, $h_2 = 0.125$

It is evident that even in this case results are very satisfactory.
In more details, the numerical solution in $x = 1$ is computed with an error equal to $1.0E - 2$ when $h_2 = 0.25$ and with an error equal to $6.6E - 4$ when $h_2 = 0.125$.

References

[1] Bellen A. and Micula G., *Spline approximations for neutral delay differential equations*, Revue d'Analyse Numer. et de Theorie de l'Approximation **23** (1994), 117–125.

[2] Caliò F., Marchetti E., Micula G. and Pavani R., *A new deficient spline functions collocation method for the second order delay differential equations*, Pure Math. Appl. **13** 1-2 (2003), 97–109.

[3] Caliò F., Marchetti E. and Pavani R., *On existence and uniqueness of first order NDDE solution obtained by deficient spline collocation*, Dip. di Matematica, Politecnico di Milano, Internal Report n. 482/P, 2001.

[4] CALIÒ F., MARCHETTI E. AND PAVANI R., *Peculiar spline collocation method to solve rough and stiff delay problems*, to appear on Mathematica-Rev. Anal. Numer. Theorie Approx (2002).

[5] CALIÒ F., MARCHETTI E. AND PAVANI R., *An algorithm based on deficient splines to solve delay neutral differential problems*, submitted 2001.

[6] CALIÒ F., MARCHETTI E., MICULA G. AND PAVANI R., *About some Volterra problems solved by a particular spline*, to appear on Studia Univ. Babes-Bolyai (2003).

[7] DRIVER R.D., *Ordinary and delay differential equations*, Springer-Verlag, Berlin 1977.

[8] MICULA G. AND AKCA H., *Approximate solutions of the second order differential equations with deviating argument*, Mathematica-Rev. Anal. Numer. Theorie Approx. **30** (1988), 37–46.

[9] MICULA G., *Numerische Behandlung der Volterra-Integralgleichungen mit Splines* Studia Univ. Babes-Bolyai, Mathematica **XXIV** 2 (1979).

[10] HU Q.Y., *Multilevel correction for discrete collocation solutions of Volterra integral equations with delay arguments*, Appl. Numer. Math. **31** (1999), 159–171.

[11] SHAMPINE L.F. AND THOMPSON S., *Solving delay differential equations with dde23*, tutorial available in: *http://www.runet.edu/ ˜thompson/webddes/*.

AMS Subject Classification: 65L05, 65R20.

Franca CALIÒ, Elena MARCHETTI, Raffaella PAVANI
Dipartimento di Matematica
Politecnico di Milano
Piazza Leonardo da Vinci 32
20133 Milano, ITALIA
e-mail: `franca.calio@mate.polimi.it`
`elena.marchetti@mate.polimi.it`
`raffaella.pavani@mate.polimi.it`

Rend. Sem. Mat. Univ. Pol. Torino
Vol. 61, 3 (2003)
Splines and Radial Functions

C. Conti - L. Gori - F. Pitolli

SOME RECENT RESULTS ON A NEW CLASS OF BIVARIATE REFINABLE FUNCTIONS

Abstract. In this paper a new class of bivariate refinable functions is presented and some of its properties are investigated. The new class is constructed by convolving a tensor product refinable function of special type with $\chi_{[0,1]}$, the characteristic function of the interval [0, 1]. As in the case of box splines, the convolution product here used is the directional convolution product.

1. Introduction

It is well know that refinable functions play a key role in different fields like, just to mention two of the most significative, subdivision algorithms and wavelets. That is why there is an enormous amount of literature analyzing properties and applications of refinable functions in both the univariate and multivariate setting. In spite of their importance in many applications, the explicit form of refinable functions known in the literature reduces, in practice, to the two celebrated cases of B-splines and box-splines on uniform grids and of Daubechies refinable functions (see [2], [3], [6], and [7], for example). This is especially true in the multivariate setting where tensor product of univariate refinable functions are mainly taken into account. The considerations above motivated us in constructing and investigating a new family of bivariate non tensor-product refinable functions. Thus, starting with a bivariate function which is a tensor-product of finitely supported totally positive refinable functions, the new functions are obtained by using the directional convolution product with the characteristic function of the interval [0, 1]. The idea is definitely borrowed from box-splines but the bivariate function we start with is not the characteristic function of $[0, 1]^2$. The univariate functions used to construct the tensor product belong to a large class of refinable functions introduced in [9], [8] by the two last authors so that they will be called GP functions. The class of GP functions contains as a particular case the cardinal B-splines with which they share many useful properties. The differences between the B-splines and the GP functions are mainly due to the fact that the refinement mask is characterized by one or more extra parameters that afford additional degrees of freedom which reveals its effectiveness in several applications.

The outline of the paper is as follows. In Section 2 we first recall the definition of the directional convolution product of a bivariate function and a univariate function. Then, we investigate which properties of the bivariate function are preserved after the directional convolution with $\chi_{[0,1]}$, the characteristic function of $[0, 1]$, is made. In Section 3 the new class of bivariate refinable functions is characterized. In the closing Section 4 a few examples are presented.

2. Directional convolution

We start this Section by recalling the definition of the directional convolution product (see also [7] for a possible use of it).

DEFINITION 1. *Let $F : \mathbb{R}^2 \to \mathbb{R}$, $\quad g : \mathbb{R} \to \mathbb{R}$ be a bivariate and a univariate function, respectively, and let $\mathbf{e} \in \mathbb{Z}^2$ be a direction vector. The* convolution product *between F and g along the direction $\mathbf{e}$ is defined as*

$$(1) \qquad (F *_{\mathbf{e}} g)(x) := \int_{\mathbb{R}} F(x - \mathbf{e}t)g(t)\, dt, \qquad x \in \mathbb{R}^2.$$

Next, let Φ be a bivariate refinable function that is a solution of a refinement equation of type

$$(2) \qquad \Phi(x) = \sum_{\alpha \in \mathbb{Z}^2} a_\alpha \Phi(2x - \alpha), \qquad x \in \mathbb{R}^2,$$

where the set of coefficients a_α forms the so called refinement mask $\mathbf{a} = \{a_\alpha, \alpha \in \mathbb{Z}^2\}$. The mask $\mathbf{a}$ is supposed to be of compact support and satisfying $\sum_{\alpha \in \mathbb{Z}^2} a_{\alpha+2\gamma} = 1$ for all $\gamma \in \{0, 1\}^2$. Furthermore, we assume that the Fourier transform of Φ satisfies $\widehat{\Phi}(0) = 1$. Here we define the Fourier transform of a given function F as

$$(3) \qquad \hat{F}(\omega) := \int_{\mathbb{R}^2} F(x) e^{-i\omega \cdot x}\, dx.$$

Using the above introduced directional convolution product we defined the bivariate function $\Psi : \mathbb{R}^2 \to \mathbb{R}$

$$(4) \qquad \Psi(x) := (\Phi *_{\mathbf{e}} \chi_{[0,1]})(x) = \int_0^1 \Phi(x - \mathbf{e}t)\, dt,$$

where $\mathbf{e} \in \{-1, 0, 1\}^2$, and $\chi_{[0,1]}$, in the following for shortness χ, is the characteristic function of the unit interval $[0, 1]$.

PROPOSITION 1. *Let Φ be a refinable function with refinement mask* $\mathbf{a}$ *such that $\sum_{\alpha \in \mathbb{Z}^2} \Phi(\cdot - \alpha) = 1$. Then, the function Ψ defined in (4) is refinable with refinement mask*

$$\mathbf{b} = \{b_\alpha = \frac{a_\alpha + a_{\alpha-\mathbf{e}}}{2},\ \alpha \in \mathbb{Z}^2\}\,.$$

Furthermore, the integer translates of Ψ *form a partition of unity, namely* $\sum_{\alpha\in\mathbb{Z}^2} \Psi(\cdot - \alpha) = 1$.

Proof. By the Ψ definition we get

$$\begin{aligned}
\Psi(x) \quad &= \textstyle\int_0^1 \Phi(x - \mathbf{e}t)\, dt = \displaystyle\sum_{\alpha\in\mathbb{Z}^2} a_\alpha \int_0^1 \Phi(2(x - \mathbf{e}t) - \alpha)dt \\
&= \tfrac{1}{2} \sum_{\alpha\in\mathbb{Z}^2} a_\alpha \int_0^2 \Phi(2x - \mathbf{e}t - \alpha)dt \\
&= \tfrac{1}{2} \sum_{\alpha\in\mathbb{Z}^2} a_\alpha \left[\int_0^1 \Phi(2x - \mathbf{e}t - \alpha)dt + \int_0^1 \Phi(2x - \mathbf{e}t - \alpha - \mathbf{e})dt \right] \\
&= \tfrac{1}{2} \sum_{\alpha\in\mathbb{Z}^2} a_\alpha [\Psi(2x - \alpha) + \Psi(2x - \alpha - \mathbf{e})] \\
&= \sum_{\alpha\in\mathbb{Z}^2} \frac{1}{2}(a_\alpha + a_{\alpha-\mathbf{e}})\Psi(2x - \alpha)
\end{aligned}$$

so that Ψ is refinable with refinement mask $\mathbf{b} = \{b_\alpha = \frac{a_\alpha + a_{\alpha-\mathbf{e}}}{2},\ \alpha \in \mathbb{Z}^2\}$.
Next, since the Fourier transform of Ψ is $\widehat{\Psi}(\omega) = \widehat{\Phi}(\omega)\widehat{\chi}(\mathbf{e}\cdot\omega)$ for all $\omega \in \mathbb{R}^2$, from $\widehat{\Phi}(0) = 1$ it trivially follows that $\sum_{\alpha\in\mathbb{Z}^2} \Psi(\cdot - \alpha) = \widehat{\Psi}(0) = \widehat{\Phi}(0) = 1$ which is the partition of unity for the function Ψ. $\square$

A theorem is now dealing with the stability of Ψ. We recall that the function Ψ is L_2-stable if there exist two constants $0 < A \le B < \infty$ such that

$$0 < A||\mathbf{c}||_2 \le || \sum_{\alpha\in\mathbb{Z}^2} c_\alpha \Psi(\cdot - \alpha)||_2 \le B||\mathbf{c}||_2 \tag{5}$$

for any real sequence $\mathbf{c} = \{c_\alpha,\ \alpha \in \mathbb{Z}^2\}$ in $\ell^2(\mathbb{Z}^2)$.

THEOREM 1. *Let* $\{\Phi(\cdot - \alpha), \alpha \in \mathbb{Z}^2\}$ *be linear independent and such that* $\widehat{\Phi}(2\pi k) = \delta_{0,k}$*, where* $\delta_{0,k}$ *is the Kronecker symbol. Then, the integer translates of* Ψ *are linearly independent. Furthermore,* $\{\Psi(\cdot - \alpha), \alpha \in \mathbb{Z}^2\}$ *is a* L_2*-stable basis.*

Proof. To prove the linear independence, it is sufficient to show that the set of the complex periodic zeros of $\widehat{\Psi}$ is empty, that is

$$Z_{\Psi}^{C} = \{\theta \in \mathbb{C}^2 | \widehat{\Psi}(\theta + 2\pi k) = 0, \forall k \in \mathbb{Z}^2\} = \{\emptyset\}$$

(see [12] for details). Now, since

$$\widehat{\Psi}(\theta + 2\pi k) = \widehat{\Phi}(\theta + 2\pi k)\widehat{\chi}(\mathbf{e} \cdot (\theta + 2\pi k)),$$

if θ is not a multiple of 2π, then $\theta + 2\pi k \notin 2\pi\mathbb{Z}^2$ and $\widehat{\Phi}(\theta + 2\pi k) \neq 0$, $\widehat{\chi}(\mathbf{e} \cdot (\theta + 2\pi k)) \neq 0$, so that θ is not a periodic zero. If θ is a multiple of 2π, then $\theta + 2\pi k \in 2\pi\mathbb{Z}^2$ and

$$\widehat{\Phi}(\theta + 2\pi k) = \begin{cases} 0, & \text{if } k \neq K, \\ 1, & \text{if } k = K, \end{cases}$$

where $K := -\frac{\theta}{2\pi}$. Now, for $k = K$ one has $\widehat{\chi}(\mathbf{e} \cdot (\theta + 2\pi K)) = \widehat{\chi}(0) = 1$, so that θ is not a periodic zero. It follows the set Z_{Ψ}^{C} is empty.
We conclude with the observation that, obviously, also the set of the *real periodic zeros* of $\widehat{\Psi}$ is empty, that is

$$Z_{\Psi}^{R} = \{\theta \in \mathbb{R}^2 | \widehat{\Psi}(\theta + 2\pi k) = 0, \forall k \in \mathbb{Z}^2\} = \{\emptyset\},$$

which implies the L_2-stable stability of the system of the integer translates of Ψ as shown, again, in [12]. □

As a consequence of Theorem 1, the following corollary holds.

COROLLARY 1. *The refinable function Ψ generates a multi-resolution analysis on $L^2(\mathbb{R}^2)$.*

3. A new class of bivariate refinable functions

Aim of this Section is the construction of a specific class of refinable functions having all the properties of the Ψ function discussed in the previous section. As Φ refinable function we consider a tensor product of particular univariate functions, that is

$$\Phi^{\mathbf{H}_1,\mathbf{H}_2}(x) := \varphi^{\mathbf{H}_1}(x_1)\varphi^{\mathbf{H}_2}(x_2), \tag{6}$$

where $\mathbf{H}_1 = (n_1, h_1)$, $\mathbf{H}_2 = (n_2, h_2)$, and $x = (x_1, x_2)$, and where $\varphi^{\mathbf{H}_1}$, $\varphi^{\mathbf{H}_2}$ are univariate functions belonging to the class of one parameter refinable functions introduced in [9]. We recall that the refinement mask of a GP function of type $\varphi^{\mathbf{H}}$, $\mathbf{H} = (n, h)$, is supported on $[0, n+1]$ and has positive entries

$$a_{\alpha}^{\mathbf{H}} = \frac{1}{2^h}\left[\binom{n+1}{\alpha} + 4(2^{h-n} - 1)\binom{n-1}{\alpha-1}\right], \quad \alpha = 0, \dots, n+1, \tag{7}$$

so that, whenever $n = h$, the function $\varphi^{\mathbf{H}}$ reduces to the B-splines of degree n.

It is worthwhile to note that the real parameter h, $h \geq n \geq 2$, is an additional parameter which turns out to be useful for getting higher flexibility in the applications.

It is easy to see that the symbol associated with the refinement mask in (7) is

$$p^{\mathbf{H}}(z) = \frac{1}{2^h}(1+z)^{n-1}(z^2 + (2^{h-n+2} - 2)z + 1)\,. \tag{8}$$

For any n and h, ($h \geq n > 2$), the function $\varphi^{\mathbf{H}}$ belongs to $C^{n-2}(\mathbb{R})$, is centrally symmetric and the function system $\{\varphi^{\mathbf{H}}(x-\alpha), \alpha \in \mathbb{Z}\}$ is linearly independent, stable and satisfies $\sum_{\alpha\in\mathbb{Z}} \varphi^{\mathbf{H}}(x-\alpha) = 1$ for all $x \in \mathbb{R}$. Moreover, the Fourier transform $\widehat{\varphi}^{\mathbf{H}}(\omega)$ vanishes if and only if $\omega \in 2\pi\mathbb{Z} \setminus \{0\}$.

With the $\varphi^{\mathbf{H}}$ refinable functions at hand we are able to construct a new class of bivariate refinable functions using the direction convolution product with direction $\mathbf{e} = (1, 1)$. We define the function $\Psi^{\mathbf{H}_1,\mathbf{H}_2}$ as

$$\begin{aligned} \Psi^{\mathbf{H}_1,\mathbf{H}_2}(x) &:= (\Phi^{\mathbf{H}_1,\mathbf{H}_2} *_{\mathbf{e}} \chi)(x) \\ &= \int_0^1 \Phi^{\mathbf{H}_1,\mathbf{H}_2}(x - \mathbf{e}t)\, dt = \int_0^1 \varphi^{\mathbf{H}_1}(x_1 - t)\varphi^{\mathbf{H}_2}(x_2 - t)\, dt\,. \end{aligned} \tag{9}$$

Note that the support of $\Psi^{\mathbf{H}_1,\mathbf{H}_2}$ satisfies

$$supp(\Psi^{\mathbf{H}_1,\mathbf{H}_2}) \subset supp(\Phi^{\mathbf{H}_1,\mathbf{H}_2}) + [0,1]^2\,,$$

where $supp(\Phi^{\mathbf{H}_1,\mathbf{H}_2}) = [0, n_1 + 1] \times [0, n_2 + 1]$. Moreover, the function $\Psi^{\mathbf{H}_1,\mathbf{H}_2}$ is such that

$$\widehat{\Psi}^{\mathbf{H}_1,\mathbf{H}_2}(\omega_1, \omega_2) = \widehat{\varphi}^{\mathbf{H}_1}(\omega_1)\widehat{\varphi}^{\mathbf{H}_2}(\omega_2)\widehat{\chi}(\omega_1 + \omega_2)\,, \tag{10}$$

and its refinement mask and associated symbol are

$$\begin{aligned} \mathbf{b}^{\mathbf{H}_1,\mathbf{H}_2} &= \left\{ \frac{(a_\alpha^{\mathbf{H}_1,\mathbf{H}_2} + a_{\alpha-\mathbf{e}}^{\mathbf{H}_1,\mathbf{H}_2})}{2},\ \alpha \in \mathbb{Z}^2 \right\}, \\ P^{\mathbf{H}_1,\mathbf{H}_2}(z) &= p^{\mathbf{H}_1}(z_1)p^{\mathbf{H}_2}(z_2)\tfrac{1}{2}(1 + z_1 z_2)\,, \end{aligned} \tag{11}$$

where $\mathbf{a}^{\mathbf{H}_1,\mathbf{H}_2} = \{a_{\alpha_1}^{\mathbf{H}_1} a_{\alpha_2}^{\mathbf{H}_2},\ \alpha = (\alpha_1, \alpha_2) \in \mathbb{Z}^2\}$ is the mask of the tensor product.

Last, due to the results in Section 2, $\Psi^{\mathbf{H}_1,\mathbf{H}_2}$ has linearly independent integer translates and it generates a multi-resolution analysis on $L^2(\mathbb{R}^2)$.

4. Examples

In this Section we show the refinement masks and the graphs of some refinable functions constructed using the directional convolution strategy.

We start by setting $\mathbf{H}_1 = \mathbf{H}_2 = (3, h)$, and $h \geq 3$. The refinement mask $\mathbf{a}^{(3,h)}$ of the univariate refinable functions for different values of h are listed below while the graphs of these functions, obtained by performing five steps of the subdivision algorithm, are shown in Figure 1.

$$\mathbf{a}^{(3,3)} = \frac{1}{2^3}\{1, 4, 6, 4, 1\}, \quad \mathbf{a}^{(3,4)} = \frac{1}{2^4}\{1, 8, 14, 8, 1\},$$

$$\mathbf{a}^{(3,8)} = \frac{1}{2^8}\{1, 128, 254, 128, 1\}.$$

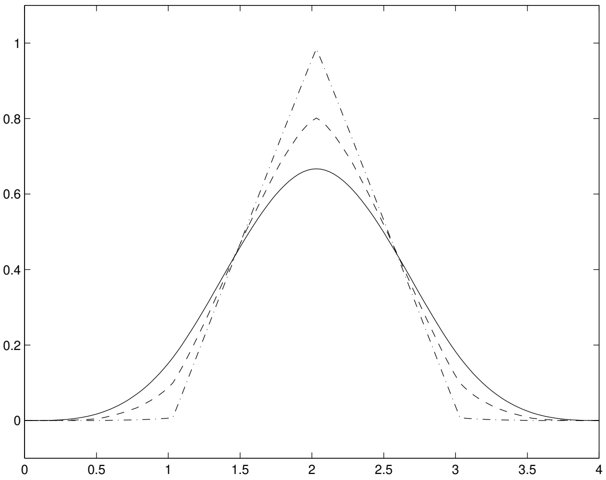


Figure 1: Graphs of the functions $\varphi^{(3,3)}(-)$, $\varphi^{(3,4)}(--)$ and $\varphi^{(3,8)}(.-)$

Note that $\varphi^{(3,3)}$ is just the cubic B-spline with uniform knots.

The bivariate refinement masks corresponding to the tensor product refinable functions $\Phi^{\mathbf{H}_1,\mathbf{H}_2}$ we construct from the previous functions for $\mathbf{H}_1 = \mathbf{H}_2 = (3, 3)$ and $\mathbf{H}_1 = \mathbf{H}_2 = (3, 4)$ are

$$\mathbf{a}^{(3,3),(3,3)} = \frac{1}{2^6}\begin{bmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & 36 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix},$$

$$\mathbf{a}^{(3,4),(3,4)} = \frac{1}{2^8}\begin{bmatrix} 1 & 8 & 14 & 8 & 1 \\ 8 & 64 & 112 & 64 & 8 \\ 14 & 112 & 196 & 112 & 14 \\ 8 & 64 & 112 & 64 & 8 \\ 1 & 8 & 14 & 8 & 1 \end{bmatrix}$$

while for $\mathbf{H}_1 = \mathbf{H}_2 = (3, 8)$ the refinement mask is

$$\mathbf{a}^{(3,8),(3,8)} = \frac{1}{2^{16}} \begin{bmatrix} 1 & 128 & 254 & 128 & 1 \\ 128 & 16384 & 32512 & 16384 & 128 \\ 254 & 32512 & 64516 & 32512 & 254 \\ 128 & 16384 & 32512 & 16384 & 128 \\ 1 & 128 & 254 & 128 & 1 \end{bmatrix}.$$

The associated refinable functions obtained by three steps of the corresponding subdivision algorithm are shown in Fig. 2, Fig. 3 and Fig. 4 where, for shortness, the function $\Phi^{\mathbf{H}_1,\mathbf{H}_2}$ with $\mathbf{H}_1 = \mathbf{H}_2$ is denoted just as $\Phi^{\mathbf{H}_1}$.

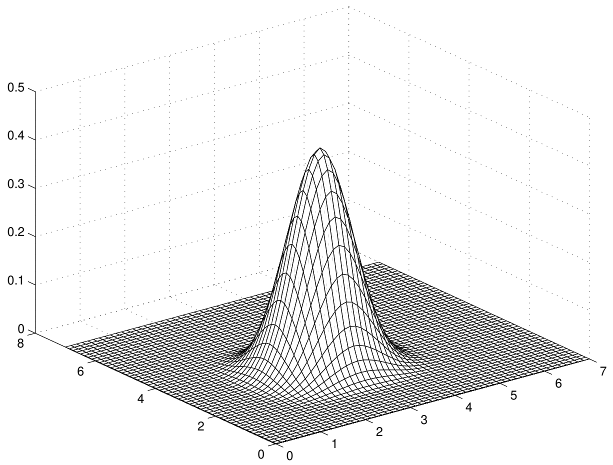


Figure 2: Graph of the function $\Phi^{(3,3)}$

Finally, the refinement mask of the convolved functions for $\mathbf{H}_1 = \mathbf{H}_2 = (3, 3)$ and $\mathbf{H}_1 = \mathbf{H}_2 = (3, 4)$ are

$$\mathbf{b}^{(3,3),(3,3)} = \frac{1}{2^{7}} \begin{bmatrix} 0 & 1 & 4 & 6 & 4 & 1 \\ 1 & 8 & 22 & 28 & 17 & 4 \\ 4 & 22 & 48 & 52 & 28 & 6 \\ 6 & 28 & 52 & 48 & 22 & 4 \\ 4 & 17 & 28 & 22 & 8 & 1 \\ 1 & 4 & 6 & 4 & 1 & 0 \end{bmatrix},$$

$$\mathbf{b}^{(3,4),(3,4)} = \frac{1}{2^{9}} \begin{bmatrix} 0 & 1 & 8 & 14 & 8 & 1 \\ 1 & 16 & 78 & 120 & 65 & 8 \\ 8 & 78 & 224 & 260 & 120 & 14 \\ 14 & 120 & 260 & 224 & 78 & 8 \\ 8 & 65 & 120 & 78 & 16 & 1 \\ 1 & 8 & 14 & 8 & 1 & 0 \end{bmatrix}$$

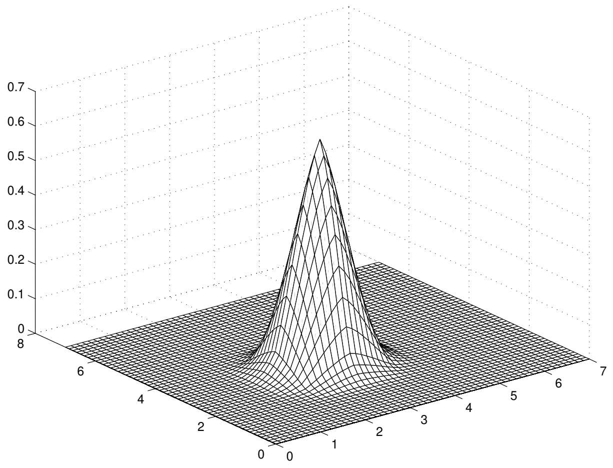


Figure 3: Graph of the function $\Phi^{(3,4)}$

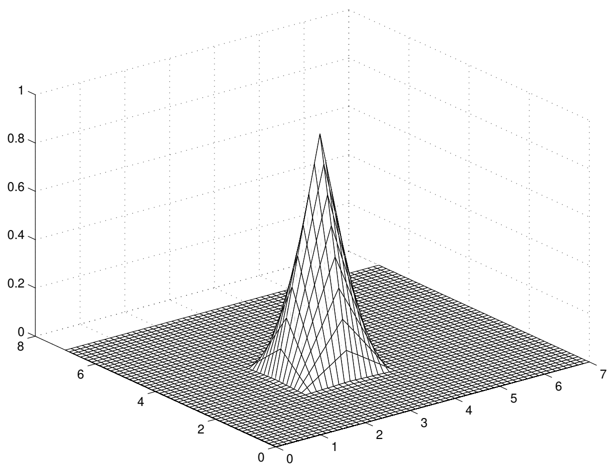


Figure 4: Graph of the function $\Phi^{(3,8)}$

and for $\mathbf{H}_1 = \mathbf{H}_2 = (3, 8)$

$$\mathbf{b}^{(3,8),(3,8)} = \frac{1}{2^{17}} \begin{bmatrix} 0 & 1 & 128 & 254 & 128 & 1 \\ 1 & 256 & 16638 & 32640 & 16385 & 128 \\ 128 & 16638 & 65024 & 80900 & 32640 & 254 \\ 254 & 32640 & 80900 & 65024 & 16638 & 128 \\ 128 & 16385 & 32640 & 16638 & 256 & 1 \\ 1 & 128 & 254 & 128 & 1 & 0 \end{bmatrix},$$

with corresponding graphs in Fig. 5, Fig. 6 and Fig. 7 (obtained, again, by three steps of the corresponding subdivision algorithm).

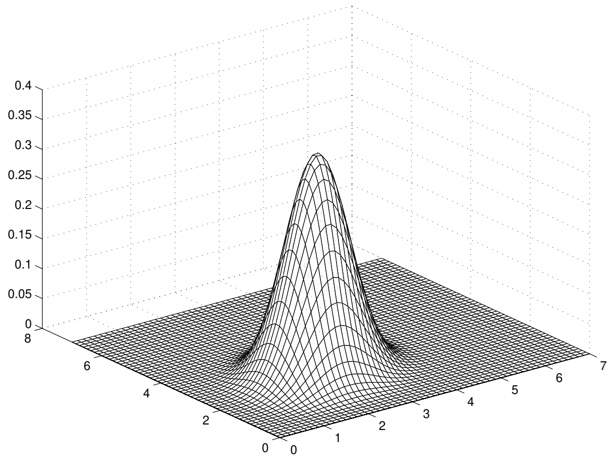


Figure 5: Graph of the function $\Psi^{(3,3)}$

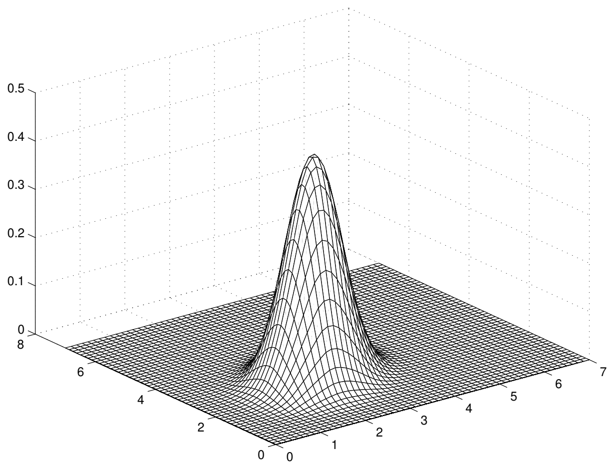


Figure 6: Graph of the function $\Psi^{(3,4)}$

Applications of the new refinable functions of type $\Psi^{\mathbf{H}_1,\mathbf{H}_2}$ are presently under investigation.

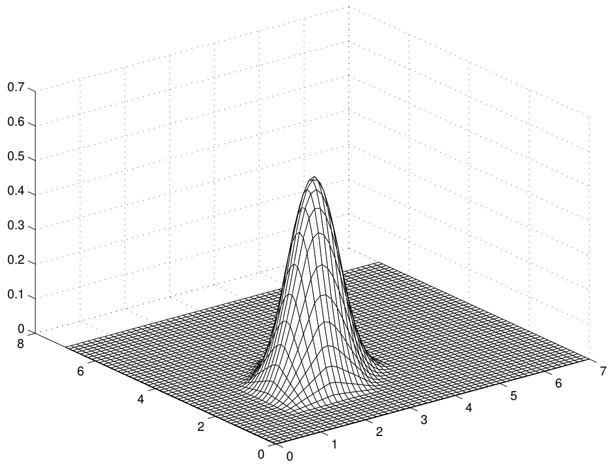


Figure 7: Graph of the function $\Psi^{(3,8)}$

References

[1] CAVARETTA A.S., DAHMEN W. AND MICCHELLI C.A., *Stationary subdivision*, Mem. Am. Math. Soc. **93**, Amer. Math. Soc., New York 1991.

[2] CHUI C.K., *Multivariate splines*, CBMS-NSF Series Applied Mathematics **54**, SIAM Publications, Philadelphia 1988.

[3] DE BOOR C., HÖLLIG K. AND RIEMENSCHNEIDER S., *Box-splines*, Springer-Verlag, New York 1993.

[4] DAHMEN W. AND MICCHELLI C.A., *Translates of multivariate splines*, Lin. Alg. Appl. **52-53** (1983), 217–234.

[5] DAHMEN W. AND MICCHELLI C.A., *On the solutions of certain systems of partial difference equations and linear dependence of trsnslates of box splines*, Trans. Amer. Math. Soc. **292** (1985), 305–320.

[6] DAUBECHIES I., *Ten lectures on wavelets*, SIAM, Philadelphia, Pennsylvania 1992.

[7] DYN N. AND LEVIN D., *Subdivision schemes in geometric modelling*, Acta Numerica, Cambridge University Press 2002, 73–144.

[8] GORI L. AND PITOLLI F., *A class of totally positive refinable functions*, Rendiconti di Matematica, Serie VII **20** (2000), 305–322.

[9] GORI L. AND PITOLLI F., *Multiresolution analysis based on certain compactly supported refinable functions*, in: "Proceeding of International Conference on

Approximation and Optimization" (Eds. Coman G., Breckner W.W. and Blaga P.), Transilvania Press 1997, 81–90.

[10] JIA R.Q. AND MICCHELLI C.A., *On linear independence on integer translates of a finite number of functions*, Proc. Edin. Math. Soc. **36** (1992), 69–85.

[11] MICCHELLI C.A., *The mathematical aspects of geometric modeling*, SIAM, Philadelfia, Pennsylvania 1995.

[12] RON A., *A necessary and sufficient condition for the linear independence of the integer translates of a compactly supported distribution*, Constr. Approx. **5** (1989), 297–308.

AMS Subject Classification: 65D15, 65D17, 41A63.

Costanza CONTI
Dip. di Energetica "Sergio Stecco"
Università di Firenze
Via Lombroso 6/17
50134 Firenze, ITALY
e-mail: `c.conti@ing.unifi.it`

Laura GORI, Francesca PITOLLI
Dip. Me.Mo.Mat.
Università di Roma "La Sapienza"
Via A. Scarpa 10
00161 Roma, ITALY
e-mail: `gori@dmmm.uniroma1.it`
`pitolli@dmmm.uniroma1.it`

Rend. Sem. Mat. Univ. Pol. Torino
Vol. 61, 3 (2003)
Splines and Radial Functions

C. Dagnino - V. Demichelis - E. Santi

ON OPTIMAL NODAL SPLINES AND THEIR APPLICATIONS

Abstract. We present a survey on optimal nodal splines and some their applications. Several approximation properties and the convergence rate, both in the univariate and bivariate case, are reported.

The application of such splines to numerical integration has been considered and a wide class of quadrature and cubature rules is presented for the evaluation of singular integrals, Cauchy principal value and Hadamard finite-part integrals. Convergence results and condition number are given.

Finally, a nodal spline collocation method, for the solution of Volterra integral equations of the second kind with weakly singular kernel, is also reported.

1. Introduction

It is well known that the polynomial spline approximation operators for real-valued functions are of great usefulness in the applications.

In their construction, it is desirable to obtain some nice properties as in particular:

1. the operator can be applied to a wide class of functions, including, for example, continuous or integrable functions;
2. they are local in the sense that can depend only on the values of f in a small neighbourhood of the evaluation point x;
3. the operators allow to approximate smooth functions f with an order of accuracy comparable to the best spline approximation. The key for obtaining operators with such property is to require that they reproduce appropriate class of polynomials.

The approximating splines obtained by applying the quasi-interpolatory operator defined in [24] satisfy the above properties and, recently, they have been widely used in the construction of integration formulas and in the numerical solution of integral and integro-differential equations, see, for instance, [3,4,7,10,13,22,27,23,28,30,32] and references therein.

This review paper is concerning the optimal nodal spline operators that, besides the properties 1., 2., 3., have the advantage of being interpolatory. These splines, introduced by DeVilliers and Rohwer [17,18] and studied in [12,14,16,19], have been utilized for constructing integration rules for the evaluation of weakly and strongly singular integrals also defined in the Hadamard finite part sense, in one or two dimensions and, more recently, for a collocation method producing the numerical solution of weakly singular Volterra integral equations.

In Section 2., after a brief outline of the construction of one-dimensional nodal spline operators, we shall present the tensor product of optimal nodal splines, recalling also some convergence results.

Section 3. is devoted to the application of the nodal spline operators in the approximation of different kind of 1D or 2D integrals and the main convergence results of the corresponding integration formulas are reported.

Finally, Section 4. deals with a collocation method, based on nodal splines, for the numerical solution of linear Volterra equation with weakly singular kernel.

2. Optimal nodal splines and their tensor product

2.1. One dimensional nodal splines

Let $J = [a, b]$ be a given finite interval of the real line $\mathbb{R}$, for a fixed integer $m \geq 3$ and $n \geq m-1$, we define a partition Π_n of J by

$$\Pi_n : a = \tau_0 < \tau_1 < ... < \tau_n = b \; ,$$

generally called "primary partition". We insert $m-2$ distinct points throughout $(\tau_\nu, \tau_{\nu+1})$, $\nu = 0, ..., n-1$ obtaining a new partition of J

$$X_n : a = x_0 < x_1, < ... < x_{(m-1)n} = b,$$

where $x_{(m-1)i} = \tau_i$, $i = 0, ..., n$. Let

$$R_n = \max_{\substack{0 \leq k, j \leq n-1 \\ |k-j|=1}} \frac{\tau_{k+1} - \tau_k}{\tau_{j+1} - \tau_j} \; , \tag{1}$$

we say that the sequence of partitions $\{\Pi_n; n = m-1, m, ...\}$ is locally uniform (l.u.) if, for all n, there exists a constant $A \geq 1$ such that $R_n \leq A$, i.e.

$$\frac{1}{A} \leq \frac{\tau_{k+1} - \tau_k}{\tau_{j+1} - \tau_j} \leq A \; , \quad k, j = 0, 1, ..., n-1 \text{ and } |k-j| = 1 \; . \tag{2}$$

Since the convergence results of the nodal splines we shall consider are based on the local uniformity property of the primary partitions sequence and one of our objectives is the use of graded meshes, the following proposition shows that a sequence of primary graded partitions is l.u. [8]. For the definition of graded partitions see for example [2].

PROPOSITION 1. *Let $[a, b]$ be a finite interval. The sequence of partitions $\{\Pi_n\}$, obtained by using graded meshes of the form*

$$\tau_i = a + \left(\frac{i}{n}\right)^r (b-a) \; , \quad 0 \leq i \leq n \; ,$$

with grading exponent $r \in \mathbb{R}$ *assumed* ≥ 1*, is l.u., i.e. it satisfies (2) with* $A = 2^r - 1$.

Now, after introducing two integers [16]

$$i_0 = \begin{cases} \frac{1}{2}(m+1) & m \quad \text{odd} \\ \\ \frac{1}{2}m + 1 & m \quad \text{even} \end{cases} \qquad \text{and} \quad i_1 = (m+1) - i_0$$

and two integer functions

$$p_\nu = \begin{cases} 0 & \nu = 0, 1, ..., i_1 - 2 \\ \nu - i_1 + 1 & \nu = i_1 - 1, ..., n - i_0 \\ n - (m-1) & \nu = n - i_0 + 1, ..., n - 1 \end{cases}$$

$$q_\nu = \begin{cases} m - 1 & \nu = 0, 1, ..., i_1 - 2 \\ \nu + i_0 & \nu = i_1 - 1, ..., n - i_0 \\ n & \nu = n - i_0 + 1, ..., n - 1 \end{cases}$$

consider the set $\{w_i(x); i = 0, 1, ..., n\}$ of functions defined as follows [17-19]

$$(3) \qquad w_i(x) = \begin{cases} l_i(x) & x \in [\tau_0, \tau_{i_1-1}], & i \leq m-1 \\ s_i(x) & x \in (\tau_{i_1-1}, \tau_{n-i_0+1}), & n \geq m \\ \bar{l}_i(x) & x \in [\tau_{n-i_0+1}, \tau_n], & i \geq n-(m-1) \end{cases}$$

where

$$l_i(x) = \prod_{\substack{k=0 \\ k \neq i}}^{m-1} \frac{x - \tau_k}{\tau_i - \tau_k}$$

$$\bar{l}_i(x) = \prod_{\substack{k=0 \\ k \neq n-i}}^{m-1} \frac{x - \tau_{n-k}}{\tau_i - \tau_{n-k}}$$

$$s_i(x) = \sum_{r=0}^{m-2} \sum_{j=j_0}^{j_1} \alpha_{i,r,j} B_{(m-1)(i+j)+r}(x)$$

with $j_0 = \max\{-i_0, i_1 - 2 - i\}$, $j_1 = \min\{-i_0 + m - 1, n - i_0 - i\}$. The coefficients $\alpha_{i,r,j}$ are given in [19] and the B-spline sequence is constructed from the set of the normalized B-splines for $i = (m-1)(i_1-2), (m-1)(i_1-2)+1, ..., (m-1)(n-i_0+1)$. Then, the following locality property holds [17]

$$(4) \qquad s_i(x) = 0 \quad , \quad x \notin [\tau_{i-i_0}, \tau_{i+i_1}].$$

Each $w_i(x)$ is nodal with respect to Π_n, in the sense that

$$w_i(\tau_j) = \delta_{i,j} \quad , \quad i, j = 0, 1, ..., n .$$

Therefore, being $\det[w_i(\tau_j)] \neq 0$, the functions $w_i(x), i = 0, 1, \dots, n$, are linearly independent. Let $\mathbf{S}_{\Pi_n} = \text{span}\{w_i(x); i = 0, 1, ..., n\}$, it is proved in [18] that, for all $s \in \mathbf{S}_{\Pi_n}$, one has $s \in \mathbf{C}^{m-2}(J)$.

For all $g \in \mathbf{B}(J)$, where $\mathbf{B}(J)$ is the set of real-valued functions on J, we consider the spline operator $W_n : \mathbf{B}(J) \to \mathbf{S}\Pi_n$, so defined

$$W_n g = \sum_{i=0}^{n} g(\tau_i) w_i(x) \quad , \quad x \in J \, .$$

By (4), for $0 \leq \nu < n$ we can write:

$$W_n g = \sum_{i=p_\nu}^{q_\nu} g(\tau_i) w_i(x), \quad x \in [\tau_\nu, \tau_{\nu+1}] \, . \tag{5}$$

Moreover $W_n p = p$, for all $p \in \mathbb{P}_m$, where $\mathbb{P}_m$ denotes the set of polynomials of order m (degree $\leq m - 1$), and $W_n g(\tau_i) = g(\tau_i)$, for $i = 0, 1, ..., n$, i.e. W_n is an interpolatory operator [17,18].

Using the results in [17-19] we deduce that, for l.u. $\{\Pi_n\}$, W_n is a bounded projection operator in S_{Π_n}. In fact, it is easy to show that

$$W_n s = s \quad , \quad \text{for all } s \in \mathbf{S}_{\Pi_n}$$

and, if we denote:

$$||W_n|| = \sup\{||W_n h||_\infty : h \in \mathbf{C}(J), ||h||_\infty < 1\},$$

with $||h||_\infty = \max\limits_{x \in I} |h(x)|$, considering that

$$||W_n|| \leq (m+1) \left[\sum_{\lambda=1}^{m-1} (R_n)^\lambda \right]^{m-1} \quad ,$$

where R_n is defined in (1), from (2), if $\{\Pi_n\}$ is l.u., we obtain $||W_n|| < \infty$.

We remark that if we assume the $(m - 2)$ points equally spaced throughout $(\tau_\nu, \tau_{\nu+1})$, $\nu = 0, 1, \dots, n - 1$, then the local uniformity constant of $\{X_n\}$ will be equal to that of $\{\Pi_n\}$.

Finally for all $g \in \mathbf{C}^{s-1}(J)$, with $1 \leq s \leq m$, we introduce the following quantity

$$E_{\nu s} = \begin{cases} D^\nu(g - W_n g) & , \quad 0 \leq \nu < s \\ D^\nu W_n g & , \quad s \leq \nu < m. \end{cases}$$

If $\{X_n\}$ is l.u., for $0 \leq \nu \leq s - 1$ there results [14,19]

$$||E_{\nu s}||_\infty = O\big(H_n^{s-\nu-1} \omega(D^{s-1} g; H_n; J)\big) \tag{6}$$

where

$$H_n = \max_{0 \leq i \leq n-1} (\tau_{i+1} - \tau_i) \tag{7}$$

and for all $f \in \mathbf{C}(J)$, $\quad \omega(f;\delta;J) = \max\limits_{\substack{x,\,x+h\,\in\,J\\ 0<h\le\delta}} |f(x+h)-f(x)|.$

For $s \le \nu < m$ in [14] a bound for $|E_{\nu s}|$ is given.

Furthermore, for every $t \subseteq J$ and $g \in \mathbf{C}^{\nu}(J)$, $0 \le \nu < m-1$ [27]

$$\omega(D^{\nu} W_n g; t; J) = O\big(\omega(D^{\nu} g; t; J)\big).$$

2.2. Tensor product of optimal nodal splines

Let D be the $\mathbb{R}^2$ subset defined by $[a,b] \times [\tilde{a},\tilde{b}]$. We consider partitions Π_n and X_n on which we construct the spline functions of order m $\{w_i(x), i = 0,\dots,n\}$ defined in (3).

Then we consider similar partitions of $[\tilde{a},\tilde{b}]$, $\tilde{\Pi}_{\tilde{n}}$ and $\tilde{X}_{\tilde{n}}$ and we construct the corresponding functions of order $\tilde{m}$ $\{\tilde{w}_{\tilde{i}}(\tilde{x}), \tilde{i} = 0,\dots,\tilde{n}\}$.

Now we may generate a set of bivariate splines

$$w_{i,\tilde{i}}(x,\tilde{x}) = w_i(x) w_{\tilde{i}}(\tilde{x})$$

tensor product of the (3) ones.

Let $\mathbf{B}(D)$ denote the set of bounded real-valued functions on D. Then, for any $f \in \mathbf{B}(D)$ we may define the following spline interpolating operator for $(x,\tilde{x}) \in [\tau_j, \tau_{j+1}] \times [\tilde{\tau}_{\tilde{j}}, \tilde{\tau}_{\tilde{j}+1}]$,

$$W^*_{n\tilde{n}} f(x,\tilde{x}) = \sum_{i=p_j}^{q_j} \sum_{\tilde{i}=\tilde{p}_{\tilde{j}}}^{\tilde{q}_j} w_{i,\tilde{i}}(x,\tilde{x}) f(\tau_i, \tilde{\tau}_{\tilde{i}}), \tag{8}$$

with $j = 0,1,\dots,n-1$ and $\tilde{j} = 0,1,\dots,\tilde{n}-1$.

In order to obtain the maximal order polynomial reproduction, we can assume $m = \tilde{m}$, i.e. we use splines of the same order on both axes. We list in the following the main properties of $W^*_{n\tilde{n}}$.

(a) $W^*_{n\tilde{n}}$ is local, in the sense that $W^*_{n\tilde{n}} f(x,\tilde{x})$ depends only on the values of f in a small neighbourhood of $(x,\tilde{x})$;

(b) $W^*_{n\tilde{n}}$ interpolates f at the primary knots, i.e. $W^*_{n\tilde{n}} f(\tau_i, \tilde{\tau}_{\tilde{i}}) = f(\tau_i, \tilde{\tau}_{\tilde{i}})$;

(c) $W^*_{n\tilde{n}}$ has the optimal order polynomial reproduction property, that means $W^*_{n\tilde{n}} p = p$, for all $p \in \mathbb{P}^2_m$, where $\mathbb{P}^2_m$ is the set of bivariate polynomials of total order m.

For $f \in \mathbf{C}^{s-1}(D)$, $1 \le s < m$ we introduce the following quantity

$$E_{\nu\tilde{\nu}s} = \begin{cases} D^{\nu,\tilde{\nu}}(f - W^*_{n\tilde{n}} f) & \text{if} \quad 0 \le \nu + \tilde{\nu} < s \\ D^{\nu,\tilde{\nu}} W^*_{n\tilde{n}} f & \text{if} \quad s \le \nu + \tilde{\nu} < m \end{cases}$$

where $D^{\nu,\tilde{\nu}}$ is the usual partial derivative operator.

Now we say that a collection of product partitions $\{X_n \times \tilde{X}_{\tilde{n}}\}$ of D is quasi uniform (q.u.) if there exists a positive constant σ such that

$$\frac{\Delta}{\hat{\delta}}, \frac{\Delta}{\tilde{\delta}}, \frac{\tilde{\Delta}}{\hat{\delta}}, \frac{\tilde{\Delta}}{\tilde{\delta}} \leq \sigma ,$$

where $\Delta = \max_{1\leq i\leq n(m-1)}(x_i - x_{i-1})$, $\hat{\delta} = \min_{1\leq i\leq n(m-1)}(x_i - x_{i-1})$ and $\tilde{\Delta} = \max_{1\leq \tilde{i}\leq \tilde{n}(m-1)}(\tilde{x}_{\tilde{i}} - \tilde{x}_{\tilde{i}-1})$, $\tilde{\delta} = \min_{1\leq \tilde{i}\leq \tilde{n}(m-1)}(\tilde{x}_{\tilde{i}} - \tilde{x}_{\tilde{i}-1})$.

We set

$$H^* = H_n + \tilde{H}_{\tilde{n}} \qquad \text{and} \qquad \Delta^* = \Delta + \tilde{\Delta} \tag{9}$$

where H_n is defined in (7) and likewise $\tilde{H}_{\tilde{n}}$.

Assuming that $f \in \mathbf{C}^{s-1}(D)$ with $1 \leq s < m$ and that $\{W_{n\tilde{n}} f\}$ is a q.u. sequence of nodal splines, then for $\nu, \tilde{\nu}$ such that $0 \leq \nu + \tilde{\nu} \leq s - 1$

$$||E_{\nu\tilde{\nu}s}||_\infty = O\big(H^{*s-\nu-\tilde{\nu}-1}\omega(D^{s-1} f; H^*; D)\big).$$

In [9] local bounds of $|E_{\nu\tilde{\nu}s}|$ are derived and local and global bounds of $|E_{\nu\tilde{\nu}s}|$, $s \leq \nu + \tilde{\nu} < m$, are also given.

Furthermore, for $f \in \mathbf{C}^p(D)$, $0 \leq p < m - 1$, and for a q.u. sequence of nodal splines $\{W^*_{n\tilde{n}} f\}$, there results for any non empty subset T of D

$$\omega(D^p W^*_{n\tilde{n}} f; T; D) = O\big(\omega(D^p f; T; D)\big) .$$

In the following we shall consider l.u. partitions in the one dimensional case and q.u. partitions in the 2D one and we shall suppose always that the norm of the partitions converges to zero as $n \to \infty$ or $n, \tilde{n} \to \infty$.

3. Numerical integration based on nodal spline operators

This section will deal with the numerical evaluation of some singular one-dimensional integrals and of certain 2D singular integrals.

3.1. Product integration of singular integrands

Consider integrals of the form

$$J(kf) = \int_I k(x) f(x) dx \tag{10}$$

where $kf \in \mathbf{L}_1(I)$, but f is unbounded in $I = [-1, 1]$.

In [26] product integration have been proposed, by substituting f by a sequence of interpolatory nodal splines $\{W_n f\}$ defined in (5), under different hypotheses on f.

By using (6) with $\nu = 0$, the author gets, firstly, the convergence of the quadrature sum $J(kW_n f)$, i.e.:

$$J(kW_n f) \to J(kf) \text{ as } \quad n \to \infty \tag{11}$$

by supposing $f \in \mathbf{C}(I)$, $k \in \mathbf{L}_1(I)$ and $H_n \to 0$ as $n \to \infty$.

We recall that a computational procedure to generate the weights $\{v_i(k) = \int_I k(x)w_i(x)dx\}$ of the above quadrature is given in [6].

Moreover in [26] the case when $f \in \mathbf{P}C(I)$, $k \in L_1(I)$ is studied and the convergence of the quadrature rules sequence is proved.

We remark that in [11] the convergence (11) has been proved also for $f \in \mathbf{R}(I)$, the class of Riemann integrable functions on I and $k \in \mathbf{L}_1(I)$.

When the function f in (10) is singular in $z \in [-1, 1)$ in [25] the author defines the family of real valued functions $M_d(z; k)$:

$$\begin{array}{l} M_d(z;k) = \{f : f \in \mathbf{P}C(z,1], \exists F : F = 0 \\ \text{ on } [-1,z],\ F \text{ is non negative, continuous} \\ \text{and nonincreasing on } (z,1),\ kF \in \mathbb{L}_1(I) \text{ and } |f| \le F \text{ on } I\} \end{array} . \tag{12}$$

He supposes that k satisfies one of the following conditions A, B:

(A) There exists $\delta > 0 : |k(x)| \le K(x)$, $\forall x \in (z, z+\delta]$, K is positive nonincreasing in that interval and KF, F defined in (12), is a $\mathbf{L}_1$ function in I.

(B) Given $q_0 \in (0, 1)$, $\exists \delta, T$, positive numbers (possibly depending on q_0), such that

$$\int_c^{c+h} |k(x)|dx \le hT|k(c+qh)|$$

$\forall q \in [q_0, 1]$, $\forall c$ and h satisfying $z \le c < c+h \le z+\delta$. Besides $|k(x)f(x)| \le G(x)$, $\forall x \in (z, z+\delta]$, where G is a positive non increasing $\mathbf{L}_1$ function in that interval.

The following theorem can be proved.

THEOREM 1. *Assume that $f \in M_d(-1; k)$ and k satisfies (A) or (B). If the sequence of partitions $\{\Pi_n\}$ is l.u. and the norm converges to zero as $n \to \infty$, then (11) holds.*

As consequence of that theorem if $z = -1$ the singularity can be ignored, provided k satisfies (A) or (B).

In the case when z is an interior singularity, it must, in general, be avoided, i.e. we must define a new integration rule

$$J^*(kW_n f) = \sum_{i=J}^{n} v_i(k) f(\tau_i)$$

where J is the smallest integer such that $z \leq \tau_{J-\lambda}$, where $\tau_{J-\lambda}$ is the left bound of the support of $s_J(x)$ and, if we assume that n is so large that $J \geq m$, then $w_i = s_i$ and $v_i(k)$ is given by:

$$v_i(k) = \int_{\tau_{i-\lambda}}^{\tau_{i+\mu}} k(x)s_i(x)dx \ ,$$

with $\lambda = i_0$ and $\mu = i_1$.

Therefore, assuming that $f \in M_d(z;k)$, $z > -1$, and k satisfying (A) or (B). If $\{\Pi_n\}$ is locally uniform and the norm tends to zero as $n \to \infty$, then

$$J^*(kW_n f) \to J(Kf) \quad \text{as} \quad n \to \infty \, .$$

If one wishes to use $J(kW_n f)$ rather than $J^*(kW_n f)$ then k must be restricted in $[-1, z)$ as well as in $(z, 1]$, for satisfying one of the following conditions $(\hat{A})$ or $(\hat{B})$.

($\hat{\mathbf{A}}$) : (A) holds and, in addition, $|k_z(x)| \leq K(x)$ in $(z, z+\delta]$, where $k_z \in L_1(2z-1, 2z+1)$ is defined by $k_z(z+y) = k(z-y)$.

($\hat{\mathbf{B}}$) : (B) holds and so does (B) with k replaced by k_z.

THEOREM 2. *Let $f \in M_d(z;k)$, $z > -1$. Assume that k satsfies $(\hat{A})$ or $(\hat{B})$ and that $\{\Pi_n\}$ is l.u. and the norm converges to zero as $n \to \infty$.*

Define

$$\hat{J}(kW_n f) = J(kW_n f) - v_\rho f(\tau_\rho)$$

where τ_ρ is the value of $\tau_i \geq z$ closest to z. Then

$$\hat{J}(kW_n f) \to J(kf) \quad \text{as} \quad n \to \infty \, .$$

In particular, if $\tau_\rho = z$ then (11) holds. If z is such that for all n, $\tau_\rho - z > C(\tau_\rho - \tau_{\rho-1})$, then (11) holds.

3.2. Cauchy principal value integrals

Consider the numerical evaluation of the Cauchy principal value (CPV) integrals

$$J(kf;\lambda) = ⨍_{-1}^{1} k(x)\frac{f(x)}{x-\lambda}dx, \quad \lambda \in (-1,1). \tag{13}$$

In [11] the problem has been investigated, following the “subtracting singularity” approach.

Assuming that $J(k;\lambda)$ exists for $\lambda \in (-1,1)$, the integral (13) can be written in the form

$$\begin{aligned} J(kf;\lambda) &= \int_{-1}^{1} k(x)g_\lambda(x)dx + f(\lambda)J(k;\lambda) \\ &= \mathcal{I}(kg_\lambda) + f(\lambda)J(k;\lambda), \end{aligned}$$

where

$$g_\lambda(x) = g(x;\lambda) = \begin{cases} \frac{f(x)-f(\lambda)}{x-\lambda} & x \neq \lambda \\ f'(\lambda) & x = \lambda \text{ and } f'(\lambda) \text{ exists} \\ 0 & \text{otherwise}. \end{cases}$$

Therefore, approximating $\mathcal{I}(kg_\lambda)$ by $\mathcal{I}(kW_n g_\lambda)$ we can write [11]

$$J(kf;\lambda) = J_n(kf;\lambda) + E_n(kf;\lambda),$$

where

$$J_n(kf;\lambda) = \mathcal{I}(kW_n g_\lambda) + f(\lambda)J(k;\lambda).$$

For any $\lambda \in (-1, 1)$ we define a family of functions $\bar{M}_d(z;k) = \{g \in C(I\backslash\lambda), \exists G : G$ is continuous nondecreasing in $[-1;\lambda)$, continuous non increasing in $(\lambda, 1]; kG \in \mathbf{L}_1(I), |g| < G$ in $I\}$.

We assume

$$N_\delta(\lambda) = \{x : \lambda - \delta \leq x \leq \lambda + \delta\},$$

where $\delta > 0$ is such that $N_\delta(\lambda) \subset I$.

We denote by $\mathbf{H}_\mu(I)$, $\mu \in (0, 1]$, the set of Hölder continuous functions

$$\begin{aligned} \mathbf{H}_\mu(I) &= \{g \in C(I) : |g(x_1) - g(x_2)| \\ &\leq L|x_1 - x_2|^\mu, \forall x_1, x_2 \in I, L > 0\} \end{aligned}$$

and by $\mathbf{D}T(I)$ the set of Dini type functions

$$\mathbf{D}T(I) = \{g \in \mathbf{C}(I) : \int_0^{l(I)} \omega(g;t)t^{-1}dt < \infty\}$$

where $l(I)$ is the length of I and ω denotes the usual modulus of continuity.

The following convergence results for the quadrature rules $J_n(kf;\lambda)$, under different hypotheses for the function f, are derived in [11].

THEOREM 3. *For any $\lambda \in (-1, 1)$, let $f \in \mathbf{H}_1\big(N_\delta(\lambda) \cap \mathbf{R}(I)\big)$ and $k \in \mathbf{L}_1(I)$. Then, for l.u. $\{\Pi_n\}$, $E_n(kf;\lambda) \to 0$ as $n \to \infty$.*

THEOREM 4. *Let $f \in \mathbf{H}_\mu(I)$, $0 < \mu < 1$, $k \in \mathbf{L}_1(I) \cap \mathbf{C}\big(N_\delta(\lambda)\big)$. Let h and p be the greatest and the smallest integers such that $\tau_h < \lambda$, $\tau_p > \lambda$. We denote by τ^* the node closest to λ*

$$\tau^* = \begin{cases} \tau_h & \text{if} \quad \lambda - \tau_h \leq \tau_p - \lambda \\ \tau_p & \text{if} \quad \lambda - \tau_h > \tau_p - \lambda \end{cases}$$

and we suppose that there exists some positive constant C, such that

$$|\tau^* - \lambda| > C \max\{(\tau_h - \tau_{h-1}), (\tau_{p+1} - \tau_p)\},$$

then, for l.u. $\{\Pi_n\}$,

$$E_n(kf;\lambda) \to 0$$

as $n \to \infty$.

THEOREM 5. *Let* $f \in \mathbf{C}^1(I)$, $k \in \mathbf{L}_1(I)$. *Then*

$$E_n(kf;\lambda) \to 0 \text{ uniformly in } \lambda, \text{ as } n \to \infty.$$

However, if $k \in \mathbf{L}_1(I) \cap \mathbf{D}T(-1,1)$, *then* $J(kf;\lambda)$ *exists for all* $\lambda(-1,1)$. *Besides*

$$J_n(kf;\lambda) \to J(kf;\lambda) \text{ as } n \to \infty$$

uniformly for all $\lambda \in (-1,1)$.

Moreover in [14] it has been proved that $J(\omega_{\alpha,\beta}W_n;\lambda) \to J(kf;\lambda)$ uniformly with respect to $\lambda \in (-1,1)$, for $\omega_{\alpha,\beta}(x) = (1-x)^\alpha(1+x)^\beta$, $\alpha,\beta > -1$, and $f(x) \in \mathbf{H}_\rho(-1,1)$, $0 < \rho \leq 1$.

3.3. The Hadamard finite part integrals

We consider the evaluation of the finite part integrals of the form

$$\bar{J}(\omega_{\alpha,\beta}f) = \mathop{=\!\!\!\!\!\!\int}_I \frac{\omega_{\alpha,\beta}(x)f(x)}{x+1}dx, \tag{14}$$

where $\alpha > -1$, $-1 < \beta \leq 0$ and $=\!\!\!\!\!\!\int$ denotes the Hadamard finite part (HFP).

It is well known that a sufficient condition so that (14) exists is

$$f \in \mathbf{H}_\mu(I), \quad 0 < \mu \leq 1, \quad \mu + \beta > 0\,.$$

We recall that [25]

$$\bar{J}(\omega_{\alpha,\beta}f) = \int_{-1}^{1} \omega_{\alpha,\beta}(x)\frac{f(x)-f(-1)}{x+1}dx + f(-1)\mathop{=\!\!\!\!\!\!\int}_{-1}^{1} \frac{\omega_{\alpha,\beta}(x)}{x+1}dx, \tag{15}$$

where, denoting $c_j = \frac{d^j}{dx^j}\frac{(1-x)^j}{j!}\Big|_{x=-1}$, $j = 0,1,\ldots$, we obtain for the HFP in (15),

$$\mathop{=\!\!\!\!\!\!\int}_{-1}^{1} \frac{\omega_{\alpha,\beta}(x)}{x+1}dx = \begin{cases} \log 2 & \text{if } \alpha=\beta=0 \\ c_0\log 2 + \sum_{j=1}^{\infty}\frac{c_j}{j!}2^j & \text{if } \beta=0, \alpha\neq 0 \\ \frac{\alpha+\beta+1}{\beta}2^{\alpha+\beta}\frac{\Gamma(\alpha+1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta+2)} & \text{if } \alpha>-1,\ -1<\beta<0, \end{cases}$$

where Γ is the gamma function.

Approximating f by $W_n f$ in (14) we obtain the quadrature rule [5]:

$$\bar{J}(\omega_{\alpha,\beta}f) = \bar{J}_n(f) + \bar{E}_n(f), \tag{16}$$

where

$$\bar{J}_n(f) = \sum_{i=0}^{n} \bar{v}_i(\omega_{\alpha,\beta}) f(\tau_i)$$

with $\bar{v}_i(\omega_{\alpha,\beta}) = \bar{J}(\omega_{\alpha,\beta} w_i)$, and

$$\bar{E}_n(f) = \bar{J}(\omega_{\alpha,\beta}(f - W_n f)).$$

A computational procedure for evaluating $\bar{v}_i(\omega_{\alpha,\beta})$ is given in [6].

Denoting by $\mathbf{H}^s_\mu(I)$ the set of the functions $f \in \mathbf{C}^s(I)$ having $f^{(s)} \in \mathbf{H}_\mu(I)$, in [5] the following theorem has been proved.

THEOREM 6. *Let $f \in \mathbf{H}^s_\mu(I)$, $0 \le s \le m-1$, and $\mu + \beta > 0$ if $s = 0$. Then, as $n \to \infty$:*

$$\|\bar{E}_n(f)\|_\infty = \begin{cases} O(H_n^{s+\mu+\beta}) & \text{if } \beta < 0 \\ O(H_n^{s+\mu}|\log H_n|) & \text{if } \beta = 0\,. \end{cases}$$

Consider now HFP integrals of the form:

$$J^*(\omega_{\alpha,\beta} f; \lambda; p) = \oint_I \omega_{\alpha,\beta}(x) \frac{f(x)}{(x-\lambda)^{p+1}}, \quad \lambda \in [-1,1], \quad p \ge 1 \tag{17}$$

If $f \in \mathbf{H}^p_\mu(I)$, then $J^*(\omega_{\alpha,\beta} f; \lambda; p)$ exists.

In [20, 21] quadrature rules for the numerical evaluation of (17), based on some different type of spline approximation, including the optimal nodal splines, are considered and studied.

In [29] the following theorem has been proved.

THEOREM 7. *Assume that in (17) $\lambda \in (-1,1)$, $p \in \mathbf{N}$ and $f \in H^p_\mu$. Let $\{f_n\}$ be a given sequence of functions such that $f_n \in \mathbf{C}^p(I)$ and*

i) - $\|D^j r_n\|_\infty = o(1)$ as $n \to \infty$ $\quad j = 0, 1, \ldots, p$, *where* $r_n = f - f_n$

ii) - $D^j r_n(-1) = 0 \quad 0 \le j \le p - \beta; \qquad D^j r_n(1) = 0 \quad 0 \le j \le p - \alpha$

iii) - $r_n \in \mathbf{H}^p_\sigma(I), \quad \forall n, \quad 0 < \sigma \le \mu, \quad \sigma + \min(\alpha,\beta) > 0.$

Then

$$J^*(\omega_{\alpha,\beta} f_n; \lambda; p) \to J^*(\omega_{\alpha,\beta} f; \lambda; p) \quad \text{as } n \to \infty \tag{18}$$

uniformly for $\forall \lambda \in (-1,1)$.

If we consider a sequence of optimal nodal splines for approximating the function f, in order to obtain the uniform convergence in (18) of integration rules, we must modify the sequence $\{W_n\}$ in the sequence $\{\hat{W}_n f\}$, for which condition *ii)* is satified.

Therefore, in [15], for $0 \leq s, t \leq p$, are defined two sets of B-splines $\bar{B}_i, \bar{B}_{N-i}$ on the knot sets

$$\{x_0, \ldots x_0, x_1, \ldots, x_{s+1}\}, \quad \{x_{N-t-1}, \ldots, x_{N-1}, x_N, \ldots, x_N\}$$

respectively, where $N = (m-1)n$ and x_0, x_N are repeated exactly m times.

Considering that $W_n f(\tau_i) = f(\tau_i)$, $i = 0, n$, one defines

$$g_n(x) := \begin{cases} \sum_{i=1}^{s} d_i \bar{B}_i(x) & x \in [x_0, \ldots, x_{s+1}] \\ 0 & x \in (x_{s+1}, \ldots, x_{N-t-1}) \\ \sum_{i=1}^{t} \tilde{d}_i \bar{B}_{N-i}(x) & x \in [x_{N-t-1}, \ldots, x_N] \end{cases}$$

where $d_i, \tilde{d}_i$ are determined by solving two non-singular triangular systems obtained by imposing

$$\begin{aligned} g^{(j)}(\tau_0) &= r_n^{(j)}(\tau_0) \quad j = 1, 2, \ldots, s \\ g_n^{(j)}(\tau_n) &= r_n^{(s)}(\tau_n) \quad j = 1, 2, \ldots, t \end{aligned}$$

For the sequence $\{\hat{W}_n f = W_n f + g_n\}$, it is possible to prove the following:

THEOREM 8. *Let* $\{\hat{W}_n f\}$ *be a sequence of modified optimal nodal splines and set* $\hat{r}_n = f - \hat{W}_n f$, *then*
$\hat{W}_n f(\tau_i) = f(\tau_i) \quad i = 0, \ldots, n$;
$D^j \hat{r}_n(-1) = 0$, $0 \leq j \leq p - \beta$; $D^j \hat{r}_n(1) = 0$, $0 \leq j \leq p - \alpha$,
$\hat{W}_n g = g$ *if* $g \in \mathbb{P}_m$.
Besides supposing $f \in \mathbf{C}^r(I_k)$, $I_k = [\tau_k, \tau_{k+1}]$, $h_k = \tau_{k+1} - \tau_k$, *for any* $x \in I_k$ *there results:*
$|D^\nu \hat{r}_n(x)| \leq \tilde{k}_\nu h_k^{r-\nu} \omega(D^r f; h_k; I_k), \quad \nu = 0, \ldots, r$
$|D^{r+1} \hat{W}_n f(x)| \leq \tilde{k}_{r+1} h_k^{-1} \omega(D^r f; h_k; I_k)$,
$\hat{r}_n \in \mathbf{H}_\mu^r(I)$.

Therefore all the conditions of theorem 3.3.2 being satisfied, if $\mu + \min(\alpha, \beta) > 0$, then

$$J^*(\omega_{\alpha,\beta} \hat{W}_n f; \lambda; p) \to J(\omega_{\alpha,\beta} f; \lambda; p) \quad \text{as } n \to \infty$$

uniformly for $\forall \lambda \in (-1, 1)$.

3.4. Integration rules for 2-D CPV integrals

In this section we will consider the numerical evaluation of the following two types of CPV integrals:

$$J_1(f; x_0, y_0) = \fint_R \omega_1(x)\omega_2(y) \frac{f(x,y)}{(x-x_0)(y-y_0)} dx dy \tag{19}$$

where $R = [a, b] \times [\tilde{a}, \tilde{b}]$, $x_0 \in (a, b)$, $y_0 \in (\tilde{a}, \tilde{b})$, and we assume $\omega_1(x) \in \mathbf{L}_1[a, b] \cap \mathbf{D}T(N_\delta(x_0))$, $\omega_2(y) \in \mathbf{L}_1[\tilde{a}, \tilde{b}] \cap \mathbf{D}T(N_\delta(y_0))$; and

$$(20) \qquad J_2(\phi; P_0) = \fint_D \Phi(P_0, P)dP, \quad P_0 \in D$$

where D denotes a polygonal region and $\Phi(P_0, P)$ is an integrable function on D except at the point P_0 where it has a second order pole.

For numerically evaluating (19), in [9] the following cubatures based on a sequence of nodal splines (8) have been proposed:

$$J_1(W_{n\tilde{n}} f; x_0, y_0) = \sum_{i=0}^{n} \sum_{\tilde{i}=0}^{\tilde{n}} v_i(x_0)\tilde{v}_{\tilde{i}}(y_0) f(\tau_i, \tilde{\tau}_{\tilde{i}}) ,$$

where $v_i(x_0) = \fint_a^b \omega_1(x) \dfrac{w_i(x)}{x - x_0} dx$, and $\tilde{v}_{\tilde{i}}(y_0) = \fint_{\tilde{a}}^{\tilde{b}} \omega_2(y) \dfrac{\tilde{w}_{\tilde{i}}(y)}{y - y_0} dy$.

We denote by $\mathbf{H}^p_{\mu,\mu}(R)$ the set of continuous functions having all partial derivatives of order $j = 0, \ldots, p$, $p \geq 0$ continuous and each derivative of order p satisfying a Hölder condition, i.e.:

$$|f^{(p)}(x_1, y_1) - f^{(p)}(x_2, y_2)| \leq C(|x_1 - x_2|^\mu + |y_1 - y_2|^\mu), \quad 0 < \mu \leq 1$$

for some constant $C > 0$, and we assume

$$(21) \qquad E_{n\tilde{n}}(f; x_0, y_0) = J_1(f; x_0, y_0) - J_1(W_{n\tilde{n}} f; x_0, y_0).$$

In [9] the following convergence theorem has been proved.

THEOREM 9. *Let $f \in \mathbf{H}^p_{\mu,\mu}$, $0 < \mu \leq 1$, $0 \leq p < m - 1$. For the remainder term in (21), there results:*

$$E_{n\tilde{n}}(f; x_0, y_0) = O\big((\Delta^*)^{p+\mu-\gamma}\big),$$

where $\gamma \in \mathbb{R}$, $0 < \gamma < \mu$, small as we like and Δ^* has been defined in (9).

In many practical applications it is necessary that rules, uniformly converging for $\forall (x_0, y_0) \in (-1, 1) \times (-1, 1)$, are available, in particular considering the Jacobi weight type functions

$$\omega_1(x) = (1 - x)^{\alpha_1}(1 + x)^{\beta_1}, \quad \omega_2(y) = (1 - y)^{\alpha_2}(1 + y)^{\beta_2}$$

with $\alpha_i, \beta_i > -1$, $i = 1, 2$, $(x, y) \in R = [-1, 1] \times [-1, 1]$.

In order to obtain uniform convergence for approximating rules numerically evaluating (19), can be useful to write the integral in the form

$$\begin{aligned} J_1(f; x_0, y_0) &= \fint_R \omega_1(x)\omega_2(y) \frac{f(x, y) - f(x_0, y_0)}{(x - x_0)(y - y_0)} dx dy \\ (22) \qquad &+ f(x_0 y_0) J(\omega_1; x_0) J(\omega_2; y_0) \end{aligned}$$

where $J(\omega_1; x_0) = \fint_{-1}^{1} \frac{\omega_1(x)}{x - x_0} dx, \quad J(\omega_2; y_0) = \fint_{-1}^{1} \frac{\omega_2(y)}{y - y_0} dy.$

We exploit the results in [31] where, considering a sequence of linear operators $F_{n\tilde{n}}$ approximating f, the integration rule for (22):

$$\begin{aligned} J_1(F_{n\tilde{n}}; x_0, y_0) &= \fint_R \omega_1(x)\omega_2(y) \frac{F_{n\tilde{n}}(x, y) - F_{n\tilde{n}}(x_0, y_0)}{(x - x_0)(y - y_0)} dx dy \\ &+ f(x_0, y_0) J(\omega_1; x_0) J(\omega_2; y_0) \end{aligned}$$

has been constructed. Denoting $r_{n\tilde{n}} = f - F_{n\tilde{n}}$, and $\Delta_{n\tilde{n}}$ the norm of the partition, with $\lim_{\substack{n \to \infty \\ \tilde{n} \to \infty}} \Delta_{n\tilde{n}} = 0$, the following general theorem of uniform convergence has been proved.

THEOREM 10. *Let $f \in H^0_{\mu\mu}(R)$, and assume that the approximation $F_{n\tilde{n}}$ to f is such that*

i) $r_{n\tilde{n}}(x, \pm 1) = 0 \quad \forall x \in [-1, 1], r_{n\tilde{n}}(\pm 1, y) = 0 \quad \forall y \in [-1, 1],$

ii) $||r_{n\tilde{n}}||_\infty = O(\Delta^\nu_{n\tilde{n}}), \quad 0 < \nu \le \mu,$

iii) $r_{n\tilde{n}} \in H^0_\sigma(R), \quad 0 < \sigma \le \mu.$

If $\rho + \gamma - \bar{\varepsilon} > 0$, where $\rho = \min(\sigma, \nu)$, $\gamma = \min(\alpha_1, \alpha_2, \beta_1, \beta_2)$ and $\bar{\varepsilon}$ is a positive real number as small as we like, then, for the remainder term, $E_{n\tilde{n}} = J_1(f; x_0, y_0) - J_1(F_{n\tilde{n}}; x_0, y_0)$, there results:

$$E_{n\tilde{n}}(f; x_0.y_0) \to 0 \quad \text{as} \quad n \to \infty, \tilde{n} \to \infty$$

uniformly for $\forall (x_0, y_0) \in (-1, 1) \times (-1, 1)$.

If we consider $F_{n\tilde{n}} = W_{n\tilde{n}}(f; x, y)$ only the conditions *ii), iii)*, with $\Delta_{n,\tilde{n}} = \Delta^*$, are satisfied, but we can modify $W_{n\tilde{n}}$ in the form

$$\begin{aligned} \bar{W}_{n\tilde{n}}(f; x, y) &= W_{n\tilde{n}}(f; x, y) + [f(-1, y) - W_{n\tilde{n}}(f; -1, y)] B_{1-m}(x) \\ &+ [f(1, y) - W_{n\tilde{n}}(f; 1, y)] B_{(m-1)n-1}(x) \\ &+ [f(x, -1) - W_{n\tilde{n}}(f; x, -1)] \tilde{B}_{1-m}(y) \\ &+ [f(x, 1) - W_{n\tilde{n}}(f; x, 1)] B_{(m-1)\tilde{n}-1}(y) \, . \end{aligned}$$

Assuming $\bar{r}_{n\tilde{n}}(x, y) = f(x, y) - \bar{W}_{n\tilde{n}}(f; x, y)$, all the condition $i) - iii)$ are verified and then

$$J_1(\bar{W}_{n\tilde{n}}; x_0, y_0) \to J_1(f; x_0, y_0) \quad \text{as } n, \tilde{n} \to \infty$$

uniformly for $\forall (x_0, y_0) \in (-1, 1) \times (-1, 1)$.

Now we consider the integral (20) for which we refer to the results in [5,6]. Since the polygon D can be thought as the union of triangles, each one with the singularity at one vertex, by introducing polar coordinates (r, ϑ) with origin at the singularity P_0, the evaluation of (20) can be reduced to the evaluation of

$$J_2^*(f) = \int_{\vartheta_1}^{\vartheta_2} \left(\fint_0^{R(\vartheta)} \frac{f(r,\vartheta)}{r} dr \right) d\vartheta, \tag{23}$$

where

$$\fint_0^{R(\vartheta)} \frac{f(r,\vartheta)}{r} d\vartheta = \int_0^{R(\vartheta)} \frac{f(r,\vartheta) - f(0,\vartheta)}{r} dr + f(0,\vartheta)\log(R(\vartheta));$$

the integration domain is a triangle (Fig. 1)

$$T = \{(r,\vartheta) : 0 \le r \le R(\vartheta), \quad \vartheta_1 \le \vartheta \le \vartheta_2\}$$

with

$$R(\vartheta) = \begin{cases} \frac{d}{\sin\vartheta - \cos\vartheta} & \text{if } s : y = cx + d \\ \frac{d}{\cos\vartheta} & \text{if } s : x = d\,. \end{cases}$$

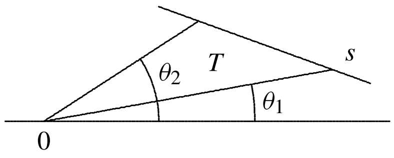


Figure 1. Domain of integration T.

The outer integral in (23) will be approximated by rules of the form considered in section 3.1 with nodes $\Pi_n = \{\tau_i\}_{i=0}^n$ and weights $\{v_i\}_{i=0}^n$; for the inner one we consider rules of the form (16), with $\alpha = \beta = 0$, based on optimal nodal splines of order $\bar{m} \ge 3$, primary knots $\bar{\Pi}_N = \{\bar{\tau}_i = \bar{y}_{(\bar{m}-1)i}\}_{i=0,\dots,N}$ corresponding to the partition

$$\bar{Y}_N = \{-1 = \bar{y}_0 < \bar{y}_1 \cdots < \bar{y}_{(\bar{m}-1)N} = 1\}$$

and we suppose that the norms H_n and $\bar{H}_N$, of Π_n and $\bar{\Pi}_N$, respectively, converges to 0 as n and $N \to \infty$.

We obtain the following rules

$$J_{2,n,N}^*(f) = \frac{\vartheta_2 - \vartheta_1}{2} \sum_{i=0}^n v_{in} \left[\sum_{k=0}^N \bar{v}_{kN} f(r_{ki}, \xi_i) + f(0,\xi_i) \log\left(\frac{R(\xi_i)}{2}\right) \right] + R_{n,N}(f),$$

where

$$\begin{cases} \xi_i = [(\vartheta_2 - \vartheta_1)/2]\tau_i + (\vartheta_2 + \vartheta_1)/2 & i = 0,\dots,n \\ r_{ki} = [R(\xi_i)/2](\bar{\tau}_{kN}) + [R(\xi_2)/2](\bar{\tau}_{kN} + 1) & i = 0,\dots,N. \end{cases}$$

Let us assume $R = \max_{\vartheta\in[\vartheta_1,\vartheta_2]} |R(\vartheta)|$, $\mathcal{R} = [0, R] \times [\vartheta_1, \vartheta_2]$ and define $m^* = \min(m, \bar{m})$.

We can prove the following theorem:

THEOREM 11. *If* $f \in \mathbf{H}^s_{\mu,\mu}(\mathcal{R})$, $0 < \mu \leq 1$ *and* $0 \leq s \leq m^* - 1$, $\{\Pi_n\}$ *and* $\{\bar{Y}_N\}$ *are sequence of locally uniform partitions, then*

$$||R_{n,N}(f)||_\infty = O(\bar{H}_N^{s+\mu} |\log(\bar{H}_N)| + H_n^{s+\mu-\varepsilon})$$

where ε *is a positive real as small as we like.*

4. A collocation method for weakly singular Volterra equations

Consider the Volterra integral equation of the second kind

$$y(x) = f(x) + \int_0^x k(x, s)y(s)ds \quad x \in I \equiv [0, X] \tag{24}$$

where k is weakly singular kernel, in particular of convolution type of the form $k(x-s)$, where $k \in \mathbf{C}(0, X] \cap \mathbf{L}_1(0, X)$, but $k(t)$ can become unbounded as $t \to 0$.

In [8], for numerically solving (24) a product collocation method, based on optimal nodal splines, has been constructed, for which error analysis and condition number are given.

If we consider a spline $y_n \in \mathbf{S}_{\pi_n}$, written in the form

$$y_n(x) = \sum_{j=0}^{n} \alpha_j w_j(x) \quad \alpha_j \in \mathbb{R}, \quad j = 0, \ldots, n,$$

and we substitute such function in (24), we obtain

$$y_n(x) - \int_0^x k(x, s)y_n(s)ds + r_n(x) = f(x)$$

where $r_n(x)$ is the residual term obtained in approximating y by y_n.

The values α_j are determined by imposing

$$r_n(\tau_j) = 0 \quad j = 0, \ldots, n, \tag{25}$$

i.e. as solution of a linear system of the form

$$\alpha_j[1 - \mu(\tau_j)] - \sum_{\substack{i=0 \\ i \neq j}}^{n} \mu_i(\tau_j)\alpha_i = f(\tau_j) \quad j = 0, \ldots, n,$$

where $\mu_i(\tau_j) = \displaystyle\int_0^{\tau_j} k(\tau_j, s)w_i(s)ds.$

In the quoted paper the explicit form of $\mu_i(\tau_j)$ for different values of i is provided.

Exploiting the properties of the operator W_n, which is a bounded interpolating projection operator, the condition (25) can be rewritten in the form

$$(I - W_n \tilde{K}) y_n = W_n f, \tag{26}$$

where $\tilde{K} y = \int_I \tilde{k}(x,s) y(s) ds$, with

$$\tilde{k}(x,s) = \begin{cases} k(x,s) & 0 \le s \le x \\ 0 & s > x, \end{cases}$$

is a bounded compact operator on $\mathbf{C}(I)$ [1]. Therefore we can deduce that equation (26) has a unique solution and

THEOREM 12. *For all n sufficiently large, say $n \ge N$, the operator $(I - W_n \tilde{K})^{-1}$ from $\mathbf{C}(I)$ to $\mathbf{C}(I)$ exists.*

Moreover it is uniformly bounded, i.e.:

$$\sup_{n \ge N} ||(I - W_n \tilde{K})^{-1}|| \le M < \infty$$

and

$$||y - y_n||_\infty \le ||(I - W_n \tilde{K})^{-1}|| \, ||y - W_n y||_\infty.$$

This leads to $||y - y_n||_\infty$ converging to zero exactly with the same rate of the norm of the nodal spline approximation error.

References

[1] ATKINSON K.E., *The numerical solution of integral equations of the second kind*, Cambridge Mon. on Appl. Comp. Math., Cambridge University Press **33** (1997), 59–672.

[2] BRUNNER H., *The numerical solution of weakly singular Volterra integral equations by collocation on graded meshes*, Math. Comp. **45** (1985), 417–437.

[3] CALIÓ F. AND MARCHETTI E., *On an algorithm for the solution of generalized Prandtl equations*, Numerical Algorithms **28** (2001), 3–10.

[4] CALIÓ F. AND MARCHETTI E., *An algorithm based on Q. I. modified splines for singular integral models*, Computer Math. Appl. **41** (2001), 1579–1588.

[5] DAGNINO C. AND DEMICHELIS V., *Nodal spline integration rules of Cauchy principal value integrals*, Intern. J. Computer Math. **79** (2002), 233–246.

[6] DAGNINO C. AND DEMICHELIS V., *Computational aspects of numerical integration based on optimal nodal spline*, Intern. J. Computer Math. **80** (2003), 243–255.

[7] DAGNINO C., DEMICHELIS V. AND SANTI E., *Numerical integration of singular integrands using quasi-interpolatory splines*, Computing **50** (1993), 149–163.

[8] DAGNINO C., DEMICHELIS V. AND SANTI E., *A nodal spline collocation method for weakly singular Volterra integral equations*, Studia Univ. Babes-Bolyai Math. **48** 3 (2003), 71–82.

[9] DAGNINO C., PEROTTO S. AND SANTI E., *Convergence of rules based on nodal splines for the numerical evaluation of certain Cauchy principal value integrals*, J. Comp. Appl. Math. **89** (1998), 225–235.

[10] DAGNINO C. AND RABINOWITZ P., *Product integration of singular integrands using quasi-interpolatory splines*, Computer Math. Applic. **33** (1997), 59–672.

[11] DAGNINO C. AND SANTI E., *Numerical evaluation of Cauchy principal integrals by means of nodal spline approximation*, Revue d'Anal. Num. Theor. Approx. **27** (1998), 59–69.

[12] DAHMEN W., GOODMAN T.N.T. AND MICCHELLI C.A., *Compactly supported fundamental functions for spline interpolation*, Numer. Math. **52** (1988), 641–664.

[13] DEMICHELIS V., *Uniform convergence for Cauchy principal value integrals of modified quasi-interpolatory splines*, Intern. J. Computer Math. **53** (1994), 189–196.

[14] DEMICHELIS V., *Convergence of derivatives of optimal nodal splines*, J. Approx. Th. **88** (1997), 370–383.

[15] DEMICHELIS V. AND RABINOWITZ P., *Finite part integrals and modified splines*, to appear in BIT .

[16] DE VILLIERS J.M., *A convergence result in nodal spline interpolation*, J. Approx. Theory **74** (1993), 266–279.

[17] DE VILLIERS J.M. AND ROHWER C.H., *Optimal local spline interpolants*, J. Comput. Appl. Math. **18** (1987), 107–119.

[18] DE VILLIERS J.M. AND ROHWER C.H., *A nodal spline generalization of the Lagrange interpolant*, in: "Progress in Approximation Theory" (Eds. Nevai P. and Pinkus A.), Academic Press, Boston 1991, 201–211.

[19] DE VILLIERS J.M. AND ROHWER C.H., *Sharp bounds for the Lebesgue constant in quadratic nodal spline interpolation*, International Series of Num. Math. **115** (1994), 1–13.

[20] DIETHELM K., *Error bounds for spline-based quadrature methods for strongly singular integrals*, J. Comput. Appl. Math. **89** (1998), 257–261.

[21] DIETHELM K., *Error bounds for spline-based quadrature methods for strongly singular integrals*, J. Comp. Appl. Math. **142** (2002), 449–450.

[22] GORI C. AND SANTI E., *Spline method for the numerical solution of Volterra integral equations of the second kind*, in: "Integral and Integro-differential equation" (Eds. Agarwal R. and Regan D.O.), Ser. Math. Anal. Appl. **2**, Gordon and Breach, Amsterdam 2000, 91–99.

[23] GORI C., SANTI E. AND CIMORONI M.G., *Projector-splines in the numerical solution of integro-differential equations*, Computers Math. Applic. **35** (1998), 107–116.

[24] LYCHE T. AND SCHUMAKER L.L., *Local spline approximation methods*, J. Approx. Theory **15** (1975), 294–325.

[25] MONEGATO G., *The numerical evaluation of a 2-D Cauchy principal value integral arising in boundary integral equation methods*, Math. Comp. **62** (1994), 765–777.

[26] RABINOWITZ P., *Product integration of singular integrals using optimal nodal splines*, Rend. Sem. Mat. Univ. Pol. Torino **51** (1993), 1–9.

[27] RABINOWITZ P., *Application of approximating splines for the solution of Cauchy singular integral equations*, Appl. Num. Math. **15** (1994), 285–297.

[28] RABINOWITZ P., *Optimal quasi-interpolatory splines for numerical integration*, Annals Numer. Math. **2** (1995), 145–157.

[29] RABINOWITZ P., *Uniform convergence results for finite-part integrals*, in: "Workshop on Analysis celebrating the 60^{th} birthday of Peter Vértesi and in memory of Ottò Kis and Apard Elbert", Budapest 2001.

[30] RABINOWITZ P. AND SANTI E., *On the uniform convergence of Cauchy principal values of quasi-interpolating splines*, BIT **35** (1995), 277–290.

[31] SANTI E., *Uniform convergence results for certain two-dimensional Cauchy principal value integrals*, Portugaliae Math. **57** (2000), 191–201.

[32] SANTI E. AND CIMORONI M.G., *On the convergence of projector-splines for the numerical evaluation of certain two-dimensional CPV integrals*, J. Comp. Math. **20** (2002), 113–120.

AMS Subject Classification: 65D30, 65D07.

Catterina DAGNINO, Vittoria DEMICHELIS
Dipartimento di Matematica
Università di Torino
via Carlo Alberto 8
10123 Torino, ITALIA
e-mail: catterina.dagnino@unito.it
vittoria.demichelis@unito.it

Elisabetta SANTI
Dipartimento di Energetica
Università dell' Aquila
Monteluco di Roio
67040 L'Aquila, ITALIA
e-mail: esanti@dsiaq1.ing.univaq.it

Rend. Sem. Mat. Univ. Pol. Torino
Vol. 61, 3 (2003)
Splines and Radial Functions

O. Davydov - R. Morandi - A. Sestini

SCATTERED DATA APPROXIMATION WITH A HYBRID SCHEME

Abstract. A local hybrid radial–polynomial approximation scheme is introduced here to modify the scattered data fitting method presented in [6], generating C^1 or C^2 approximating spline surfaces. As for the original method, neither triangulation nor derivative estimate is needed, and the computational complexity is linear. The reported numerical experiments relate to two well known test functions and confirm both the accuracy and the shape recovery capability of the proposed hybrid scheme.

1. Introduction

In this paper we investigate the benefit obtainable using local hybrid radial–polynomial approximations to modify the scattered data fitting method introduced in [6] which is based on direct extension of local polynomials to bivariate splines. The hybrid approach here considered is motivated by the well known excellent quality of scattered data radial basis function approximation [2]. Polynomial terms are also admitted in order to improve the approximation in the subdomains with polynomial-like behaviour of the data (e.g. almost flat areas).

Both the original and our hybrid scheme do not need data triangulation because the standard four directional mesh covering the domain is used. In addition, they do not require derivative estimates because only functional values are required. Clearly, the hybrid approximations must be converted in local polynomials for making possible their extension to splines. However, the additional computational cost of the conversion phase is negligible with respect to the whole cost of the method. In addition, as well as for the original method, the computational complexity of the scheme is linear, and this is obviously a very important feature particularly when large data sets have to be handled. A C^1 cubic or a C^2 sextic final spline approximation is produced, which is advantageous for CAGD applications since splines are a standard tool for that purpose [7].

In our modified approach, thanks to the usage of radial terms only in the local setting, the related local hybrid approximations can be computed without using special numerical techniques because the subset of data used for each of them is small and its size is assumed a priori bounded, which results in avoiding large matrices completely. In addition, for the same reason a simple and no–cost adaptation of the scaling parameter characterizing the radial terms of the hybrid approximations is possible. We note that local scaling adaptation is a nice feature of the scheme because, as proved by the researches reported by various authors (e.g. [1, 13, 14]), the use of different scaling parameters can be very proficuous in particular relating to shape recovery, but it is not easy when global radial schemes are used.

In this paper, in order to investigate the accuracy and the shape recovery capability of the method, we have experimented its performances by means of two reference mathematical test functions, that is the well known Franke [9] and Nielson [15] functions. For both the reported test functions the results highlight the good behaviour of the proposed hybrid scheme.

The paper is organized as follows. In Section 2 the original bivariate spline approximation method is summarized and in Section 3 the local hybrid approximation scheme is introduced. Finally in Section 4 the numerical results related to the two considered test functions are presented.

2. The original method

In this section we give some basic information about the original scattered data approximation scheme introduced in [6, 12] which is a two-stage method extending local approximations to the final global spline approximating surface. In fact, our scheme is obtained acting on the first stage of the original method, that is modifying the local approximations. On the other hand, the philosophy of the method and its second stage, devoted to the spline computation, are unchanged.

First, let us introduce some fundamental definitions (see for details [8]).

The **Bernstein-Bézier representation** of a bivariate polynomial p of total degree $\leq d$ is

$$p = \sum_{i+j+k=d} c_{ijk} \, B^d_{ijk}, \tag{1}$$

where B^d_{ijk}, $i + j + k = d$, $i, j, k \in \mathcal{N}$ are the Bernstein polynomials of degree d related to the reference triangle T with vertices $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$.

Each coefficient c_{ijk}, $i + j + k = d$ in (1) is associated with the **domain point** $\eta_{ijk} \in T$,

$$\eta_{ijk} := \frac{i}{d}\mathbf{a} + \frac{j}{d}\mathbf{b} + \frac{k}{d}\mathbf{c}.$$

The set of all the domain points associated with T is denoted by $\mathcal{D}_{d,T}$ and the set of all the domain points related to the triangles of the considered triangulation Δ is denoted by $\mathcal{D}_{d,\Delta}$.

A set $\mathcal{M} \subset \mathcal{D}_{d,\Delta}$ is called a **minimal determining set** for the linear subspace $\mathcal{S}$ of the spline space $\mathcal{S}^0_d(\Delta)$ if, setting the coefficients of $s \in \mathcal{S}$ associated with the domain points in $\mathcal{M}$ to zero implies that all the coefficients of s vanish and no proper subset of $\mathcal{M}$ exists with the same property.

We now summarize the original method we refer to, relating to [6] for a complete description. In this approach local polynomials are extended to bivariate splines producing a C^1 or C^2 approximating surface using cubic or sextic splines respectively. The extension to bivariate splines is done in the second stage by using the smoothness conditions between adjacent Bézier triangular patches [8]. A uniform four directional mesh Δ covering the domain $\Omega \subset \mathbb{R}^2$ is used and local polynomials are computed by discrete least squares using the stable Bernstein-Bézier representation form. The computational complexity of the method grows linearly with the number N of data points $\{(\mathbf{X}_i, f_i),\ i = 1, \ldots, N, \mathbf{X}_i \in \Omega \subset \mathbb{R}^2\}$. Thus, large data and many different data distributions can be efficiently handled, as shown in [6]. The efficiency of the method mainly depends on the theoretical determination of minimal determining sets $\mathcal{M}$ for the spline approximating spaces which consist of all domain points belonging to a set $\mathcal{T}$ of uniformly distributed triangles of Δ. In fact, using this result, local polynomial Bèzier patches can be separately computed for each triangle belonging to $\mathcal{T}$ and then univocally extended to the final spline approximation.

Concerning the local polynomial approximations, it is clear that their accuracy and shape quality heavily influences the corresponding attributes of the spline approximation. As a consequence, an important point is the selection of the data used for defining through the least squares procedure each local polynomial p_T of total degree $\leq d$ ($d = 3$ for cubics and 6 for sextics) on each triangle $T \in \mathcal{T}$. So, they initially correspond to locations $\mathbf{X}_i$ inside a circle Ω_T centered at the barycenter of T and with radius equal to the grid size. However, if they are few, the radius is suitably increased and if they are too many, in order to accelerate the computational process, their number N_T is decreased using a grid-type thinning algorithm. A lower and an upper bound M_{Min} and M_{Max} for N_T are assumed as input parameters provided by the user. Another important input parameter of the method is the tolerance κ_P used to control the inverse of the minimal singular value $\sigma_{min,d,T}$ of the collocation matrix $M_{d,T}$ related to the least-squares local polynomial approximation defined on each $T \in \mathcal{T}$. In fact, as proved in [4], imposing an upper bound for $\sigma^{-1}_{min,d,T}$ allows a direct control on the approximation power of the least-squares scheme, besides guaranteeing its numerical stability. An adaptive degree reduction procedure for guaranteeing this bound is used, producing constant approximations in the worst case.

3. The local hybrid scheme

As we already said in the introduction, the idea of our hybrid method is to enhance the approximation quality of the local approximations by using linear combinations of polynomials and radial basis functions. Once a local hybrid approximation g_T is computed on a triangle $T \in \mathcal{T}$, it is trasformed into a polynomial approximation of degree d computing the discrete least squares polynomial approximation of degree d with respect to the evaluations of g_T at all the $\binom{D+2}{2}$ domain points on T, where it is assumed $D = 2d$. On this concern, we remark that the additional cost related to this conversion phase is negligible with respect to the whole cost of the method mainly for two reasons. First, the collocation matrix associated with each local conversion hybrid–to–polynomial is the same for all triangles $T \in \mathcal{T}$. Second, it has a small

$\sigma^{-1}_{min,d,T}$ (2.87 for $D = 6$ and 21.74 for $D = 12$), so guaranteeing that the least squares polynomial of degree d is a good approximation of g_T [4].

Let $\Xi_T = \{\mathbf{X}_1, \ldots, \mathbf{X}_{N_T}\}$ denote the set of locations related to the triangle T (its definition is based on the same strategy used in the original method described in the previous section). The local mixed approximation g_T has the form

$$g_T(\cdot) = \sum_{j=1}^{m} a_j^T\, p_j^T(\cdot) + \sum_{j=1}^{n_T} b_j^T\, \phi_T(\|\cdot - \mathbf{Y}_j^T\|_2) \tag{2}$$

where span$\{p_1^T, \ldots, p_m^T\}$ is the space Π_q^2 of bivariate polynomials of degree $q \geq 0$ and $m = \binom{q+2}{2} \leq N_T$. The function $\phi_T : \mathbb{R}_{\geq 0} \to \mathbb{R}$ can be any suitably smooth positive definite function or a conditionally positive definite function of order at most $q + 1$ on $\mathbb{R}^2$ (see [2]). The approximation g_T is constructed minimizing the ℓ_2-norm of the residual on Ξ_T,

$$\Big(\sum_{i=1}^{N_T} (f_i - g_T(\mathbf{X}_i))^2\Big)^{1/2}, \tag{3}$$

where $0 \leq n_T \leq N_T - m$, and the set of knots $Y_T = \{\mathbf{Y}_j\,,\ j = 1, \ldots, n_T\}$ is a subset of Ξ_T.

We do not consider the additional orthogonality constraints

$$\sum_{j=1}^{n_T} b_j^T\, p(\mathbf{Y}_j^T) = 0, \qquad \text{all } p \in \Pi_q^2, \tag{4}$$

usually required in radial approximation ([2]), because we want to exploit in full the approximation power of the linear space

$$\mathcal{H}_T := \text{span}\left\{p_1^T, \ldots, p_m^T, \phi_T(\|\cdot - \mathbf{Y}_1^T\|_2), \ldots, \phi_T(\|\cdot - \mathbf{Y}_{n_T}^T\|_2)\right\}.$$

So we have to check the uniqueness of the solution of our least squares problem and this is done requiring that

$$\sigma^{-1}_{min}(C_T) \leq \kappa_H, \tag{5}$$

where κ_H is a user specified tolerance and $\sigma_{min}(C_T)$ is the minimal singular value of the collocation matrix C_T defined by

$$\begin{bmatrix} p_1^T(\mathbf{X}_1) & \ldots & p_m^T(\mathbf{X}_1) & \phi_T(\|\mathbf{X}_1 - \mathbf{Y}_1^T\|_2) & \ldots & \phi_T(\|\mathbf{X}_1 - \mathbf{Y}_{n_T}^T\|_2 \\ \vdots & & \vdots & \vdots & & \vdots \\ p_1^T(\mathbf{X}_{N_T}) & \ldots & p_m^T(\mathbf{X}_{N_T}) & \phi_T(\|\mathbf{X}_{N_T} - \mathbf{Y}_1^T\|_2) & \ldots & \phi_T(\|\mathbf{X}_{N_T} - \mathbf{Y}_{n_T}^T\|_2 \end{bmatrix}.$$

An adaptive ascending iterative strategy is used for defining n_T and the related set of knots Y_T. For the description of the details of such a strategy, the reader is referred to the forthcoming paper [5]. However here we just mention that this strategy is based on the inequality (5). The reason why we control the unique solvability of our least squares problem using (5) instead of a cheaper criterion avoiding the computation of $\sigma_{min}(C_T)$ ([16]) is because it allows us to control also the approximation error $\|f - g_T\|_{C(T)}$, where we are here assuming that $f_i = f(\mathbf{X}_i)$, $i = 1, \dots, N_T$, being f a continuous function. In fact, if (5) holds and N_T is upper bounded, assuming that the polynomial basis $\{p_1^T, \dots, p_m^T\}$ and ϕ_T are properly scaled, it can be proved that ([4, 5]) there exists a constant c_T such that

$$\|f - g_T\|_{C(T)} \le c_T \; E(f, \mathcal{H}_T)_{C(T)}, \tag{6}$$

where $E(f, \mathcal{H}_T)_{C(T)}$ is the error of the best approximation of f from $\mathcal{H}_T$,

$$E(f, \mathcal{H}_T)_{C(T)} := \inf_{g \in \mathcal{H}_T} \|f - g\|_{C(T)}.$$

4. Numerical results

The features of our local hybrid approximation scheme are investigated incorporating it into the two-stage scattered data fitting algorithm of [6]. More precisely, the method RQ_2^{av} of [6, Section 5] has been always used in the reported experiments, producing a C^2 piecewise polynomial spline of degree $d = 6$ with respect to the four-directional mesh. For our experiments in (2) we have always considered

$$\phi_T(r) \quad = \quad -\delta d_T \, \phi_{MQ}\left(\frac{r}{\delta d_T}\right) = \sqrt{(\delta d_T)^2 + r^2}, \tag{7}$$

where

$$d_T := \max_{1 \le i, j \le N_T} \|\mathbf{X}_i - \mathbf{X}_j\|_2$$

is the diameter of Ξ_T and δ is a scaling parameter. As this radial basis function is conditionally positive definite of order 1, we take $q = 0$, and thus the polynomial part in (2) is just a constant.

The input parameters to the method are the grid size $n_x \times n_y$ on a rectangular domain, the inverse minimal singular value tolerance κ_H, the minimum and maximum numbers M_{min}, M_{max} of data points belonging to each Ξ_T, the scaling coefficient δ used in (7), the upper bound n_{max} on the knot number n_T used in (2).

We consider here two tests, relating to the Franke (Test 1) and Nielson (Test 2) reference functions reported in Figure 1. Each displayed approximation is depicted together with the related data sample. For both considered tests a uniform 101×101 grid is used for the visualization and for the computation of the maximum (*maxg*) and root mean square (*rmsg*) errors. In all experiments below $n_{max} = 2\binom{d+2}{2} - 1$ and no upper bound for N_T is assigned, that is $M_{max} = N$. The lower bound M_{min} is always 20 and the scaling parameter δ in (7) is 0.4. The tolerance κ_H in (5) is taken to be equal to 10^5.

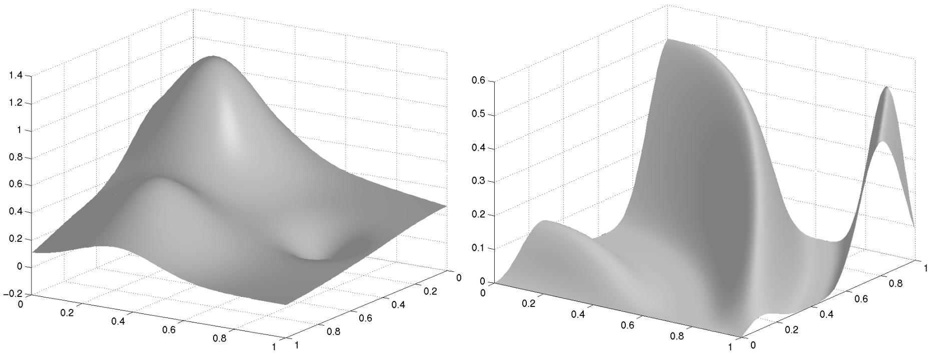


Figure 1: Franke and Nielson parent surfaces on the left and on the right, respectively.

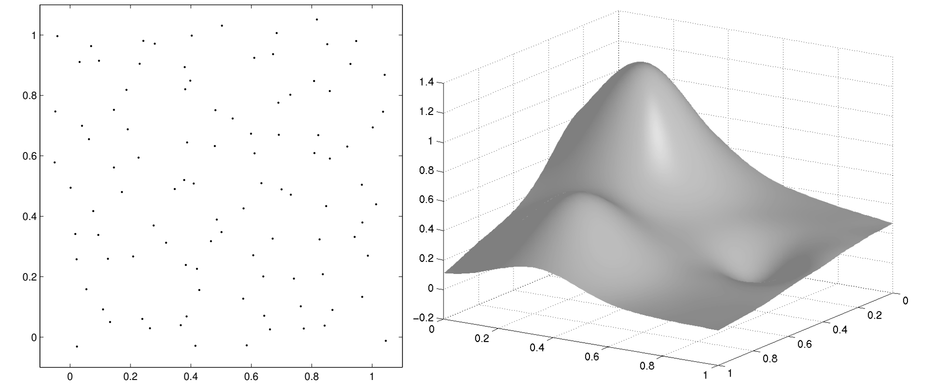


Figure 2: On the left the locations of the 100 data points for Test 1. On the rigth the related approximation.

In our first test, related to the Franke function, a small set of $N = 100$ data points is used. It is available from [10] as `ds3` and is shown on the left of Figure 2. The approximation depicted on the right of Figure 2 has been obtained using a uniform grid of size $n_x = n_y = 5$. The average number of knots used for the local hybrid approximations is 23.9. The related grid errors are $maxg = 1.5 \cdot 10^{-2}$ and $rmsg = 2.7 \cdot 10^{-3}$. For comparison, using the same grid size the errors obtained with the original method and reported in [6] are $maxg = 3.8 \cdot 10^{-2}$ and $rmsg = 7.6 \cdot 10^{-3}$ (see Table 3 of that paper). In addition, we found in the literature the following errors for the interpolation of this data with the global multiquadric method: $maxg = 2.3 \cdot 10^{-2}$ and $rmsg = 3.6 \cdot 10^{-3}$ in the famous Franke's report [9], and $rmsg = 2.6 \cdot 10^{-3}$ in [3]. (In both cases a uniform 33×33 grid was used to compute the error.) Note that the above error from [3] corresponds to the case when a parameter value for multiquadric was found by optimization.

Our second test relates to the Nielson function. First we have considered a small

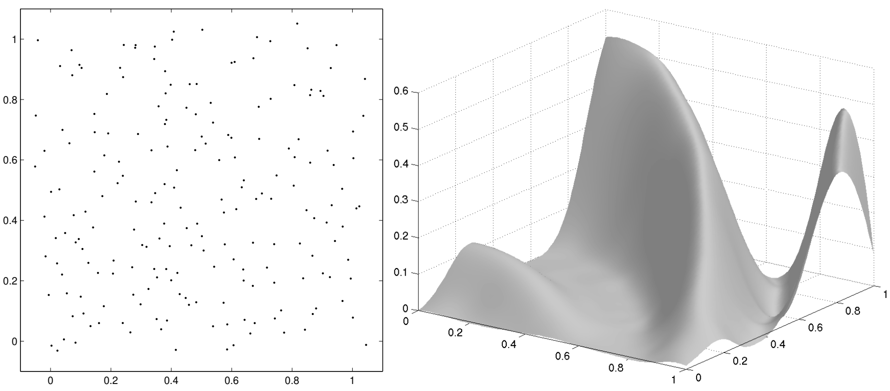


Figure 3: On the left the locations of the 200 data points for Test 2. On the rigth the related approximation.

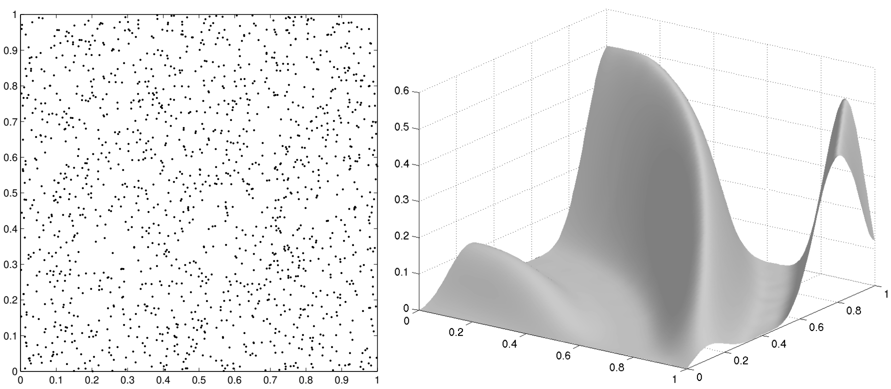


Figure 4: On the left the locations of the 1500 data points for Test 2. On the rigth the related approximation.

set of 200 data points obtained evaluating this function on the locations corresponding to the data points available from [10] as `ds4`. These locations are shown on the left of Figure 3. Again a uniform grid of size $n_x = n_y = 5$ is used. In this case the approximation shown on the right of Figure 3 is obtained using an average knot number equal to 22 and the related grid errors are $maxg = 6.9 \cdot 10^{-2}$ and $rmsg = 1.4 \cdot 10^{-2}$. For comparison, we mention that the same data set is used in [11] to test a least squares approximation method based on multiquadrics and parameter domain distortion. The best root mean square error (computed using a uniform 33×33 grid) reported in [11] is $1.3 \cdot 10^{-2}$ (see Table 1 and Figure 6 of that paper). Even if Figure 3 clearly shows some artifacts, we evaluate positively the results related to this first experiment for Test 2. In fact the accuracy and the shape recovery capability of our scheme are both comparable with those obtained in the best case reported in [11]. We would like also to say on this concern that, even if the results given in [11] have been obtained with remarkably few degrees of freedom, it should be taken into account that the parametric domain distortion method may encounter difficulties when applied to real data, as the authors admit [11, Section 4]. Finally, we get full shape recovery also for this challenging test function when we consider a denser set of 1500 scattered data depicted on the left of Figure 4 and use a finer spline grid by taking $n_x = n_y = 8$. The shape of the corresponding approximation depicted on the right of Figure 4 is almost perfect now and the related grid errors are $maxg = 3.8 \cdot 10^{-2}$ and $rmsg = 1.3 \cdot 10^{-3}$. The mean number of knots used in this case is 22.4.

References

[1] BOZZINI M. AND LENARDUZZI L., *Bivariate knot selection procedure and interpolation perturbed in scale and shape*, to appear in: “Proceedings of the Fifth Int. Conference on Curves and Surfaces”, Saint Malo 2002.

[2] BUHMANN M. D., *Radial basis functions*, Acta Numerica (2000), 1–38.

[3] CARLSON R. E. AND FOLEY T. A., *Interpolation of track data with radial basis methods*, Comp. Math. with Appl. **24** (1992), 27–34.

[4] DAVYDOV O., *On the approximation power of local least squares polynomials*, in: “Algorithms for Approximation IV”, (Eds. Levesley J., Anderson I.J. and Mason J.C.) 2002, 346–353.

[5] DAVYDOV O.,MORANDI R. AND SESTINI A., *Local hybrid approximation for scattered data fitting with bivariate splines*, preprint 2003.

[6] DAVYDOV O. AND ZEILFELDER F., *Scattered data fitting by direct extension of local polynomials to bivariate splines*, Adv. Comp. Math., to appear.

[7] FARIN G., HOSCHEK J. AND KIM M. S., (Eds.), *Handbook of computer aided geometric design*, Elsevier Science, North Holland 2002.

[8] FARIN G., *Curves and surfaces for computer aided geometric design*, Academic Press, San Diego 1993.

[9] FRANKE R., *A critical comparison of some methods for interpolation of scattered data*, Report NPS-53-79-003, Naval Postgraduate School 1979.

[10] FRANKE R., Homepage `http://www.math.nps.navy.mil/~rfranke/`, Naval Postgraduate School.

[11] FRANKE R. AND HAGEN H., *Least squares surface approximation using multiquadrics and parametric domain distortion*, Comput. Aided Geom. Design **16** (1999), 177–196.

[12] HABER J., ZEILFELDER F., DAVYDOV O. AND SEIDEL H.-P., *Smooth Approximation and Rendering of Large Scattered Data Sets*, in: "Proceedings of IEEE Visualisation 2001" (Eds. Ertl T., Joy K. and Varshney A.), 2001 341–347, 571.

[13] KANSA E. J. AND CARLSON R. E., *Improved accuracy of multiquadric interpolation using variable shape parameters*, Comp. Maths. with Appls. **24** (1992), 99–120.

[14] MORANDI R. AND SESTINI A., *Geometric knot selection for radial scattered data approximation*, in: "Algorithms for Approximation IV", (Eds. Levesley J., Anderson I.J. and Mason J.C.), 2002, 244–251.

[15] NIELSON G. M., *A first order blending method for triangles based upon cubic interpolation*, Int. J. Numer. Meth. Engrg. **15** (1978), 308–318.

[16] STEWART G. W., *Matrix Algorithms; Volume I: Basic Decompositions*, SIAM, Philadelphia 1998.

AMS Subject Classification: 41A15, 65D10.

Oleg DAVYDOV
Mathematisches Institut
Justus-Liebig-Universität Giessen
D-35392 Giessen, GERMANY
e-mail: `oleg.davydov@math.uni-giessen.de`

Rossana MORANDI, Alessandra SESTINI
Dipartimento di Energetica
Università di Firenze,
Via Lombroso 6/17
50134 Firenze, ITALY
e-mail: `morandi@de.unifi.it`
`sestini@de.unifi.it`

Rend. Sem. Mat. Univ. Pol. Torino
Vol. 61, 3 (2003)
Splines and Radial Functions

S. De Marchi*

ON OPTIMAL CENTER LOCATIONS FOR RADIAL BASIS FUNCTION INTERPOLATION: COMPUTATIONAL ASPECTS

Abstract. The problem of choosing "good" nodes is a central one in polynomial interpolation. Made curious from this problem, in this work we present some results concerning the computation of optimal points sets for interpolation by radial basis functions. Two algorithms for the construction of near-optimal set of points are considered. The first, that depends on the radial function, compute optimal points by adding one of the maxima of the *power function* with respect to the preceding set. The second, which is independent of the radial function, is shown to generate near-optimal sets which correspond to *Leja extremal points*. Both algorithms produce point sets almost similar, in the sense of their mutual *separation distances*. We then compare the interpolation errors and the growth of the Lebesgue constants for both point sets.

1. Introduction

First some introductory material and definitions concerning the interpolation problem with radial basis functions. Take a set $X = \{x_1, \ldots, x_N\} \subseteq \Omega \subseteq \mathbb{R}^d$ of N distinct points coming from a compact subset Ω of $\mathbb{R}^d$. The points $\{x_i\}$ are usually referred as the *data sites* and the set X as the *data set*. Suppose further that N data values $f_1, \ldots, f_N$ should be interpolated at the data sites. Fix then a basis function $\phi : [0, \infty) \to \mathbb{R}$, a simple way to define an interpolant $s_{f,X}$ to f at X is by linear combinations of the form

$$s_{f,X}(x) = \sum_{j=1}^{N} \alpha_j \phi(\|x - x_j\|) \tag{1}$$

where $\|\cdot\|$ is the Euclidean norm, and the coefficients $\{\alpha_j\}$ are uniquely determined by the interpolation conditions

$$s_{f,X}(x_i) = f_i, \quad i = 1, \ldots, N \tag{2}$$

if the interpolation matrix $A_{\phi,X} := (\phi(\|x_i - x_j\|)_{1 \le i,j \le N}$ is invertible. Furthermore, for various reasons it is sometimes necessary to add the space $\mathbb{P}_m^d$ of polynomials of

*This work has been done with the support of ex-60% funds of the University of Verona, year 2002.

degree $\leq m$ in $\mathbb{R}^d$ to the interpolating functions. Interpolation is then uniquely possible with the further requirement: if $p \in \mathbb{P}_m^d$ satisfies

$$p(x_i) = 0, \text{ for all } x_i \in X \;\Rightarrow\; p = 0$$

and if ϕ is *conditionally positive definite* (shortly CPD) of order m on Ω (cf. e.g. [16]). If $A_{\phi,X}$ is *positive definite* $\forall\, X \subseteq \Omega$, then ϕ is said *positive definite* (shortly PD), that is conditionally positive definite of *order m=0*. Instead of ϕ, we can consider the symmetric kernel function $\Phi(x, y) = \phi(\|x - y\|)$, so that $\Phi : \Omega \times \Omega \to \mathbb{R}$, which is the notation used later on in the paper.

In the paper we mainly focus to the case of *positive definiteness*, since every *CPD kernel* has an associated *normalized PD* kernel (cf. e.g. [2, 17]).

The problem of finding good interpolation points for RBF interpolations has been addressed only recently (cf. [3, 4, 8]). In particular, in [4] the authors showed how difficult is the problem just in the one dimensional setting because one has to globally minimize a highly nonlinear function of Nd unknowns which is usually a *hard* problem.

In our previous paper [7] we have already discussed the problem of finding good or near-optimal interpolation points for radial basis function interpolation essentially by minimizing the *power function* associated to the symmetric kernel Φ. The main result there was that those points are *asymptotically uniformly distributed* in the Euclidean norm. That is why we called them *near-optimal points*.

The paper is organized as follows. In section 2 we essentially describe what we consider *near-optimal points* for radial basis function interpolation and we introduce the tools we shall use in the rest of the paper. In section 3 after presenting two algorithms for computing near-optimal points, one depending on Φ and one independent, i.e. *data-independent*, we investigate on some computational aspects and consequences related to the problem presenting in particular for the dimension $d = 2$, the connection between these near-optimal points and *Leja extremal sequences*. In section 4 we present numerical results: in particular we show the interpolation errors when interpolants are built on near-optimal point sets, and the corresponding Lebesgue constants. In section 5 we conclude noticing that the most reliable near-optimal points are the ones connected to the proper Φ even if the data-independent ones are proved to be competitive.

2. Interpolation error, power function and Lebesgue constant

Given $\Phi : \Omega \times \Omega \to \mathbb{R}$, a positive definite kernel, the recovery of functions from function values $f(x_j)$ on the set $X = \{x_1, \ldots, x_N\} \subset \Omega$ of N different data sites, can be done via interpolants of the form

(3) $$s_{f,X} = \sum_{j=1}^{N} \alpha_j \Phi(\cdot, x_j)\ .$$

This interpolant, as in classical polynomial interpolation, can also be written in terms of *cardinal functions* $u_j \in V_X = \text{span}\{\Phi(\cdot, x) : x \in X\}$ such that $u_j(x_k) = \delta_{j,k}$. Then, the interpolant (3) takes the usual Lagrangian form

$$s_{f,X} = \sum_{j=1}^{N} f(x_j) u_j . \tag{4}$$

It is well-known that local error estimates for interpolation by radial basis functions have the form (cf. e.g. [15])

$$|f(x) - s_{f,X}(x)| \le \kappa \quad P_{\Phi,X}(x) \tag{5}$$

with κ a positive constant depending only on f and $P_{\Phi,X}$ being the *power function* that takes the explicit form

$$P^2_{\Phi,X}(x) = \Phi(x,x) - 2\sum_{j=1}^{N} u_j(x)\Phi(x, x_j) + \sum_{j,k=1}^{N} u_j(x) u_k(x) \Phi(x_j, x_k).$$

Moreover, letting $\mathbf{u} = (-1, u_1(x), \ldots, u_N(x))$ we have the alternative representation

$$P^2_{\Phi,X}(x) = \mathbf{u}\, A_{\Phi,Y}\, \mathbf{u}^T \ , \tag{6}$$

as a quadratic form, where $Y = X \cup \{x\}$ and $A_{\Phi,Y}$ is the interpolation matrix corresponding to the set Y. This representation says immediately that the power function is non-negative since the vector $\mathbf{u}$ annihilates all polynomials $\mathbb{P}^d_m$ due to the polynomial reproduction property.

For *positive definite* kernels, given the set X where the numbering of its points is fixed, for a second ordered set $Y = \{y_1, ..., y_N\}$ we consider the matrix $A_{\Phi,X}(y_1, ..., y_N) = \big(\Phi(y_i, x_j)\big)_{1\le i,j\le N}$. We note that this matrix is symmetric and has determinant that is independent of the order of the points in X. Moreover, since Φ is positive definite, the matrix is positive definite and has positive determinant that we denote by $\det_{\Phi,X}(y_1, ..., y_N) = \det\big(\Phi(y_i, x_j)\big)_{1\le i,j\le N}$. Thus, the *cardinal functions* have the useful representation

$$u_k(x) = \frac{\det_{\Phi,X}(x_1, ..., x_{k-1}, x, x_{k+1}, ..., x_N)}{\det_{\Phi,X}(x_1, ..., x_N)} , \tag{7}$$

which reminds the determinantal form of the elementary Lagrange polynomials in polynomial interpolation. Moreover, from the representations (6) and (7), the power function can also be rewritten as

$$P^2_{\Phi,X}(x) = \frac{\det_{\Phi,Y}(x, x_1, ..., x_N)}{\det_{\Phi,X}(x_1, ..., x_N)} . \tag{8}$$

In other words, the power function is nothing but the norm of the pointwise error functional, and it can be numerically evaluated from the Lagrange basis.

Typically, error estimates and convergence rates lead to the problem of bounding the power function in terms of the *fill distance*,

$$h_{X,\Omega} = \sup_{x \in \Omega} \min_{x_j \in X} \|x - x_j\|_2 .$$

We will not discuss the details here: the interested reader can refer to [19]. Instead, we remark that this minimization property has another consequence. Letting X and Y the point sets above defined, then the associated power functions must necessarily satisfy

$$P^2_{\Phi,X}(x) \geq P^2_{\Phi,Y}(x), \qquad x \in \Omega ,$$

due to the *maximality property* of the power function and the fact that the $P_{\Phi,X}$ vanishes only at the points of X (cf. [15, §4]) and this inequality holds pointwise and everywhere in Ω. The above inequality will be an important ingredient for the Algorithm 1 to be presented in the next section.

Also the *separation distance*

$$q_X = \frac{1}{2} \min_{\substack{x_i, x_j \in X \\ x_i \neq x_j}} \|x_i - x_j\| ,$$

plays a role in finding good points for radial basis function interpolation. In fact, in [8], the author studied point sets $X \subset \Omega$ which maximize the *uniformity*

$$\rho_{X,\Omega} = \frac{q_X}{h_{X,\Omega}} = \sup_{Y \in \mathcal{X}_\Omega} \rho_{Y,\Omega} ,$$

among all point sets $Y \in \mathcal{X}_\Omega$, $\mathcal{X}_\Omega$ consisting of Voronoi vertices used to decompose $\mathbb{R}^d$ into Voronoi tiles. The result there was that point sets that optimally balance $h_{X,\Omega}$ against q_X, are optimally distributed in the domain Ω.

Finally, our last tool is the *Lebesgue constant*. As in the (univariate) polynomial case, from the representation (4) we consider the Lebesgue function $\lambda_N(x) := \sum_{j=1}^N |u_j(x)|$. Its maximum value,

$$\Lambda_N := \max_{x \in \Omega} \lambda_N(x) = \max_{x \in \Omega} \sum_{j=1}^{N} |u_j(x)| , \tag{9}$$

is referred to as the associated Lebesgue constant and gives the norm of the interpolating projector $\mathcal{P}_n : \mathcal{C}(\Omega) \to V_\Omega$, with $V_\Omega = \mathrm{span}\{\Phi(\cdot, x) : x \in \Omega\}$, both spaces equipped with the sup-norm. As well-known in the polynomial case, optimal points are not known explicitly, therefore in applications we can restrict to *near-optimal* points, that is, roughly speaking, points whose Lebesgue constant grows asymptotically like the optimal one. Therefore, near-optimal points should be found among the ones that minimize Λ_N. In the framework of interpolation by polynomials, points that minimize the Lebesgue constant by maximizing the *Vandermonde determinant*, are known as *Fekete points*. Fekete points are well-known and widely studied for polynomial interpolation also in the multi-dimensional setting. For radial basis functions, only recently and only in the univariate case there were some attempts to find Fekete-like points [4]. The main conclusion of that paper was that, surprisingly w.r.t. the polynomial case in which Fekete points have the arccosine distribution, *optimal points for radial basis function interpolation are asymptotically equidistributed*. Actually, a similar conclusion for 2-dimensional domains was also obtained in the paper [8]. Iske considered perturbations of the data sites in order to improve the performance of the interpolation process, showing that good points realize a balance between the quantities q_X and $h_{X,\Omega}$. Moreover, the same author in [9] has shown that the Lebesgue constant Λ_N for interpolation by polyharmonic splines is indeed the *condition number* w.r.t. the sup-norm of the interpolation operator and that this constant is invariant under uniform scalings, rotations and translations of the domain.

On the basis of these arguments, using the representation by cardinal functions u_k of the interpolant $s_{f,X}$, we can try to minimize the Lebesgue constant by maximizing the denominator of each function u_k in (7). Unfortunately these Vandermonde-like matrices, which depend on Φ, are not always well-conditioned.

Hence, to find near-optimal points for radial basis function interpolation we can proceed along the following lines:

- by minimizing the power function, which depends on Φ, in order to minimize the error in (5);
- by finding a representation of the u_k by well-conditioned matrices (for instance using some kind of stable orthogonal expansions) and maximizing the corresponding Vandermonde matrix, like for Fekete points, in order to minimize the Lebesgue constant of the interpolating operator.

In this paper we have explored the first instance and in the next section we present two methods that allow to compute near-optimal interpolation points: the first minimizes the power function associated to the kernel Φ; the second, based on geometric considerations, is completely independent on Φ and related to *Leja extremal sequences*.

3. On computing near-optimal point locations and Leja sequences

In the recent paper [7] we presented a numerical method that produces well–distributed point sets based on a *greedy algorithm* that generates larger and larger point sets by adding at each step one of the point where the power function attains its maxima with respect to the preceding set. The algorithm steps are as follows.

Algorithm 1

1. Initial step: $X_1 = \{x_1\}$ for some $x_1 \in \Omega$ arbitrary chosen.
2. Iterative step:

(10) $$X_j := X_{j-1} \cup \{x_j\} \text{ with } P_{\Phi,X_{j-1}}(x_j) = \|P_{\Phi,X_{j-1}}\|_{L_\infty(\Omega)}, \ \ j \geq 2.$$

Note that practically, we maximized over some very large discrete set $X \subset \Omega$ instead of maximizing on Ω. Letting $P_j := P_{\Phi, X_j}$, this algorithm converges in the sense that $\lim_{j \to \infty} \|P_j\|_{L_\infty(\Omega)} = 0$. In fact, since the point x_{j+1} is such that $P_j(x_{j+1}) = \|P_j\|_{L_\infty(\Omega)}$ and since $X_j \subseteq X_{j+1}$, we have $P_j(x) \geq P_{j+1}(x) \geq 0$ for all $x \in \Omega$.

The convergence and the speed of convergence of the Algorithm 1 are stated in the following Theorem.

THEOREM 1. (cf. [7, §4]) *Suppose $\Omega \subseteq \mathbb{R}^d$ is compact and satisfies an interior cone condition. Suppose further that $\Phi \in \mathcal{C}^2(\Omega_1 \times \Omega_1)$ is a positive definite kernel defined on a convex and compact region $\Omega_1 \supseteq \Omega$. Then, the greedy algorithm defined in (10) converges at least like*

$$\|P_j\|_{L_\infty(\Omega)} \leq C\, j^{-1/d}$$

with a constant $C > 0$.

REMARKS. The Theorem holds for positive definite kernels: this is not a big restriction since, as already pointed out, every CPD kernel has an associated NPD kernel (cf. Introduction). We also observe that the positive constant C is independent of j and that the power function depress to zero quite slowly, as will appear clearer from Examples.

3.1. A geometric greedy method

From experiments we have noted that Algorithm 1, that minimizes the power function $P_{\Phi,X}$, practically fills the currently largest hole in the data by placing a new data point close to the center of that hole and as a surprise, *independently* of the function Φ. Therefore, this observation suggested a new algorithm that we termed *geometric greedy algorithm* since the construction of *optimal* points is simply based on geometric considerations.

Algorithm 2

1. Let Ω be a compact set in $\mathbb{R}^d$, and consider $X_0 = \{x_0\}$ where x_0 belongs to the boundary of Ω.
2. If $X_n \subset \Omega$ is finite and consisting of n points, choose $x_{n+1} \in \Omega \setminus X_n$ so that its distance to X_n is maximal. Thus, $X_{n+1} := X_j \cup \{x_{n+1}\}$.

REMARKS. As before, for numerical purposes we should consider a discretization of Ω that is a finite set, say Ω_N, with *cardinality* N. Then, each step of the algorithm can be carried out in $\mathcal{O}(N)$ operations, since for each $x \in \Omega_N \setminus X_n$ we should compute the distance to its nearest neighbor within X_n. To update this array of length N, it requires firstly computing the $N - n$ values $\|x - x_i\|_2, \quad i = 1, ..., N - n$ and then taking the componentwise minimum within the i-th array of distances. The next point x_{n+1} is then easily found by picking the maximum of the array of the minima.

Defining the separation distance for points in X_n by

$$q_n := \frac{1}{2} \min_{\substack{x,\, y \,\in\, X_n \\ x \neq y}} \|x - y\|_2$$

and the corresponding fill distance

$$h_n := \max_{x \in \Omega} \min_{y \in X_n} \|x - y\|_2 = \min_{y \in X_n} \|x_{n+1} - y\|_2 = h_{X_n,\Omega} \ .$$

PROPOSITION 1. *Algorithm 2 produces point sets which are* **quasi-uniform** *in the Euclidean distance, that is*

$$h_n \geq q_n \geq \frac{1}{2} h_{n-1} \geq \frac{1}{2} h_n \ , \quad \forall n \geq 2.$$

Proof. The left–hand and right–hand sides are obvious. The remaining inequalities can be settled by induction. Indeed, for X_2 we have

$$q_2 = \frac{1}{2} \|x_2 - x_1\|_2 = \frac{1}{2} \min_{y \in X_1} \|x_2 - y\|_2 = \frac{1}{2} h_1 .$$

Assuming that $q_n \geq \frac{1}{2} h_{n-1}$, then

$$q_{n+1} = \min \left\{ q_n, \frac{1}{2} \min_{x \in X_j} \|x_{n+1} - x\|_2 \right\} = \min \left\{ q_n, \frac{1}{2} h_n \right\} \ ,$$

we get $q_{n+1} \geq \min \left\{ \frac{1}{2} h_{n-1}, \frac{1}{2} h_n \right\} \geq \frac{1}{2} h_n$.

□

REMARKS.

- The above Algorithm 2 turns out to work quite well when it comes to finding subsets of Ω of cardinality n with *small* fill distance $h_{X,\Omega}$ and *large* separation distance q_X.
- The construction technique proposed in the Algorithm 2 is *independent of the Euclidean metric*. In fact, the proof does not depend on the fact that q_n and h_n are expressed by using the Euclidean metric. Hence, if μ is any metric on Ω, the Algorithm 2 can be used to compute points asymptotically equidistributed in the metric μ.

3.2. Leja sequences

Leja extremal sequences were introduced by F. Leja in his interesting paper (cf. [10]) and recently have attracted the attention of researchers for their important properties and applications (cf. e.g. [13, 1, 6]).

DEFINITION 1. Let λ_1 be arbitrarily chosen in $[a, b]$. The points $\lambda_s \in [a, b]$, $s = 2, ..., N$, such that

$$\prod_{k=1}^{s-1} |\lambda_s - \lambda_k| = \max_{x \in [a,b]} \prod_{k=1}^{s-1} |x - \lambda_k| \quad . \tag{11}$$

are called a *Leja sequence* for the interval $[a, b]$ (cf. [10]).

We recall that Leja points, in the one-dimensional case, are computationally effective for polynomial interpolation in Newton form since they provide an increasing sequence of points and they stabilize the computation of divided differences. Moreover, they can be extracted from a discretization of $[a, b]$ in a fast way (the so-called *fast Leja points*) and, like Chebyshev points, Fekete points and zeros of Jacobi orthogonal polynomials, they have the *arccosine distribution* (cf. [13, 1]).

Unfortunately the multivariate equivalent of Leja points are not yet completely explored, as it is the case for the study of near-optimal points for multivariate interpolation (cf. [14, 6, 5]). For $d = 2$ something has been done.

DEFINITION 2. *Let Ω be a compact subset of $\mathbb{C} \approx \mathbb{R}^2$ and $w : \Omega \to \mathbb{R}_+$ a real positive function on Ω called* weight function. *Let $z_0 \in \Omega$ be such that*

$$w(z_0)\|z_0\| = \max_{z \in E} w(z)\|z\| \ , \tag{12}$$

and

$$w(z_n) \prod_{k=0}^{n-1} \|z_n - z_k\| = \max_{z \in E} w(z) \prod_{k=0}^{n-1} \|z - z_k\| \ , \ z_n \in \Omega \, . \tag{13}$$

where $\|\cdot\|$ is any norm of $\mathbb{R}^2$ and $z = x + iy$, $z = (x, y)$ and $z_n = (x_n, y_n)$, $n = 1, 2, ...$ The sequence $\{z_n\}$ not-unique that satisfies (12) and (13) is called a **sequence of Leja points** *for Ω.*

The distribution of Leja points so defined, depends on the choice of the weight function w. Indeed, when $w \equiv 1$, for the maximum principle of analytic functions Leja points distribute only on the boundary of Ω while for $w \neq 1$ they lye also in the interior (cf. [14] and for more examples see [6]). A conceptually similar construction of Leja points which is *independent of the weight w*, was suggested by L. Bos (private communication to the author). The idea behind the construction is simple: *find a sequence of points that maximize a function of distances from already computed points.* The proposed distance was simply the *Euclidean distance*.

DEFINITION 3. *Let Ω_N be a discretization of a compact domain $\Omega \subset \mathbb{C} \equiv \mathbb{R}^2$ and let z_0 be arbitrarily chosen in Ω_N. The points z_n, $n = 1, 2, ...$*

$$z_n := \max_{z \in \Omega_N \setminus \{z_0, \ldots, z_{n-1}\}} \ \min_{0 \le k \le n-1} \|z - z_k\|_2 \ . \tag{14}$$

are a set of **Leja points** *for Ω_N.*

In Figure 1 we show 60 Leja points on three classical domains computed by means of (14).

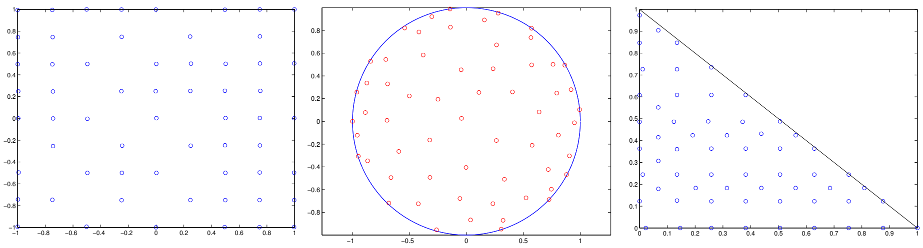


Figure 1: 60 Leja points on the square, the unit circle and the right triangle all discretized by 60^3 random points, computed by using (14).

Moreover, supported by numerical experiments, L. Bos proposed that the following claim should be true.

CLAIM. If $z_0 = \max_{z\in\Omega} \|z\|_2$, then the Leja points defined by (14) are asymptotically equidistributed w.r.t. the Euclidean metric.

REMARKS. The previous Definition 3 and the successive Claim, firstly stated in the framework of Leja sequences, reveal the connection with near-optimal points computed by Algorithm 2. From Proposition 1, we now know that the points constructed by (14) are indeed the data–independent ones. Therefore, to prove the previous Claim we simply resort to the proof of Proposition 1.

⋄⋄

Thus, by Algorithms 1 and 2 (or equivalently by (14)) we have two sets of near–optimal points for radial basis function interpolation. How close are these point sets? Which point set is “better” for interpolation purposes? These are some of the questions that we want to answer by numerical experiments in the next section.

4. Numerical results

In this section we present some examples of distribution of points as computed by Algorithms 1 and 2 in the bidimensional setting. We considered the square $\Omega = [-1, 1] \times [-1, 1]$ on which we picked 10000 random points. We have run the Algorithm 1 until the norm of the power function went below some fixed threshold $\tilde{\eta}$. As for Algorithm 2, we computed *once and for all* the necessary points up to a given number extracting them from a discretization of Ω. We have computed 406 points extracted from a discretization of 406^3 points of Ω. The number 406 corresponds to the dimension of the bivariate polynomials of degree ≤ 27 and the reason why we have extracted N points from N^3, comes from the theory of Leja sequences, as explained in the book [14]. Moreover, we stopped to 406 because of RAM limitations of the machine where computations where done. Consider that representing 406^3 reals in double precisions requires 510Mb of RAM. But, this was not a big problem, since in the following examples the points computed by Algorithm 1 were always less than 406.

In Figures 2-4 we show the distributions of the points computed both with the greedy method, Algorithm 1, and the geometric greedy method, Algorithm 2. On each figure we also show the *separation distances* among these points, making visually clearer that Algorithm 2 generates points nearly equidistributed in the Euclidean metric (as stated in Proposition 1).

By means of the Algorithm 1 applied to the Gaussian with scale 1, to reduce the power function below $\tilde{\eta} = 2 \cdot 10^{-7}$, we computed 65 points. For the Wendland's compactly supported function with scale 1, to reduce the power function below $\tilde{\eta} = 0.1$ we computed 80 optimal points and for the inverse multiquadrics with scale 1, we computed 90 points to depress the power function to $\tilde{\eta} = 2 \cdot 10^{-5}$. The choice of different $\tilde{\eta}$ depends on the decreasing rates of the associated power functions. Note that, for a given N, i.e. the number of optimal points we wish to find, so far we are not able to determine $\tilde{\eta}_\Phi(N)$ corresponding to a particular Φ.

Furthermore, given Φ_1, let X_1 be the optimal point set computed by minimizing the associated power function, say P_{Φ_1, X_1}, using Algorithm 1. Are these points optimal also for another $\Phi_2 \neq \Phi_1$? If not, are the points computed by the Algorithm 2 optimal for any given Φ, instead? In what follows, we will try to give qualitative answers to these "obvious" questions, showing in particular that the points computed by Algorithm 2 are good enough for almost all radial basis function interpolation problems.

We labeled by *g-gauss-65*, *gg-65*, *g-wend-80*, *gg-80*, *g-invm-90* and *gg-90* the point sets computed by Algorithm 1 and 2, where the prefix '*g*' recalls the word *greedy* while '*gg*' the words *geometric greedy*. The labels *gauss, wend, invm* recall instead the type of the radial function used in the minimization process. The '*gg*' point sets do not need to recall the radial function since they are independent of it.

As for interpolation, we have considered two test functions: $f_1(x, y) = \mathrm{e}^{-8(x^2+y^2)}$ and $f_2(x, y) = \sqrt{x^2 + y^2 - xy}$. The first is $\mathcal{C}^\infty$, while the second has discontinuity of the gradient. In Tables 1-3, we show the interpolation errors in the L_2-norm when the interpolant is constructed by means of the Gaussian, Wendland's and inverse multiquadrics, respectively. Each columns has an heading that recalls the set of points on which interpolation took place. The errors have been computed by sampling the functions on a regular grid of 30×30 points. While errors for the Gaussian are meaningless except in some cases, essentially due to errors occurring along boundaries, the interpolation errors for Wendland's and the inverse multiquadrics confirm, once again, that the points computed by Algorithm 2 are as good as the points computed by Algorithm 1.

	g-gauss-65	gg-65	g-wend-80	gg-80	g-invm-90	gg-90
f_1	$5.5\,10^{-1}$	**	$5.6\,10^{-1}$	**	$4.9\,10^{-1}$	**
f_2	$7.3\,10^{-1}$	**	**	**	**	**

Table 1. Errors in L_2-norm for interpolation by the Gaussian. When errors are > 1.0 we put $**$.

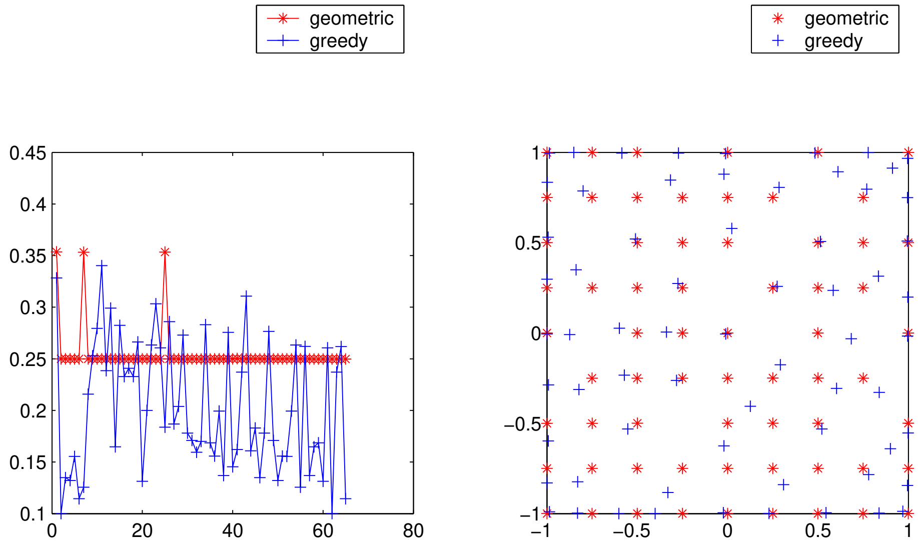


Figure 2: 65 optimal points for the Gaussian with scale 1. Right: the points as computed by the geometric greedy algorithm (*) and the greedy algorithm (+). Left: the separation distances among them.

	g-gauss-65	gg-65	g-wend-80	gg-80	g-invm-90	gg-90
f_1	$2.1\,10^{-1}$	$1.6\,10^{-1}$	$1.3\,10^{-1}$	$1.1\,10^{-1}$	$1.4\,10^{-1}$	$1.0\,10^{-1}$
f_2	$6.1\,10^{-1}$	$8.7\,10^{-1}$	$6.1\,10^{-1}$	$9.7\,10^{-1}$	$4.6\,10^{-1}$	$5.8\,10^{-1}$

Table 2. Errors in L_2-norm for interpolation by the Wendland's function.

	g-gauss-65	gg-65	g-wend-80	gg-80	g-invm-90	gg-90
f_1	$2.3\,10^{-1}$	$2.3\,10^{-1}$	$4.0\,10^{-2}$	$3.1\,10^{-2}$	$3.5\,10^{-2}$	$2.5\,10^{-2}$
f_2	$5.9\,10^{-1}$	$6.0\,10^{-1}$	$3.8\,10^{-1}$	$4.6\,10^{-1}$	$3.7\,10^{-1}$	$3.6\,10^{-1}$

Table 3. Errors in L_2-norm for interpolation by the inverse multiquadrics.

We have also computed, and plotted in Figures 5-7, the *Lebesgue constants* associated to these near-optimal point sets. The abscissas represent the *polynomial degree* and run till the maximum polynomial degree representable with the number of points in the sets. With the 65 points computed with the Gaussian the maximum degree is 9; for the 80 points for the Wendland's function and the 90 points computed for the inverse multiquadrics the maximum polynomial degree is 11. The computations of Lebesgue constants by means of (9) were done by discretizing the square $[-1, 1]^2$ with a grid of 40×40 points where we sampled the cardinal functions u_k. The graphs show that, *except for the Gaussian*, the Lebesgue constants of the optimal points computed by the greedy method grow slower than the ones of the data-independent points. Moreover, in all cases they grow approximately linearly in the polynomial degree (modulo some constants). This explains once more why the errors computed with the Gaussian are

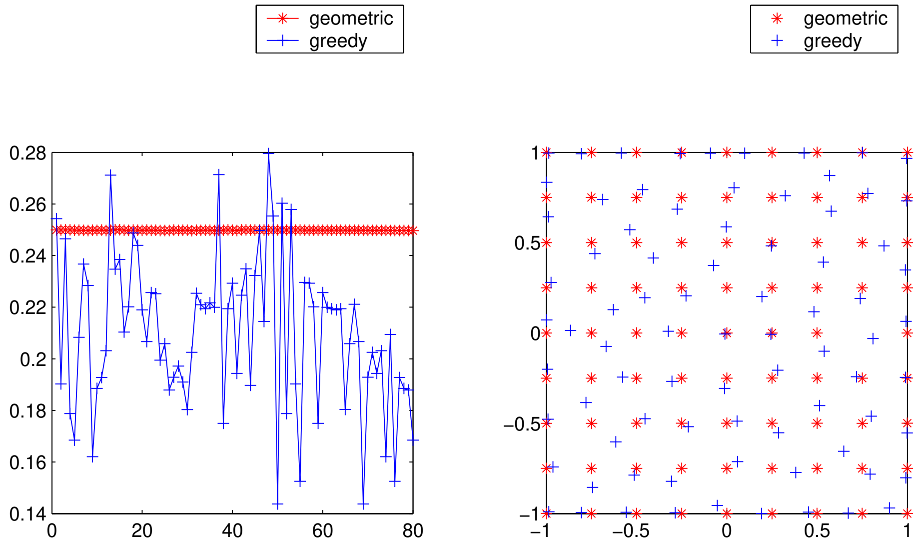


Figure 3: 80 optimal points for the Wendland's function. Right: the points as computed by the geometric greedy algorithm (*) and the greedy algorithm (+). Left: the separation distances among them.

meaningless. Of course, for a complete understanding of the asymptotic behavior of the Lebesgue constants we should go further in the computations, but we were not able due to hardware restrictions.

Concerning the computational efforts of both algorithms, we show in Table 4 the CPU time in seconds of Algorithm 1 for computing the optimal points for a given threshold. These computational costs were determined by the Matlab function `cputime`.

Gaussian scale 1, $\tilde{\eta} = 2 \cdot 10^{-7}$, 65 points, 51 sec.
Gaussian scale 2, $\tilde{\eta} = 2 \cdot 10^{-7}$, 32 points, 18 sec.
Wendland scale 1, $\tilde{\eta} = 0.1$, 80 points, 76 sec.
Wendland scale 15, $\tilde{\eta} = 2 \cdot 10^{-5}$, 100 points, 105 sec.
inverse multiquadrics scale 1, $\tilde{\eta} = 2 \cdot 10^{-5}$, 90 points, 110 sec.
inverse multiquadrics scale 2, $\tilde{\eta} = 2 \cdot 10^{-5}$, 34 points, 26 sec.

Table 4. Computational costs (`cputime` in seconds) of optimal points as computed by Algorithm 1.

Algorithm 2 was run *once and for all* to compute at once the 406 points by means of (14). These computations were done in about 5 minutes of CPU time on a PC with 900MHz Athlon processor and in this case the program was written in Fortran 77. The coordinates of the points were stored in a file and used later on with the same

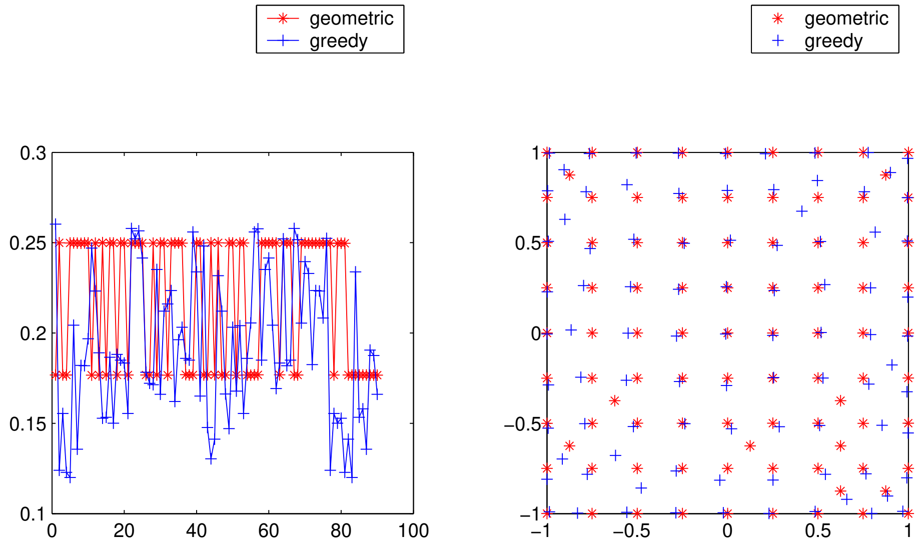


Figure 4: 90 optimal points for the inverse multiquadrics with scale 1. Right: the points as computed by the geometric greedy algorithm (*) and the greedy algorithm (+). Left: the separation distances among them.

Matlab program that we wrote for making comparison plots, computing separation and fill distances as well as Lebesgue constants with respect to the points computed by Algorithm 1.

5. Conclusions

The paper essentially presented two main results.

- Optimal points for radial basis function interpolation can be computed independently of the radial function and *once for all*. These points, in the two-dimensional case, correspond to Leja points in the Euclidean metric. They are asymptotically equidistributed with respect to the Euclidean metric (that is why we called *near-optimal*). Moreover, Algorithm 2 can be used with any metric, producing point sets asymptotically equidistributed with respect to that metric.

- From the Lebesgue constants behavior, we can conclude that data-independent points have Lebesgue constants that grow faster than data-dependet ones. Experiments on the growth of the Lebesgue constants on different nodal sets for bivariate polynomial interpolation are currently in progress and will be presented in the forthcoming paper [5]. From the results in that paper, here we only observe that quasi-uniformity is only a *necessary* condition for near-optimality of a point set. Therefore, generally speaking, the study of the growth of the Lebesgue

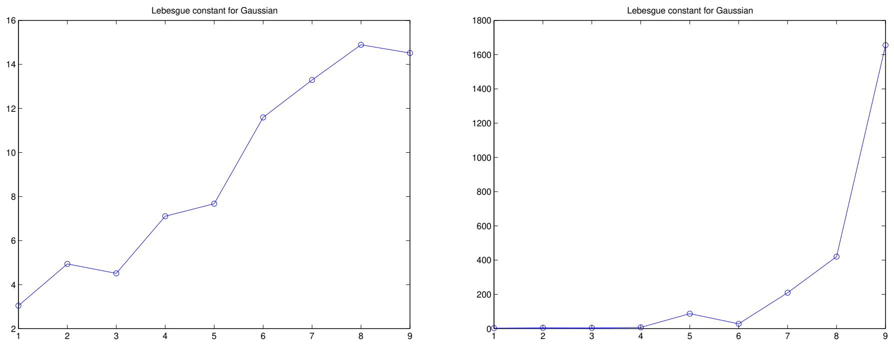


Figure 5: Lebesgue constants of the optimal points for the Gaussian (left) and the data-independent points (right).

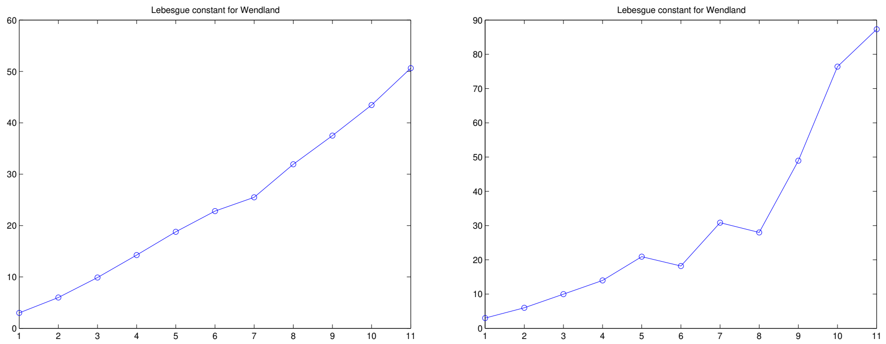


Figure 6: Lebesgue constants of the optimal points for the Wendland's function (left) and the data-independent points (right).

constant of a point set is not a general criterion to investigate on the goodness of a point set. In the univariate setting for polynomial interpolation on bounded intervals, a similar conclusion was obtained in the paper [11]. Hence, we can confirm that data-independent points should be used in radial basis function interpolation because of their general and effective computational technique and their interpolation errors which are of the same order of the near-optimal points computed by minimizing the power function.

Acknowledgments. I would like to express my thanks to Prof. Len Bos of the University of Calgary who pointed out to me the possibility to compute Leja-like sequences by means of (14). Thanks also to an anonymous referee who made punctual suggestions which essentially improved the results presented in the paper.

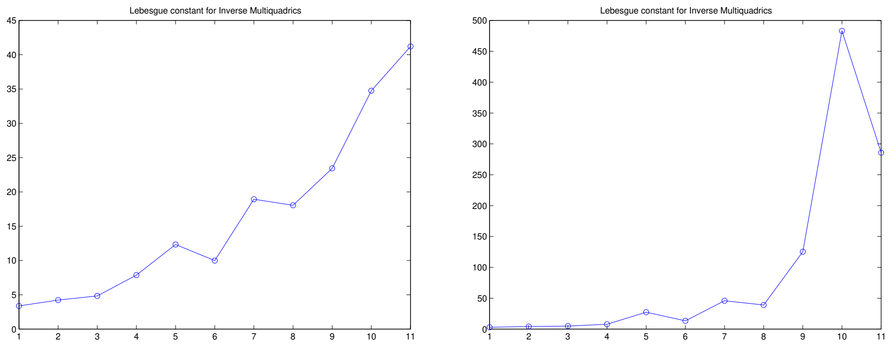


Figure 7: Lebesgue constants of the optimal points for the inverse multiquadrics (left) and the data-independent points (right).

References

[1] BAGLAMA J., CALVETTI D. AND REICHEL L., *Fast Leja points*, Electron. Trans. Numer. Anal. **7** (1998), 124–140.

[2] BEATSON R.K., LIGHT W.A. AND BILLINGS S., *Fast solution of the radial basis function interpolation equations: Domain decomposition methods*, SIAM J. Sci. Comput. **22** (2000), 1717–1740.

[3] BEYER A., *Optimale Centerverteilung bei Interpolation mit radialen Basisfunktionen*, Diplomarbeit, University of Göttingen, Göttingen 1994.

[4] BOS L.P. AND MAIER U., *On the asymptotics of points which maximize determinants of the form* $det(g(|x_i - x_j|))$, in: "Advances in Multivariate Approximation", (Eds. Haussmann W., Jetter K. and Reimer M.), Math. Res. **107**, Wiley-VCH., Berlin 1999, 107–128.

[5] CALIARI M., DE MARCHI S. AND VIANELLO M., *Bivariate Polynomial Interpolation at new nodal sets*, in preparation (2003).

[6] DE MARCHI S., *On Leja sequences: some results and applications*, in press on Appl. Math. Comput. (2003).

[7] DE MARCHI S., SCHABACK R. AND WENDLAND H., *Near-optimal data-independent point locations for radial basis function interpolation*, accepted by Adv. Comput. Math. (2003).

[8] ISKE A., *Optimal distribution of centers for radial basis function methods*, Tech. Rep. M0004, Technische Universität München, München 2000.

[9] ISKE A., *On the Approximation Order and Numerical Stability of Local Lagrange Interpolation by Polyharmonic Splines*, to appear in: "Proceedings of the 5th

International Conference on Multivariate Approximation" (Bommerholz, Sept. 2002); preprint, Technische Universität München, München 2003.

[10] LEJA F., *Sur certaines suites liées aux ensembles plans er leur application à la représentation conforme*, Ann. Polon. Math. **4** (1957), 8–13.

[11] MASTROIANNI G. AND OCCORSIO D., *Optimal systems of nodes for Lagrange interpolation on bounded intervals. A survey*, J. Comput. Appl. Math. **134** 1-2 (2001), 325–341.

[12] MICCHELLI C.A., *Interpolation of scattered data: distance matrices and conditionally positive definite functions*, Constr. Approx. **2** 1 (1986), 11–22.

[13] REICHEL L., *Newton interpolation at Leja points*, BIT Numerical Analysis **30** (1990), 332–346.

[14] SAFF E.B. AND TOTIK V., *Logarithmic potential with external fields*, Springer, Berlin 1997.

[15] SCHABACK R., *Reconstruction of Multivariate Functions from Scattered Data*, available at `www. num. math. uni-goettingen. de/ schaback/ research/ papers/ rbfbook.ps`.

[16] SCHABACK R., *Comparison of radial basis function interpolants*, in: "Multivariate Approximations: from CAGD to Wavelets", (Eds. Jetter K. and Utreras F.), World Scientific Publishing Co., 1993.

[17] SCHABACK R., *Native Hilbert spaces for radial basis functions I*, in: "New developments in approximation theory", (Eds. Müller M.W. et al.), Int. Ser. Numer. Math. **132**, Birkhäuser Verlag, Basel 1999, 255–282.

[18] WENDLAND H., *Error estimates for interpolation by compactly supported radial basis functions of minimal degree*, J. Approx. Theory **93** (1998), 258–272.

[19] WU Z. AND SCHABACK R., *Local error estimates for radial basis function interpolation of scattered data*, IMA Journal of Numerical Analysis **13** (1993), 13–27.

AMS Subject Classification: 41A05, 41A25, 41A63, 65D05, 65D15.

Stefano DE MARCHI
Department of Computer Science
University of Verona
S.da Le Grazie, 15
37134 Verona, ITALY
e-mail: `stefano.demarchi@univr.it`

Rend. Sem. Mat. Univ. Pol. Torino
Vol. 61, 3 (2003)
Splines and Radial Functions

G. Pittaluga - L. Sacripante - E. Venturino

A COLLOCATION METHOD FOR LINEAR FOURTH ORDER BOUNDARY VALUE PROBLEMS

Abstract. We propose and analyze a numerical method for solving fourth order differential equations modelling two point boundary value problems. The scheme is based on B-splines collocation. The error analysis is carried out and convergence rates are derived.

1. Introduction

Fourth order boundary value problems are common in applied sciences, e.g. the mechanics of beams. For instance, the following problem is found in [3], p. 365: The displacement u of a loaded beam of length $2L$ satisfies under certain assumptions the differential equation

$$\begin{aligned}
\frac{d^2}{ds^2}\left(EI(s)\frac{d^2u}{ds^2}\right)+Ku &= q(s), \quad -L \le s \le L,\\
u''(-L) &= u'''(-L)=0,\\
u''(L) &= u'''(L)=0.
\end{aligned}$$

Here,

$$I(s)=I_0\left(2-\left(\frac{s}{L}^2\right)\right), \quad q(s)=q_0\left(2-\left(\frac{s}{L}\right)^2\right), \quad K=\frac{40EI_0}{L^4},$$

where E and I_0 denote constants.

We wish to consider a general linear problem similar to the one just presented, namely

$$LU \equiv U^{(iv)}+a(x)U''(x)+b(x)U(x)=f(x) \tag{1}$$

for $0<x<1$, together with some suitable boundary conditions, say

$$U(0)=U_{00}, \quad U'(0)=U_{01}, \quad U'(1)=U_{11}, \quad U(1)=U_{10}. \tag{2}$$

Here we assume that $a, b \in C^0[0,1]$. In principle, the method we present could be applied also for initial value problems, with minor changes. In such case (2) could be replaced by suitable conditions on the function and the first three derivatives of the unknown function at the point $s=0$.

The technique we propose here is a $B-$spline collocation method, consisting in finding a function $u_N(x)$

$$u_N(x) = \alpha_1 \Phi_1(x) + \alpha_2 \Phi_2(x) + ... + \alpha_N \Phi_N(x)$$

solving the N×N system of linear equations

$$(3) \qquad Lu_N(x_i) \equiv \sum_{j=1}^{N} \alpha_j L\Phi_j(x_i) = f(x_i)\ ,\ 1 \le i \le N$$

where $x_1,\ x_2,\ \ldots,\ x_N$ are N distinct points of [0,1] at which all the terms of (3) are defined.

In the next Section the specific method is presented. Section 3 contains its error analysis. Finally some numerical examples are given in Section 4.

2. The method

A variety of methods for the solution of the system of differential equations exist, for instance that are based on local Taylor expansions, see e.g. [1], [2], [6], [7], [8], [16]. These in general would however generate the solution and its derivatives only at the nodes. For these methods then, the need would then arise to reconstruct the solution over the whole interval. The collocation method we are about to describe avoids this problem, as it provides immediately a formula which gives an approximation for the solution over the entire interval where the problem is formulated.

Let us fix n, define then $h = 1/n$ and set $N = 4n+4$; we can then consider the grid over [0, 1] given by $x_i = ih$, $i = 0, ..., n$. We approximate the solution of the problem (1) as the sum of B-splines of order 8 as follows

$$(4) \qquad u_N\,(x) = \sum_{i=1}^{4n+4} \alpha_i B_i\,(x)\,.$$

Notice that the nodes needed for the construction of the $B-$spline are $\{0, 0, 0, 0, 0, 0, 0, 0, h, h, h, h, 2h, 2h, 2h, 2h, \ldots, (n-1)h, (n-1)h, (n-1)h, (n-1)h, 1, 1, 1, 1, 1, 1, 1, 1\}$.

Let us now consider $\theta_j,\ j = 1, ..., 4$, the zeros of the Legendre polynomial of degree 4. Under the linear map

$$\tau_{ij} = \frac{h}{2}\theta_j + \frac{x_i + x_{i-1}}{2}, \quad i = 1, ..., n,\ j = 1, ..., 4$$

we construct their images $\tau_{ij} \in [x_{i-1}, x_i]$. This is the set of collocation nodes required by the numerical scheme. To obtain a square system for the $4n+4$ unknowns α_i, the $4n$ collocation equations need to be supplemented by the discretized boundary conditions (2).

Letting $\alpha \equiv (\alpha_1, \ldots, \alpha_{4n+4})^t$, and setting for $i = 1, ..., n,\ j = 1, ..., 4$,

$$\mathbf{F} = (U_{00}, U_{01}, f(\tau_{11}), f(\tau_{12}), ..., f(\tau_{ij}), ..., f(\tau_{n3}), f(\tau_{n4}), U_{11}, U_{10})^t,$$

we can write

$$L_h\alpha \equiv [M_4 + h^2M_2 + h^4M_0]\alpha = h^4\mathbf{F} \tag{5}$$

with $M_k \in \mathbb{R}^{(n+4)\times(n+4)}, k = 0, 2, 4$, where the index of each matrix is related to the order of the derivative from which it stems. The system thus obtained is highly structured, in block bidiagonal form. Indeed, for $k = 0, 2, 4$, $\tilde{T}_j^{(k)} \in \mathbb{R}^{2\times4},\ j = 0, 1$, $A_j^{(k)} \in \mathbb{R}^{4\times4},\ j = 0, 1$, $B_j^{(k)} \in \mathbb{R}^{4\times4},\ j = 2, \ldots, n$, $C_j^{(k)} \in \mathbb{R}^{4\times4},\ j = 1, \ldots, n-1$, we have explicitly

$$M_k = \begin{bmatrix} \tilde{T}_0^{(k)} & O_{2,4} & O_{2,4} & & O_{2,4} & O_{2,4} & & O_{2,4} & O_{2,4} & O_{2,4} \\ A_0^{(k)} & C_1^{(k)} & O & & O & O & & O & O & O \\ O & B_2^{(k)} & C_2^{(k)} & \ldots & O & O & \ldots & O & O & O \\ O & O & \ldots & \ldots & O & O & \ldots & O & O & O \\ O & O & O & \ldots & O & O & \ldots & O & O & O \\ O & O & O & \ldots & B_j^{(k)} & C_j^{(k)} & \ldots & O & O & O \\ & & & & \ldots & \ldots & & & & \\ O & O & O & \ldots & O & O & \ldots & B_{n-1}^{(k)} & C_{n-1}^{(k)} & O \\ O & O & O & \ldots & O & O & \ldots & O & B_n^{(k)} & A_1^{(k)} \\ O_{2,4} & O_{2,4} & O_{2,4} & \ldots & O_{2,4} & O_{2,4} & \ldots & O_{2,4} & O_{2,4} & \tilde{T}_1^{(k)} \end{bmatrix}$$

Unless otherwise stated, or when without a specific size index, each block is understood to be 4 by 4. Also, to emphasize the dimension of the zero matrix we write $O_m \in \mathbb{R}^{m\times m}$ or $O_{m,n} \in \mathbb{R}^{m\times n}$.

Specifically, for M_4 we have for $T_j \in \mathbb{R}^{2\times2},\ j = 0, 1$,

$$\tilde{T}_0 \equiv \tilde{T}_0^{(4)} = \begin{bmatrix} T_0 & O_2 \end{bmatrix} \qquad \tilde{T}_1 \equiv \tilde{T}_1^{(4)} = \begin{bmatrix} O_2 & T_1 \end{bmatrix} \tag{6}$$

with

$$T_0 = \begin{bmatrix} h^4 & 0 \\ -7h^3 & 7h^3 \end{bmatrix}, \qquad T_1 = \begin{bmatrix} -7h^3 & 7h^3 \\ 0 & h^4 \end{bmatrix} \tag{7}$$

Furthermore for the matrix M_4 all blocks with same name are equal to each other and we set

$$C \equiv C_1^{(4)} = C_2^{(4)} = ... = C_{n-1}^{(4)}, \quad B \equiv B_2^{(4)} = B_3^{(4)} = ... = B_n^{(4)}.$$

For the remaining blocks we explicitly find

$$A_0 \equiv A_0^{(4)} = \begin{bmatrix} 676.898959 & -2556.080843 & 3466.638660 & -1843.444245 \\ 252.6301981 & -637.2153922 & 206.4343097 & 524.0024063 \\ 30.1896807 & 63.1159957 & -181.0553956 & -258.101801 \\ 0.281162 & 10.18023913 & 107.9824229 & 137.5436408 \end{bmatrix} \tag{8}$$

$$(9)\qquad C=\begin{bmatrix} 194.1150595 & 59.18676730 & 2.650495372 & 0.03514515003 \\ -329.0767906 & -47.64856975 & 27.10012948 & 3.773710141 \\ 499.494486 & -120.6542168 & -64.5675240 & 31.57877478 \\ -664.532755 & 709.1160198 & -385.1831009 & 84.61236994 \end{bmatrix}$$

$$(10)\qquad B=\begin{bmatrix} 84.61236994 & -385.1831008 & 709.1160181 & -664.5327536 \\ 31.57877478 & -64.56752375 & -120.6542173 & 499.4944874 \\ 3.773710141 & 27.1001293 & -47.648570 & -329.076791 \\ 0.03514515003 & 2.6504944 & 59.186765 & 194.11506 \end{bmatrix}$$

$$(11)\quad A_1\equiv A_1^{(4)}=\begin{bmatrix} 137.5436422 & 107.9824252 & 10.1802390 & 0.28116115 \\ -258.1018004 & -181.0553970 & 63.1159965 & 30.18968105 \\ 524.0024040 & 206.4343108 & -637.2153932 & 252.6301982 \\ -1843.444246 & 3466.638661 & -2556.080843 & 676.8989596 \end{bmatrix}$$

Two main changes hold for the matrices M_2 and M_0, with respect to M_4; the first lies in the top and bottom corners, where $\tilde{T}_j^{(0)} = \tilde{T}_j^{(2)} = O_{2,4}, \quad j = 0, 1$. They contain then a premultiplication by diagonal coefficient matrices. Namely letting $A_{0,2}, C_2, B_2, A_{1,2}, D_i \in \mathbb{R}^{4\times 4}$, $D_i = diag(a_{i1}, a_{i2}, a_{i3}, a_{i4})$, with $a_{ij} \equiv a(\tau_{ij}), \ j = 1, 2, 3, 4, \ i = 1, 2, ..., n$, we have

$$A_0^{(2)} = D_1 A_{0,2}, \quad A_1^{(2)} = D_n A_{1,2}$$

$$C_i^{(2)} = D_i C_2, \quad i = 1, 2, ..., n-1$$

$$B_i^{(2)} = D_i B_2, \quad i = 2, 3, ..., n$$

where

$$A_{0,2}=\begin{bmatrix} 29.30827273 & -47.68275514 & 9.072282826 & 7.792345494 \\ 5.67012435 & 2.62408902 & -8.50204661 & -6.772789016 \\ 0.16439223 & 1.339974467 & 3.602756629 & 1.87349900 \\ 0.00006780 & 0.004406270526 & 0.1127212947 & 1.392658748 \end{bmatrix}$$

$$C_2=\begin{bmatrix} 1.450129518 & 0.05858914701 & 0.001126911401 & 0.8471353553\ 10^{-5} \\ 4.030419911 & 2.533012603 & 0.3966407043 & 0.02054903207 \\ -9.65902448 & -0.812682181 & 2.782318888 & 0.7087655468 \\ 4.07133215 & -8.31463385 & -0.93008649 & 3.663534093 \end{bmatrix}$$

$$B_2=\begin{bmatrix} 3.663534093 & -0.9300864851 & -8.314633880 & 4.071332233 \\ 0.7087655468 & 2.782318883 & -0.812682182 & -9.659024508 \\ 0.02054903207 & 0.396640665 & 2.53301254 & 4.0304199 \\ 0.8471353553\ 10^{-5} & 0.00112689 & 0.0585890 & 1.4501302 \end{bmatrix}$$

$$A_{1,2}=\begin{bmatrix} 1.392658814 & 0.112721477 & 0.00440599 & 0.000067777 \\ 1.873498986 & 3.602756584 & 1.33997443 & 0.164392258 \\ -6.772789012 & -8.502046610 & 2.62408899 & 5.670124369 \\ 7.792345496 & 9.072282833 & -47.68275513 & 29.30827274 \end{bmatrix}$$

Similarly, for $A_{0,0}, C_0, B_0, A_{1,0}, E_i \in \mathbb{R}^{4\times 4}$, $E_i = diag(b_{i1}, b_{i2}, b_{i3}, b_{i4})$, with $b_{ij} \equiv b(\tau_{ij})$, $j = 1, 2, 3, 4$, $i = 1, 2, ..., n$, we have

$$A_0^{(0)} = E_1 A_{0,0}, \quad A_1^{(0)} = E_n A_{1,0}$$

$$C_i^{(0)} = E_i C_0, \quad i = 1, 2, ..., n-1$$

$$B_i^{(0)} = E_i B_0, \quad i = 2, 3, ..., n$$

with

$$A_{0,0} = \begin{bmatrix} 0.604278729 & 0.3156064435 & 0.07064438205 & 0.008784901454 \\ 0.060601115 & 0.2089471273 & 0.3087560066 & 0.2534672883 \\ 0.000426270 & 0.006057945090 & 0.03689680420 & 0.1248474545 \\ 0.10\,10^{-7} & 0.7425933886\,10^{-6} & 0.00002933256459 & 0.0006554638258 \end{bmatrix}$$

$$C_0 = \begin{bmatrix} 0.0006703169101 & 0.00001503946986 & 0.1853647586\,10^{-6} & 0.9723461945\,10^{-9} \\ 0.1448636180 & 0.02163722179 & 0.001674337031 & 0.00005328376522 \\ 0.4676572160 & 0.2815769859 & 0.07496220012 & 0.007575139336 \\ 0.1985435495 & 0.4197299375 & 0.3055061349 & 0.07553484124 \end{bmatrix}$$

$$B_0 = \begin{bmatrix} 0.07553484124 & 0.3055061345 & 0.4197299367 & 0.1985435448 \\ 0.007575139336 & 0.07496219992 & 0.2815769862 & 0.4676572138 \\ 0.00005328376522 & 0.00167433770 & 0.0216372202 & 0.144863633 \\ 0.9723461945\,10^{-9} & 0.1836\,10^{-6} & 0.000015047 & 0.00067030 \end{bmatrix}$$

$$A_{1,0} = \begin{bmatrix} 0.00065546187 & 0.00002934342 & 0.72976\,10^{-6} & 0.78\,10^{-8} \\ 0.1248474495 & 0.03689679808 & 0.00605794290 & 0.0004262700 \\ 0.2534672892 & 0.3087560073 & 0.2089471283 & 0.0606011146 \\ 0.00878490146 & 0.07064438202 & 0.3156064438 & 0.6042787300 \end{bmatrix}$$

In the next Section also some more information on some of the above matrices will be needed, specifically we have

$$\begin{aligned} \|A_1\|_2 &\equiv a_1^* = 0.0321095, \\ \|B^{-1}\|_2 &\equiv b_1^* = 0.1022680, \\ \rho(B^{-1}) &\equiv b_2^* = 0.0069201. \end{aligned} \tag{12}$$

3. Error analysis

We begin by stating two Lemmas which will be needed in what follows.

LEMMA 1. *The spectral radius of any permutation matrix P is $\rho(P) = 1$ and $\|P\|_2 = 1$.*

Proof. Indeed notice that it is a unitary matrix, as it is easily verified that $P^{-1} = P^* = P$, or that $P^*P = I$, giving the second claim. Moreover, since $\rho(P^*) \equiv \rho(P^{-1}) = \rho(P) = \rho(P)^{-1}$, we find $\rho^2(P) = 1$, i.e. the first claim.

□

LEMMA 2. *Let us introduce the auxiliary diagonal matrix of suitable dimension* $\Delta_m = diag\left(1, \delta^{-1}, \delta^{-2}, ..., \delta^{1-m}\right)$ *choosing* $\delta < 1$ *arbitrarily small. We can consider also the vector norm defined by* $\|x\|_* \equiv \|\Delta x\|_2$ *together with the induced matrix norm* $\|A\|_*$. *Then, denoting by* $\rho(A) \equiv \max_{1 \le i \le n} |\lambda_i^{(A)}|$ *the spectral radius of the matrix A, where* $\lambda_i^{(A)},\ i = 1(1)n$ *represent its eigenvalues, we have*

$$\|A\|_* \le \rho(A) + O(\delta), \quad \|\Delta^{-1}\|_2 = 1.$$

Proof. The first claim is a restatement of Theorem 3, [9] p. 13. The second one is immediate from the definition of Δ.

□

Let y_N be the unique B-spline of order 8 interpolating to the solution U of problem (1). If $f \in C^4([0, 1])$ then $U \in C^8([0, 1])$ and from standard results, [4], [15] we have

$$\|D^j(U - y_N)\|_\infty \le c_j h^{8-j}, \quad j = 0, \dots, 7. \tag{13}$$

We set

$$y_N(x) = \sum_{j=1}^{4n+4} \beta_j B_j(x). \tag{14}$$

The function u_N has coefficients that are obtained by solving (5); we define the function $\mathbf{G}$ as the function obtained by applying the very same operator of (5) to the spline y_N, namely

$$\mathbf{G} \equiv h^{-4} L_h \beta \equiv h^{-4}[M_4 + h^2 M_2 + h^4 M_0]\beta. \tag{15}$$

Thus $\mathbf{G}$ differs from $\mathbf{F}$ in that it is obtained by a different combination of the very same B-splines.

Let us introduce the discrepancy vector $\sigma_{ij} \equiv \mathbf{G}(\tau_{ij}) - \mathbf{F}(\tau_{ij}),\ i = 1(1)n,\ j = 1(1)4$ and the error vector $\mathbf{e} \equiv \beta - \alpha$, with components $e_i = \beta_i - \alpha_i,\ i = 1, \dots, 4n+4$. Subtraction of (5), from (15) leads to

$$[M_4 + h^2 M_2 + h^4 M_0]\mathbf{e} = h^4 \sigma. \tag{16}$$

We consider at first the dominant systems arising from (5), (15), i.e.

$$M_4 \tilde{\alpha} = h^4 \mathbf{F}, \quad M_4 \tilde{\beta} = h^4 \mathbf{G}. \tag{17}$$

Subtraction of these equations gives the dominant equation corresponding to (16), namely

$$M_4 \tilde{\mathbf{e}} = h^4 \sigma, \quad \tilde{\mathbf{e}} \equiv \tilde{\alpha} - \tilde{\beta}. \tag{18}$$

Notice first of all, that in view of the definition of **G** and of the fact that y_N interpolates on the exact data of the function, the boundary conditions are the same both for (5) and (15). Hence $\sigma_1 = \sigma_2 = \sigma_{4n+3} = \sigma_{4n+4} = 0$. In view of the triangular structure of T_0 and T_1, it follows then that $\tilde{e}_1 = \tilde{e}_2 = \tilde{e}_{4n+3} = \tilde{e}_{4n+4} = 0$, a remark which will be confirmed more formally later.

We define the following block matrix, corresponding to block elimination performed in a peculiar fashion, so as to annihilate all but the first and last element of the second block row of M_4

$$\tilde{R} = \begin{bmatrix} I_2 & O_{2,4} & O_{2,4} & O_{2,4} & \dots & O_{2,4} & \dots & O_{2,4} & O_{2,4} & O_2 \\ O_{4,2} & I_4 & Q & Q^2 & \dots & Q^{j-2} & \dots & Q^{n-2} & Q^{n-1} & O_{4,2} \\ O_{4,2} & O & I_4 & O & \dots & O & \dots & O & O & O_{4,2} \\ & & & & \dots & & & & & \\ O_{4,2} & O & O & O & \dots & O & \dots & I_4 & O & O_{4,2} \\ O_{4,2} & O & O & O & \dots & O & \dots & O & I_4 & O_{4,2} \\ O_2 & O_{2,4} & O_{2,4} & O_{2,4} & \dots & O_{2,4} & \dots & O_{2,4} & O_{2,4} & I_2 \end{bmatrix}$$

where $Q = -CB^{-1}$. Recall once more our convention for which the indices of the identity and of the zero matrix denote their respective dimensions and when omitted each block is understood to be 4 by 4. Introduce the block diagonal matrix $\tilde{A}^{-1} = diag(I_{4n}, A_1^{-1})$. Observe then that $\tilde{R}M_4\tilde{A}^{-1} = \tilde{M}_4$, with

$$\tilde{M}_4 = \begin{bmatrix} \tilde{T}_0 & O_{2,4} & O_{2,4} & O_{2,4} & \dots & O_{2,4} & O_{2,4} & O_{2,4} \\ A_0 & O & O & O & \dots & O & O & Q^{n-1} \\ O & B & C & O & \dots & O & O & O \\ O & O & B & C & \dots & O & O & O \\ & & & & \dots & & & \\ O & O & O & O & \dots & B & C & O \\ O & O & O & O & \dots & O & B & I_4 \\ O_{2,4} & O_{2,4} & O_{2,4} & O_{2,4} & \dots & O_{2,4} & O_{2,4} & \tilde{T}_1 A_1^{-1} \end{bmatrix}$$

Let us consider now the singular value decomposition of the matrix Q, $Q = V\Lambda U^*$, [12]. Here $\Lambda = diag(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ is the diagonal matrix of the singular values of Q, *ordered from the largest to the smallest*. Now, premultiplication of $\tilde{M}_4$ by $S = diag(I_2, V^*, I_{4n-2})$ and then by the block permutation matrix

$$\tilde{P} = \begin{bmatrix} I_2 & O_{2,4} & O_{2,4n-4} & O_{2,2} \\ O_{4,2} & I_4 & O_{4,4n-4} & O_{4,2} \\ O_{2,2} & O_{2,4} & O_{2,4n-4} & I_2 \\ O_{4n-4,2} & O_{4n-4,4} & I_{4n-4} & O_{4n-4,2} \end{bmatrix}$$

followed by postmultiplication by $\tilde{S} = diag(I_{4n}, U)$ and then by

$$\hat{P} = \begin{bmatrix} I_4 & O & O_{4,4n-4} \\ O_{4n-4,4} & O_{4n-4,4} & I_{4n-4} \\ O & I_4 & O_{4,4n-4} \end{bmatrix}$$

gives the block matrix

$$\bar{E} = \begin{bmatrix} \tilde{E} & O_{8,4n-4} \\ \tilde{L} & \tilde{B} \end{bmatrix}. \tag{19}$$

Here

$$\tilde{E} = \begin{bmatrix} \tilde{T}_0 & O_{2,4} \\ V^* A_0 & \Lambda^{n-1} \\ O_{2,4} & \tilde{T}_1 A_1^{-1} U \end{bmatrix} \tag{20}$$

and

$$\tilde{L} = \begin{bmatrix} O_{4n-8,4} & O_{4n-8,4} \\ O_4 & U \end{bmatrix} \tag{21}$$

as well as

$$\tilde{B} = \begin{bmatrix} B & C & O & O & \dots & O & O \\ O & B & C & O & \dots & O & O \\ O & O & B & C & \dots & O & O \\ & & & & \dots & & \\ O & O & O & O & \dots & B & C \\ O & O & O & O & \dots & O & B \end{bmatrix}. \tag{22}$$

It is then easily seen that

$$\bar{E}^{-1} = \begin{bmatrix} \tilde{E}^{-1} & O_{8,4n-4} \\ -\tilde{B}^{-1}\tilde{L}\tilde{E}^{-1} & \tilde{B}^{-1} \end{bmatrix}. \tag{23}$$

In summary, we have obtained $\bar{E} = \tilde{P} S \tilde{R} M_4 \tilde{A}^{-1} \tilde{S} \hat{P}$. It then follows $M_4 = \tilde{R}^{-1} S^{-1} \tilde{P} \bar{E} \hat{P} \tilde{S}^{-1} \tilde{A}$, and in view of Lemma 1, system (18) becomes

$$\bar{E} \hat{P} \tilde{S}^{-1} \tilde{A} \tilde{\mathbf{e}} = h^4 \tilde{P} S \tilde{R} \sigma. \tag{24}$$

To estimate the norm of $\bar{E}^{-1}$ exploiting its triangular structure (19), we concentrate at first on (20). Recalling the earlier remark on the boundary data, we can partition the error from (16) and the discrepancy vectors as follows: $\tilde{\mathbf{e}} = (\tilde{e}_1, \tilde{e}_2, \tilde{\mathbf{e_t}}, \tilde{\mathbf{e_c}}, \tilde{\mathbf{e_b}}, \tilde{e}_{4n+3}, \tilde{e}_{4n+4})^T$, $\tilde{\mathbf{e_t}}, \tilde{\mathbf{e_b}}, \in \mathbb{R}^2$, $\tilde{\mathbf{e_c}}, \in \mathbb{R}^{4n-4}$. Define also $\tilde{\mathbf{e_{out}}} = (\tilde{\mathbf{e_t}}, \tilde{\mathbf{e_b}})^T$, $\hat{\mathbf{e}}_{\mathbf{out}} = (0, 0, \tilde{\mathbf{e_{out}}}, 0, 0)^T$, $\hat{\mathbf{e_t}} = (e_1, e_2, \tilde{\mathbf{e_t}})^T$, $\hat{\mathbf{e_b}} = (\tilde{\mathbf{e_b}}, e_{4n+3}, e_{4n+4})^T$.

Now introduce the projections Π_1, Π_2 corresponding to the top and bottom portions of the matrix (19). Explicitly, they are given by the following matrices

$$\Pi_1 = \begin{bmatrix} I_8 & O_{8,4n-4} \end{bmatrix} \quad \Pi_2 = \begin{bmatrix} O_{4n-4,8} & I_{4n-4} \end{bmatrix}. \tag{25}$$

Consider now the left hand side of the system (24). It can be rewritten in the following fashion

$$\Pi_1 \bar{E} \hat{P} \tilde{S}^{-1} \tilde{A} \tilde{\mathbf{e}} = \tilde{E} \hat{P} \tilde{S}^{-1} \tilde{A} \tilde{\mathbf{e}} = \tilde{E} \begin{bmatrix} \hat{\mathbf{e_t}} \\ U^* A_1 \hat{\mathbf{e_b}} \end{bmatrix} \tag{26}$$

The matrix in its right hand side $Z \equiv \Pi_1 \tilde{P} S \tilde{R}$ instead becomes

(27)

$$Z = \begin{bmatrix} I_2 & O_{2,4} & O_{2,4} & \dots & O_{2,4} & \dots & O_{2,4} & O_{2,4} & O_2 \\ O_{4,2} & V^* & V^*Q & \dots & V^*Q^{j-2} & \dots & V^*Q^{n-2} & V^*Q^{n-1} & O_{4,2} \\ O_2 & O_{2,4} & O_{2,4} & \dots & O_{2,4} & \dots & O_{2,4} & O_{2,4} & I_2 \end{bmatrix}.$$

From (26) using (20), we find

(28)
$$\tilde{E}\begin{bmatrix} \hat{\mathbf{e}}_{\mathbf{t}} \\ U^* A_1 \hat{\mathbf{e}}_{\mathbf{b}} \end{bmatrix} = \begin{bmatrix} \tilde{T}_0 \hat{\mathbf{e}}_{\mathbf{t}} \\ V^* A_0 \hat{\mathbf{e}}_{\mathbf{t}} + \Lambda^{n-1} U^* A_1 \hat{\mathbf{e}}_{\mathbf{b}} \\ \tilde{T}_1 \hat{\mathbf{e}}_{\mathbf{b}} \end{bmatrix} = \begin{bmatrix} T_0 \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \\ V^* A_0 \hat{\mathbf{e}}_{\mathbf{t}} + \Lambda^{n-1} U^* A_1 \hat{\mathbf{e}}_{\mathbf{b}} \\ T_1 \begin{pmatrix} e_{4n+3} \\ e_{4n+4} \end{pmatrix} \end{bmatrix}.$$

Introduce now the following matrix

$$H = \begin{bmatrix} -96.42249156 & 409.2312351 & \lambda_1^{n-1} & 0 \\ -162.6192900 & 738.3915192 & 0 & \lambda_2^{n-1} \\ 264.5383512 & -1216.139747 & 0 & 0 \\ 645.9124120 & -2179.906392 & 0 & 0 \end{bmatrix},$$

where the first two columns are the last two columns of $V^* A_0$. The matrix of the system can then be written as

$$\tilde{E} \equiv R_1^{-1} R_1 \begin{bmatrix} T_0 & O_{2,4} & O_2 \\ Y_0 & H & Y_1 \\ O_2 & O_{2,4} & T_1 \end{bmatrix} \begin{bmatrix} I_4 & O_4 \\ O_4 & U^* A_1 \end{bmatrix}$$

$$= R_1^{-1} \begin{bmatrix} T_0 & O_{2,4} & O_2 \\ Y_0 & \tilde{\Lambda}(I + N_1) & Y_1 \\ O_2 & O_{2,4} & T_1 \end{bmatrix} P^\dagger \begin{bmatrix} I_2 & O_{2,4} \\ O_{2,4} & [U^* A_1]_{1,2} \\ I_2 & O_{2,4} \\ O_{2,4} & [U^* A_1]_{3,4} \end{bmatrix} \equiv R_1^{-1} \bar{\Lambda} P^\dagger \bar{P} S,$$

where we introduced the permutation $P^\dagger$ exchanging the first two with the last two columns of the matrix H, its inverse producing a similar operation on the rows of the matrix to its right; we have denoted the first two rows of such matrix by $[U^* A_1]_{1,2}$ and a similar notation has been used on the last two. R_1 denotes the 8 by 8 matrix corresponding to the elementary row operation zeroing out the element (4, 2) of H, i.e. the element (6, 4) of $\tilde{E}$. Thus $R_1 H P_1$ is upper triangular, with main diagonal given by $\tilde{\Lambda} \equiv diag(\lambda_1^{n-1}, \lambda_2^{n-1}, r, s)$, $\lambda_1 = 5179.993642 > 1$, $\lambda_2 = 11.40188637 > 1$. It can then be written then as $R_1 H P_1 = \tilde{\Lambda}(I + N_1)$, with N_1 upper triangular and nilpotent.

The inverse of the above matrix $\bar{\Lambda}$ is then explicitly given by

$$\bar{\Lambda}^{-1} \equiv \begin{bmatrix} T_0^{-1} & O_{2,4} & O_2 \\ -T_0^{-1}(I+N_1)^{-1}\tilde{\Lambda}^{-1}Y_0 & (I+N_1)^{-1}\tilde{\Lambda}^{-1} & -T_1^{-1}(I+N_1)^{-1}\tilde{\Lambda}^{-1}Y_1 \\ O_2 & O_{2,4} & T_1^{-1} \end{bmatrix}$$

where $\tilde{N}_1$ denotes a nilpotent upper triangular matrix.

From (20) and the discussion on the boundary conditions the top portion of this system gives for the right hand side $h^4 Z\sigma = h^4 [0,0,\sigma_c,0,0]^T$. Thus from $\bar{\Lambda}^{-1}Z\sigma$ gives immediately $e_1 = e_2 = e_{4n+3} = e_{4n+4} = 0$ as claimed less formally earlier. The top part of the dominant system then simplifies by removing the two top and bottom equations, as well as the corresponding null components of the error and right hand side vectors. Introduce also the projection matrix $\Pi_3 = diag(\mathbf{0}_2, I_4, \mathbf{0}_2)$, where $\mathbf{0}_m$ denotes the null vector of dimension m. We then obtain

$$\hat{\mathbf{e}}_{\mathbf{out}} = \Pi_3\hat{\mathbf{e}}_{\mathbf{out}} = h^4\Pi_3\bar{\Lambda}^{-1}Z\sigma_c = h^4\Pi_3 S\bar{P}P^{\dagger}(I+N_1)^{-1}\tilde{\Lambda}^{-1}R_1\Pi_1\tilde{P}S\tilde{R}\sigma_c$$

from which letting $\lambda^{\dagger} \equiv \max(\lambda_1^{1-n}, \lambda_2^{1-n}, r^{-1}, s^{-1}) = \max(r^{-1}, s^{-1})$, the estimate follows using Lemmas 1 and 2

$$\begin{aligned} \|\hat{\mathbf{e}}_{\mathbf{out}}\|_* &\le h^4\|\Pi_3\|_*\|S\|_*\|\bar{P}\|_*\|P^{\dagger}\|_*\|(I+N)^{-1}\|_*\|\tilde{\Lambda}^{-1}\|_* \\ &\quad \|R_1\|_*\|\Pi_1\tilde{P}S\Delta\Delta^{-1}\tilde{R}\Delta\Delta^{-1}\sigma_c\|_* \\ &\le h^4\lambda^{\dagger}(1+O(\delta))^4[\rho(S)+O(\delta)][\rho(R_1)+O(\delta)]\| \\ &\quad \Pi_1\tilde{P}S\Delta\|_*\|\Delta^{-1}\tilde{R}\Delta\|_*\|\Delta^{-1}\sigma_c\|_* \\ &\le h^4\lambda^{\dagger}(1+O(\delta))^6\|\Pi_1\tilde{P}S\Delta\|_*\|(I+\tilde{R}_2)\|_*\|\Delta\Delta^{-1}\sigma_c\|_2 \\ &\le h^4\lambda^{\dagger}(1+O(\delta))^7\|\Pi_1\tilde{P}S\Delta\|_*\sqrt{4n-4}\|\sigma_c\|_\infty \end{aligned} \tag{29}$$

as $\tilde{R}_2$ is upper triangular and nilpotent. Now observe that the product $\tilde{P}S\Delta = diag(D_1, V^*D_2, D_3, D_4)$, where each block is as follows

$$\begin{aligned} D_1 &= diag(1, \delta^{-1}),\ D_2 = diag(\delta^{-2}, \delta^{-3}, \delta^{-4}, \delta^{-5}), \\ D_3 &= diag(\delta^{-8}, \delta^{-9}),\ D_4 = diag(\delta^{-10}, \dots, \delta^{-4n-3}, \delta^{-6}, \delta^{-7}). \end{aligned}$$

It follows that $\Pi_1\tilde{P}S\Delta = diag(D_1, V^*D_2, D_3, \mathbf{0}_{4n-4})$. Hence

$$\begin{aligned} \|\Pi_1\tilde{P}S\Delta_{4n+4}\|_* &= \|\|diag(I_2, V^*, I_2)\ diag(D_1, D_2, D_3)\|_* \\ &\le \|diag(I_2, V^*, I_2)\|_*\|diag(D_1, D_2, D_3)\|_* \\ &\le [\rho(diag(I_2, V^*, I_2)) + O(\delta)](1+O(\delta)) \le (1+O(\delta))^2 \end{aligned} \tag{30}$$

since for the diagonal matrix $\rho[diag(D_1, D_2, D_3)] = 1$ and from Lemma 1 $\rho(V^*) = 1$, the matrix V being unitary. But also,

$$\|\hat{\mathbf{e}}_{\mathbf{out}}\|_*^2 = \|\Delta_4\hat{\mathbf{e}}_{\mathbf{out}}\|_2^2 = \hat{\mathbf{e}}_{\mathbf{out}}^*\Delta_4^2\hat{\mathbf{e}}_{\mathbf{out}} = \sum_{i=1}^{4} e_i^2\delta^{2i-8} \ge \|\hat{\mathbf{e}}_{\mathbf{out}}\|_\infty^2$$

i.e. $\|\hat{\mathbf{e}}_{\mathbf{out}}\|_* \geq \|\hat{\mathbf{e}}_{\mathbf{out}}\|_\infty$. In summary combining (29) with (30) we have

$$\begin{aligned} \|\hat{\mathbf{e}}_{\mathbf{out}}\|_\infty &\leq h^{\frac{7}{2}} 2\lambda^\dagger (1 + O(\delta))^9 \|\sigma_c\|_\infty \leq h^{\frac{7}{2}} 2\lambda^\dagger (1 + O(\delta)) \|\sigma_c\|_\infty \\ &\equiv h^{\frac{7}{2}} \eta \|\sigma\|_\infty \qquad (31) \end{aligned}$$

which can be restated also as $h^{-4}\|\tilde{E}\mathbf{e}_{\mathbf{out}}\|_\infty \geq [\eta h^{\frac{7}{2}}]^{-1}\|\mathbf{e}_{\mathbf{out}}\|_\infty$ i.e. from Thm. 4.7 of [10], p. 88, the estimate on the inverse follows

$$\|\tilde{E}^{-1}\|_\infty \leq \eta n^{\frac{1}{2}}.$$

Looking now at the remaining part of (18) with the bottom portion matrix of $\bar{E}$, see (19), we can rewrite it as $\tilde{B}\tilde{\mathbf{e}}_{\mathbf{c}} = \sigma_c - \tilde{L}\hat{\mathbf{e}}_{\mathbf{out}}$. We have $\tilde{B} = E\hat{B}$, with $\hat{B} = diag(B, \ldots, B)$ and

$$E = \begin{bmatrix} I & -Q & O & O & \ldots & O & O \\ O & I & -Q & O & \ldots & O & O \\ O & O & I & -Q & \ldots & O & O \\ & & & & \ldots & & \\ O & O & O & O & \ldots & I & -Q \\ O & O & O & O & \ldots & O & I \end{bmatrix} \qquad (32)$$

and thus $\tilde{B}^{-1} = \hat{B}^{-1}E^{-1}$. Notice that E^{-1} is a block upper triangular matrix, with the block main diagonal containing only identity matrices, it can then be written as $E^{-1} = I_{4n-4} + U_0$, U_0 being nilpotent (i.e. block upper triangular with zeros on the main diagonal). Thus Lemma 2 can be applied once more. The system can then be solved to give

$$\tilde{\mathbf{e}}_{\mathbf{c}} = \hat{B}^{-1}E^{-1}[h^4\sigma - \tilde{L}\hat{\mathbf{e}}_{\mathbf{out}}].$$

Premultiplying this system by Δ^{-1} and taking norms, we obtain using (29),

$$\begin{aligned} \|\Delta^{-1}\tilde{\mathbf{e}}_{\mathbf{c}}\|_* &\leq h^4\|\Delta^{-1}\hat{B}^{-1}E^{-1}\sigma\|_* + \|\Delta^{-1}\hat{B}^{-1}E^{-1}\tilde{L}\hat{\mathbf{e}}_{\mathbf{out}}\|_* \\ &\leq h^4\|\Delta\Delta^{-1}\hat{B}^{-1}E^{-1}\sigma\|_2 + \|\Delta^{-1}\|_*\|\hat{B}^{-1}\|_*\|E^{-1}\|_*\|U\hat{\mathbf{e}}_{\mathbf{out}}\|_* \\ &\leq h^4\|\hat{B}^{-1}\|_2\|E^{-1}\sigma\|_2 + [\rho(\hat{B}^{-1}) + O(\delta)][1 + O(\delta)]\|U\|_*\|\hat{\mathbf{e}}_{\mathbf{out}}\|_* \\ &\leq h^4\|B^{-1}\|_2\sqrt{4n-4}\|E^{-1}\sigma\|_\infty + \rho(B^{-1})[1 + O(\delta)]^3\eta h^{\frac{7}{2}} \\ &\leq h^4 b_1^* 2\sqrt{n}\|E^{-1}\|_\infty\|\sigma\|_\infty + \rho(B^{-1})[1 + O(\delta)]\eta h^{\frac{7}{2}}\|\sigma\|_\infty \\ &\leq h^{\frac{7}{2}} 2b_1^* e_\infty^* \|\sigma\|_\infty + b_2^*[1 + O(\delta)]\eta h^{\frac{7}{2}}\|\sigma\|_\infty \\ &\leq h^{\frac{7}{2}}\left[2b_1^* e_\infty^* + b_2^*[1 + O(\delta)]\eta\right]\|\sigma\|_\infty \equiv h^{\frac{7}{2}}\mu\|\sigma\|_\infty \,. \qquad (33) \end{aligned}$$

On the other hand

$$\|\Delta^{-1}\tilde{\mathbf{e}}_{\mathbf{c}}\|_* = \|\Delta\Delta^{-1}\tilde{\mathbf{e}}_{\mathbf{c}}\|_2 = \|\tilde{\mathbf{e}}_{\mathbf{c}}\|_2 \geq \|\tilde{\mathbf{e}}_{\mathbf{c}}\|_\infty.$$

In summary, by recalling (12) and since $\|E^{-1}\|_\infty \equiv e_\infty^* = 72.4679$

$$\|\tilde{\mathbf{e}}_{\mathbf{c}}\|_\infty \leq \mu n^{-\frac{7}{2}}\|\sigma\|_\infty.$$

Together with the former estimate (31) on $\|\hat{\mathbf{e}}_{\mathbf{out}}\|_\infty$, we then have

$$\|\tilde{\mathbf{e}}\|_\infty \le \nu n^{-\frac{7}{2}} \|\sigma\|_\infty,$$

which implies, once again from Thm. 4.7 of ([10]), $h^{-4}\|M_4\tilde{\mathbf{e}}\|_\infty \ge \nu^{-1} n^{-\frac{1}{2}} \|\tilde{\mathbf{e}}\|_\infty$, i.e. in summary we can state the result formally as follows.

THEOREM 1. *The matrix M_4 is nonsingular. The norm of its inverse matrix is given by*

$$\|M_4^{-1}\|_\infty \le \nu n^{\frac{1}{2}}. \tag{34}$$

Now, upon premultiplication of (16) by the inverse of M_4, letting $N \equiv M_4^{-1}(M_2 + h^2 M_0)$, we have

$$\mathbf{e} = h^4 (I + h^2 N)^{-1} M_4^{-1} \sigma. \tag{35}$$

As the matrices M_2 and M_0 have entries which are bounded above, since they are built using the coefficients a and b, which are continuous functions on [0, 1], i.e. themselves bounded above, Banach's lemma, [12] p. 431, taking h sufficiently small, allows an estimate of the solution as follows.

$$\|\mathbf{e}\|_\infty \le h^4 \|(I + h^2 N)^{-1}\|_\infty \|M_4^{-1}\|_\infty \|\sigma\|_\infty \le \frac{h^4 \nu \|\sigma\|_\infty n^{\frac{1}{2}}}{1 - h^2\|N\|_\infty} \le \gamma n^{-\frac{7}{2}} \|\sigma\|_\infty, \tag{36}$$

having applied the previous estimate (34). Observe that

$$\|u_N - y_N\|_\infty \le \|\mathbf{e}\|_\infty \max_{0 \le x \le 1} \sum_{i=0}^{4n+4} B_i(x) \le \theta \|\mathbf{e}\|_\infty.$$

Applying again (13) to σ, using the definition (5) of L_h, we find for $1 \le k \le n$, $j = 1(1)4$, by the continuity of the functions $\mathbf{F}$, $\mathbf{G}$

$$|\sigma_{4k+j}| = h^4 |\mathbf{G}(\tau_{k,j}) - \mathbf{F}(\tau_{k,j})| \le \zeta_{k,j} h^4. \tag{37}$$

It follows then $\|\sigma\|_\infty \le \zeta h^4$ and from (36), $\|\mathbf{e}\|_\infty \le \gamma h^{\frac{15}{2}}$. Taking into account this result, use now the triangular inequality as follows

$$\|U - u_N\|_\infty \le \|U - y_N\|_\infty + \|y_N - u_N\|_\infty \le c_0 h^8 + \eta\gamma h^{\frac{15}{2}} \le c^* h^{\frac{15}{2}}$$

in view of (13) and (36). Hence, recalling that $N = 4n + 4$, we complete the error analysis, stating in summary the convergence result as follows

THEOREM 2. *If $f \in \mathrm{C}^4([0, 1])$, so that $U \in \mathrm{C}^8([0, 1])$ then the proposed B-spline collocation method (5) converges to the solution of (1) in the Chebyshev norm; the convergence rate is given by*

$$\|U - u_N\|_\infty \le c^* N^{-\frac{15}{2}}. \tag{38}$$

REMARK 1. The estimates we have obtained are not sharp and in principle could be improved.

4. Examples

We have tested the proposed method on several problems. In the Figures we provide the results of the following examples. They contain the semilogarithmic plots of the error, in all cases for $n = 4$, i.e. $h = .25$. In other words, they provide the number of correct significant digits in the solution.

EXAMPLE 1. We consider the equation

$$y^{(4)} - 3y^{(2)} - 4y = 4\cosh(1),$$

with solution $y = \cosh(2x - 1) - \cosh(1)$.

EXAMPLE 2. Next we consider the equation with the same operator L but with different, variable right hand side

$$y^{(4)} - 3y^{(2)} - 4y = -6\exp(-x),$$

with solution $y = \exp(-x)$.

EXAMPLE 3. Finally we consider the variable coefficient equation

$$y^{(4)} - xy^{(2)} + y\sin(x) = \frac{24}{(x+3)^5} - \frac{2x}{(x+3)^3} + \frac{\sin(x)}{x+3},$$

with solution $y = \frac{1}{x+3}$.

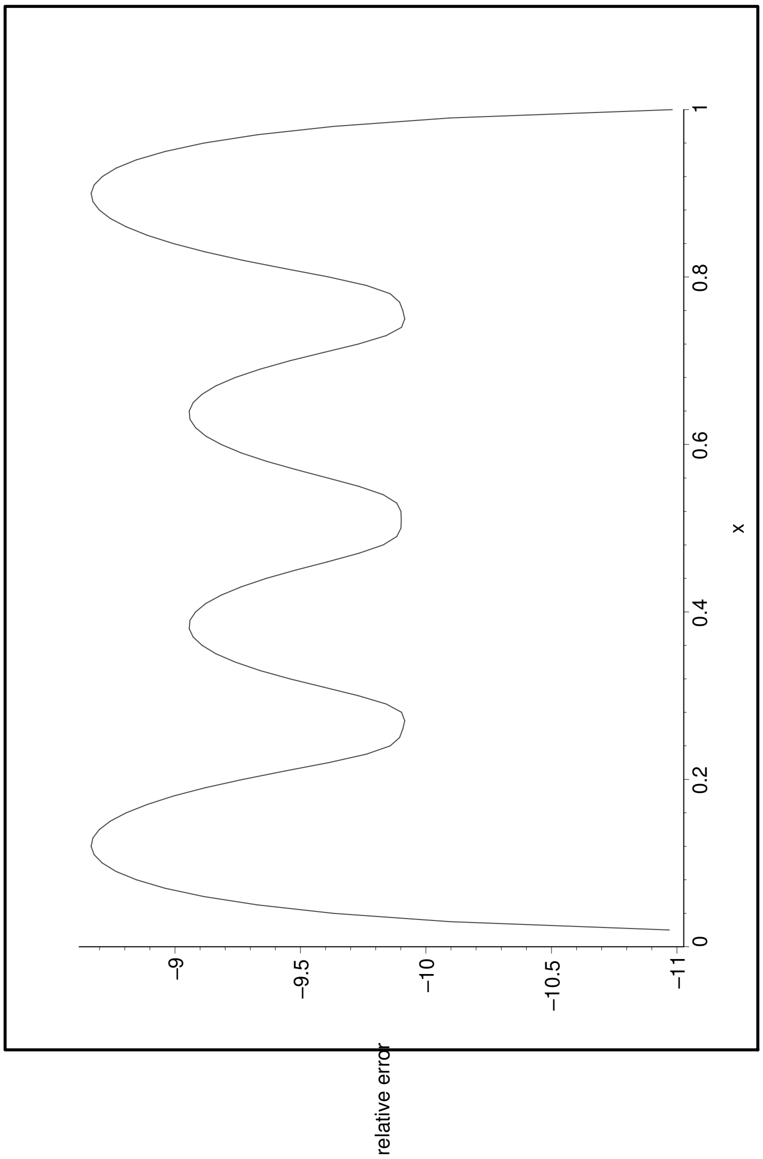


Figure 1: Semilogarithmic graph of the relative error for Example 1.

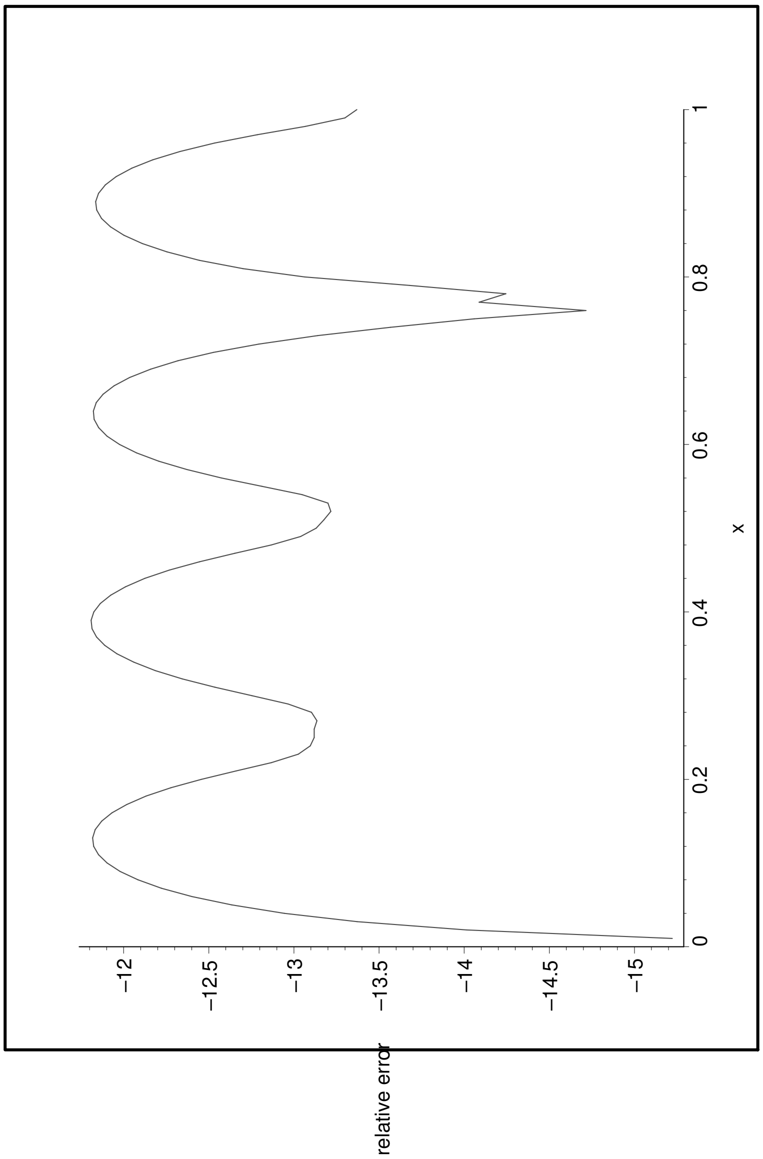


Figure 2: Semilogarithmic graph of the relative error for Example 2.

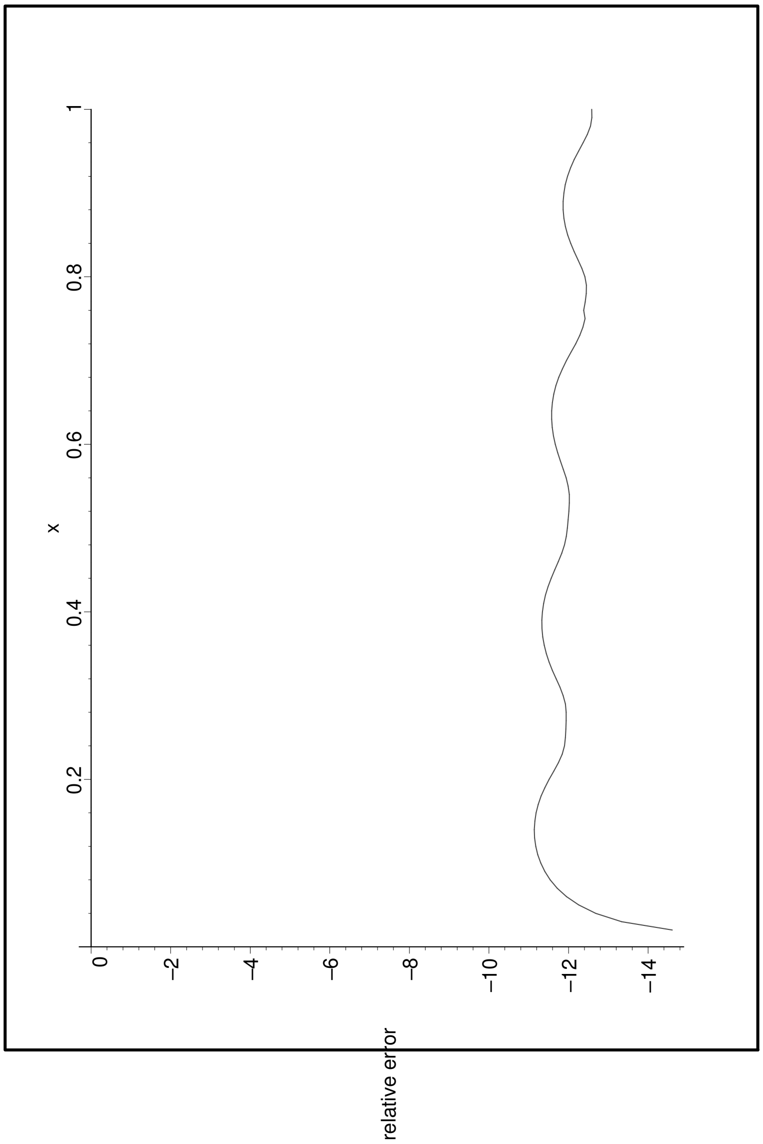


Figure 3: Semilogarithmic graph of the relative error for Example 3.

References

[1] BRUNNER H., *Implicit Runge-Kutta Nystrom methods for general Volterra integro-differential equations*, Computers Math. Applic. **14** (1987), 549–559.

[2] BRUNNER H., *The approximate solution of initial value problems for general Volterra integro-differential equations*, Computing **40** (1988), 125–137.

[3] DAHLQUIST G., BJÖRK Å. AND ANDERSON D., *Numerical methods*, Prentice-Hall 1974.

[4] DE BOOR C. AND SWARTZ B., *Collocation at Gaussian points*, SIAM J. Num. Anal. **10** (1973), 582–606.

[5] DE BOOR C., *A practical guide to splines*, Springer Verlag, New York 1985.

[6] FAWZY T., *Spline functions and the Cauchy problems. II. Approximate solution of the differential equation $y'' = f(x, y, y')$ with spline functions*, Acta Math. Hung. **29** (1977), 259–271.

[7] GAREY L.E. AND SHAW R.E., *Algorithms for the solution of second order Volterra integro-differential equations*, Computers Math. Applic. **22** (1991), 27–34.

[8] GOLDFINE A., *Taylor series methods for the solution of Volterra integral and integro-differential equations*, Math. Comp. **31** (1977), 691–707.

[9] ISAACSON E. AND KELLER H.B., *Analysis of numerical methods*, Wiley, New York 1966.

[10] LINZ, *Theoretical Numerical Analysis*, Wiley, New York 1979.

[11] LINZ, *Analytical and numerical methods for Volterra integral equations*, SIAM, Philadelphia 1985.

[12] NOBLE B., *Applied Linear Algebra*, Prentice-Hall 1969.

[13] PITTALUGA G., SACRIPANTE L. AND VENTURINO E., *Higher order lacunary interpolation*, Revista de la Academia Canaria de Ciencias **9** (1997), 15–22.

[14] PITTALUGA G., SACRIPANTE L. AND VENTURINO E., *Lacunary interpolation with arbitrary data of high order*, Ann. Univ. Sci. Budapest Sect. Comput. **XX** (2001), 83–96.

[15] PRENTER P.M., Splines and variational methods, Wiley, New York 1989.

[16] VENTURINO E. AND SAXENA A., *Smoothing the solutions of history-dependent dynamical systems*, Numerical Optimization **19** (1998), 647–666.

AMS Subject Classification: 65L05, 45J05, 65L99, 45L10.

Giovanna PITTALUGA, Laura SACRIPANTE, Ezio VENTURINO
Dipartimento di Matematica
Universita' di Torino
via Carlo Alberto 10
10123 Torino, ITALIA
e-mail: giovanna.pittaluga@unito.it
laura.sacripante@unito.it
ezio.venturino@unito.it

Rend. Sem. Mat. Univ. Pol. Torino
Vol. 61, 3 (2003)
Splines and Radial Functions

M.L. Sampoli

CLOSED SPLINE CURVES BOUNDING MAXIMAL AREA

Abstract. In this paper we study the problem of constructing a closed spline curve in $\mathbb{R}^2$, which interpolates a given set of data points, is shape-preserving and which, in addition, bounds the maximal area. The construction is done by using the so-called *Abstract Schemes* (AS). The resulting spline curve, expressed in its piecewise Beźier representation, has degree 3 and continuity C^1 and can be extended to a curve of degree 6 and continuity C^2, with similar properties.

1. Introduction

In the definition and development of mathematical models which could describe real objects or real phenomena a great deal of research has been done in the field of *constrained interpolation* and approximation. Constrained interpolation indeed arises in various applications. In industry, for instance, when we are dealing with the problem of designing the network curves constituting the tail of an aircraft we should avoid any oscillations which could affect the aerodynamic properties of the resulting surface. In these cases we give additional constraints such as *smoothing* constraints or *shape-preserving* constraints. In this context, an interesting problem, important for its applications in naval engineering and ship industry, is that of constructing a closed curve in $\mathbb{R}^2$, interpolating a given set of points, shape-preserving and bounding maximal area.

Aim of the paper is indeed to present a method to solve this problem. The proposed method is based on the application of Abstract Schemes. These schemes have been developed to solve general constrained interpolation problems (see for instance [4], [17], [7]).

The basic idea behind AS is given by observing that when we interpolate some data points by a spline, and want to fulfill other requirements, we usually dispose of several free parameters $\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_N$ ($\mathbf{d}_i \in \mathbb{R}^q$), which are associated with the knots. If we now express the constraints as conditions relative to each interval between two knots, they can be rewritten as a sequence of inclusion conditions: $(\mathbf{d}_i, \mathbf{d}_{i+1}) \in D_i \subset \mathbb{R}^{2q}$, where the sets D_i are the corresponding feasible domains. In this setting the problems of existence, construction and selection of an optimal solution can be studied with the help of Set Theory in a general way.

The remainder of the paper is organized as follows. In the next section the fundamental ideas of AS will be recalled. In order to make the paper self-contained it is more convenient to present here the basic ideas and the main results on AS, although they can be found in other papers as well. In Section 3 we shall present the application of AS for the construction of C^1 cubic curves with maximal area. The C^2 spline curves will be constructed in Section 4 and the examples and the final conclusions will be reported in Section 5.

2. Basic ideas on abstract schemes

The main idea which gave rise to abstract schemes was the observation that most methods used in constrained interpolation have a common structure, even if, at a first sight, they seem quite different from each other. This structure can be sketched as follows: first a suitable set of piecewise functions is chosen and a set of parameters $(\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_N)$ is selected. Then, each function piece is expressed using these parameters. A further step consists in rewriting the constraints in terms of these parameters and deriving a set of *admissible domains* D_i. This common procedure is thus exploited to build up a general theory, in order to check the feasibility of the problem, and then to provide a general purpose algorithm for computing a solution (given by an *optimal* sequence).

The first abstract schemes were developed some years ago by Schmidt and independently by Costantini, see for instance [16] and [3] (and the survey papers [5], [17]). In those papers an abstract algorithm to construct univariate functions subjected to separable constraints was presented.

Let us see now a rigorous formulation. Supposing we are given the sequences of sets $D_i \subset \mathbb{R}^q \times \mathbb{R}^q$, with $D_i \neq \emptyset$, for every $i = 0, 1, ..., N-1$, which define the constraint domains, we may define the *global* set

$$\mathbf{D} := \{(\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_N) \in \mathbb{R}^{q\ (N+1)} \quad s.t.\ (\mathbf{d}_i, \mathbf{d}_{i+1}) \in D_i,\ i = 0, 1, \ldots, N-1\}.$$

This is the solution set. Therefore a problem of constrained interpolation can be suitably reduced to the study of set $\mathbf{D}$. Indeed we shall consider the following problems

P1 *Is* $\mathbf{D}$ *non empty? In other words do there exist sequences* $(\mathbf{d}_0, \mathbf{d}_1, ..., \mathbf{d}_N)$ *such that*

$$(\mathbf{d}_i, \mathbf{d}_{i+1}) \in D_i,\ i = 0, 1, \ldots, N-1. \tag{1}$$

P2 *If there exist sequences fulfilling (1), is it possible to build up an algorithm which computes one among them efficiently?*

Obviously, if the solution is not unique we will select *the best one*, that is the sequence which will minimize or maximize some objective functional (we will see that in our case the functional is related with the area).

We assume the problem is *well posed*, in the sense that the sets D_i are supposed non empty for every i; we suppose also that D_i are closed sets, for every i (not necessarily compact). Moreover in general, as the solution is always made up piecewisely, we shall adopt the same notation used for piecewise curves: we call *i-th segment* the portion of the solution from the i-th to the $i+1$-st breakpoint.

A solution to these problems can be obtained using a two-sweep strategy ([4]), processing the data first in one *direction*, for instance *from left to right*, through algorithm A1 (forward sweep), and then in the opposite direction, through algorithm A2 (backward sweep). In more detail, let us denote with $\Pi_1,\ \Pi_2 : \mathbb{R}^q \times \mathbb{R}^q \to \mathbb{R}^q$ the projection maps from the "x_1x_2-plane" onto the "x_1-axis" and "x_2-axis" respectively, and let us define the sets

$$B_i := \Pi_1(D_i)\ ;\quad i = 0, 1, \ldots, N-1\ ; \qquad B_N := \mathbb{R}^q\ . \tag{2}$$

Now, in the forward sweep, as we may observe that for every parameter $\mathbf{d}_i$ either the constraint domain coming from the segment $(i-1, i)$, that is D_{i-1}, or the one coming from the segment $(i, i+1)$, that is D_i, have to be taken into account, we determine, for every parameter, the *true* admissible domain A_i. This is indeed done by algorithm A1.

Algorithm A1.

1. *Set* $A_0 := B_0$, $J := N$

2. *For* $i = 1, \ldots, N$

 2.1 *Set* $A_i := \Pi_2(D_{i-1} \cap \{A_{i-1} \times B_i\})$.

 2.2 *If* $A_i = \emptyset$ *set* $J := i$ *and stop.*

3. *Stop.*

In this connection, we have the following result, [4].

THEOREM 1. *P1 has a solution if, and only if,* $J = N$ *that is* $A_i \neq \emptyset$, $i = 0, 1, \ldots, N$. *If* $(\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_N)$ *is a solution then*

$$\mathbf{d}_i \in A_i\ ;\ i = 0, 1, \ldots, N. \tag{3}$$

We remark that, in general, a solution of P1 is not unique and that the necessary condition (3) is not sufficient. Thus, if the sequence of non empty sets $A_0, \ldots, A_N$ has been defined by algorithm A1, a first simple scheme for computing a sequence $(\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_N)$ is provided by the following algorithm (backward sweep) whose effectiveness is guaranteed by Theorem 2 (we refer again to [4] for the proof).

Algorithm A2.

1. *Choose any* $\mathbf{d}_N \in A_N$.

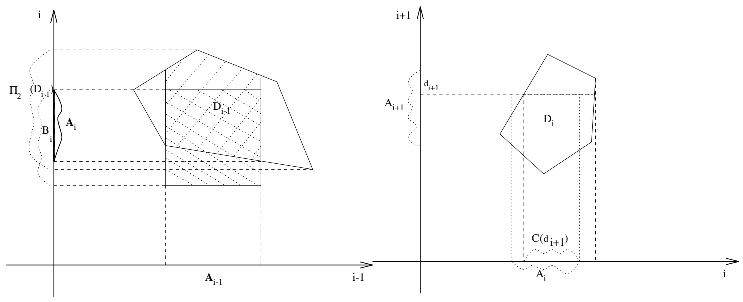
2. *For* $i = N-1, N-2, \ldots, 0$

 2.1 *set* $C_i(\mathbf{d}_{i+1}) := \Pi_1(D_i \cap \{A_i \times \{\mathbf{d}_{i+1}\}\})$

 2.2 *Choose any* $\mathbf{d}_i \in C_i(\mathbf{d}_{i+1})$

3. *Stop.*

THEOREM 2. *Let the sequence* $A_0, A_1, \ldots, A_N$ *be given by algorithm A1, with* $A_i \neq \emptyset$; $i = 0, 1, \ldots, N$. *Then algorithm A2 can be completed (that is the sets* $C(\mathbf{d}_{i+1})$ *are not empty) and any sequence* $(\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_N)$ *computed by algorithm A2 is a solution for problem P2.*



(a) (b)

Figure 1: Algorithm 1: construction of the admissible domains A_i,(a). Algorithm 2: graphical sketch of step 2.1.(b).

2.1. Boundary conditions

When we are constructing closed curves the end points should be handled as the other inner points so that the solution $\mathbf{s}(t)$ has to satisfy the condition $\mathbf{s}^{(k)}(x_0) = \mathbf{s}^{(k)}(x_N)$ for $k = 0, 1, \ldots$ up to the continuity order considered. This kind of conditions are called *non separable* boundary conditions. In terms of abstract scheme formalization the above conditions reduce to find a sequence $(\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_N)$ such that $\mathbf{d}_N = \beta(\mathbf{d}_0)$, where β is any continuous function with continuous inverse.

These conditions, giving a direct relationship between the first and last element of the sequence $(\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_N)$, destroy the sequential structure of our scheme. A possible strategy would consist in considering, among all the sequences $(\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_N)$ which belong to $\mathbf{D}$, the ones where starting with an element $\mathbf{d}_0 \in A_0$, end up with $\mathbf{d}_N = \beta(\mathbf{d}_0)$. In other words we first check whether there are in the admissible domain A_0 some $\mathbf{d}_0$ such that $\beta(\mathbf{d}_0) \in A_N$ or, equivalently ,whether $\beta(A_0) \cap A_N \neq \emptyset$. Then, in the backward sweep we pick up one of these $\mathbf{d}_N$ and go back ending in $\mathbf{d}_0 = \beta^{-1}(\mathbf{d}_N)$.

A full development of this procedure, along with the conditions under which such sequences exist and the related constructive algorithms to determine them, would require some theory of set-valued maps whose details are beyond the scope of this paper. We refer to [4] for theoretical aspects and to [6], [8] for an idea of some practical applications.

2.2. Selecting an optimal solution

It is clear from algorithms A1 and A2 that it is possible to find infinite sequences $(\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_N)$ satisfying the constraints, as in algorithm A2 the admissible sets $C_i(\mathbf{d}_{i+1})$, defined in step 2.1, do not reduce, in general, to a single point. It is therefore a natural idea to look for an *optimal* sequence, where the optimality criterion can be given as the maximum or the minimum of a suitable functional F that is

$$\max_{(\mathbf{d}_0,\mathbf{d}_1,\ldots,\mathbf{d}_N)\in\mathbf{D}} F(\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_N) \,, \tag{4}$$

and we specify the functional F according to our requirements. Although several forms of functionals could be considered, for the sake of simplicity, we shall limit ourselves to

$$F(\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_N) := \sum_{i=0}^{N-1} g_i(\mathbf{d}_i, \mathbf{d}_{i+1}), \tag{5}$$

where g_i gives the local contribution to the objective function and, in shape-preserving problems, can be connected with the shape of the resulting function (we shall consider the area bounded by a closed curve).

To solve the optimization problem we present here an approach based on dynamic programming (DP) [2]. As we will see later, this approach is well suited to deal with discrete problems. Moreover DP is extremely flexible, as many functionals and any kind of separable constraints (i.e. constraints which can be related to only one curve segment and then whic can be expressed *separately* from segment to segment) can be processed using the same algorithmic structure and, unlike other optimization methods, constraints play here a positive role, limiting the size of the *decision space*. In this regard, we may observe that the functional recurrence relations of dynamic programming can be very efficiently linked with the constraints in Algorithm A2. We refer to [9] for full details on how implement dynamic programming in Algorithm A2.

Below is reported a sketch of the algorithm where we have stored in Φ_i the cost associated with the i-th stage and in T_i is stored the optimal policy (therefore $\max_{(\mathbf{d}_0,\mathbf{d}_1,\ldots,\mathbf{d}_N)\in\mathbf{D}} F(\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_N) = \max_{\mathbf{d}_N} \Phi_N(\mathbf{d}_N)$). As a consequence, starting with the optimal $\mathbf{d}_N$, we obtain the optimal $\mathbf{d}_{N-1} := T_{N-1}(\mathbf{d}_N)$ and so on.

Algorithm A2DP.

1. *For any* $\delta_0 \in A_0$ *set* $\Phi_0(\delta_0) := 0$

2. *For* $i = 1, 2, \ldots, N$

2.1 *For any* $\delta_i \in A_i$ *compute* $C_{i-1}(\delta_i) := \Pi_1(D_{i-1} \cap \{A_{i-1} \times \{\delta_i\}\})$

2.2 *For any* $\delta_i \in A_i$ *compute* $\Phi_i(\delta_i) := \max_{\delta_{i-1}\in C_{i-1}(\delta_i)} (g(\delta_{i-1}, \delta_i) + \Phi_{i-1}(\delta_{i-1})) = g(T_{i-1}(\delta_i), \delta_i) + \Phi_{i-1}(T_{i-1}(\delta_i))$ *and the corresponding optimizing value* $T_{i-1}(\delta_i)$

3. *Compute* $\mathbf{d}_N$ *such that* $\Phi_N(\mathbf{d}_N) = \max_{\delta_N \in A_N} \Phi_N(\delta_N)$

4. *For* $i = N-1, \ldots, 0$

4.1. $\mathbf{d}_i := T_i(\mathbf{d}_{i+1})$

5. *Stop.*

2.3. Multivariate case

The two-sweep scheme, given by algorithms A1 and A2 (or A2DP), has turned out to be an effective method to solve several problems and its main attraction relies in the fact that it is *general*, being applicable to a wide range of problems.

However an its closer inspection shows us that, although there is no assumption on the subsets D_i of $\mathbb{R}^q \times \mathbb{R}^q$, which are the basic elements of the imposed constraints, the practical usage of this method has been so far confined to the case $D_i \in \mathbb{R} \times \mathbb{R}$, that means $q = 1$. This is due to the fact that in algorithms, either A1 or A2 we have to compute the projection of intersections of subsets in a product space. More precisely, we may recall, for instance, that step 2.1 of A1, which is the kernel of all the modifications and improvements later developed, requires the computation of the following set:

$$A_i := \Pi_2(D_{i-1} \cap \{A_{i-1} \times B_i\}),$$

and this leads, even in the simplest higher dimension case, that is $q = 2$, to intersections and projections of arbitrary subsets of $\mathbb{R}^2 \times \mathbb{R}^2$. Even in the case of linear inequalities for the constraints (D_i would be a polytope of $\mathbb{R}^4$), the corresponding algorithm is extremely difficult to implement and has an unaffordable computational cost. Indeed, in $\mathbb{R}^q$, the computational cost of set intersections and their projections is given by $O(n^{q-1} \; log \; n)$, where n is the number of polytope vertices, see [15] for full details.

Thus, the practical application of abstract schemes has been for many years restricted to *univariate problems*, where we have only one parameter associated with every knot (two for every segment). This limitation is rather restrictive as univariate problems suffice in general to model interpolation of *functions*, while are not suitable for interpolation of *parametric curves*, which can represent closed curves. We may see that usually parametric planar curve interpolation gives rise already to constraint domains in $\mathbb{R}^2 \times \mathbb{R}^2$.

Recent research has been therefore devoted to develop a new theory and construct new methods suitable and applicable to *multivariate* constraint problems (see for instance [7], [14]). It is worthwhile to repeat that the dimension q of the parameter space is not related to the dimension of the point space we are working in.

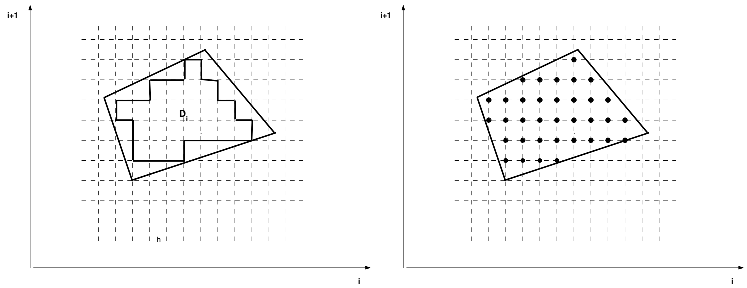
Recently a new approach has been proposed (see [9], [8]). It is based on the observation that if we consider the union of $2q$*-boxes* (i.e. rectangular parallelepipeds with facets parallel to the coordinate hyperplanes) the computational cost of their intersections and projections is reduced to $O(n \; log^{q-1} \; n)$, [15]. The basic simple idea of the new method is that of approximating the constraint domains D_i with a union of $2q$-boxes $\widetilde{D}_i$

We refer to [9] for full details on this new approach. For the sake of completeness we report here the main ideas on how this approximation is performed.

For every domain D_i, we suppose we are able to give an estimate of a lower and/or upper bound for each dimension. Then we may choose a step size $\mathbf{h} = (h_1, h_2, \ldots, h_q)$ and, starting from the lower (upper) bound, construct a multidimensional grid in $\mathbb{R}^q \times \mathbb{R}^q$ whose dimension is assigned. We thus approximate every domain D_i with the union of those boxes whose vertices are contained in D_i. By construction we have $\widetilde{D}_i \subseteq D_i$ and we may easily see that, given $\mathbf{h}$, for $\mathbf{h} \to \mathbf{0}$, we have $meas\ (D_i \setminus \widetilde{D}_i) \to 0$.

The next step consists in making a further approximation. Once we have obtained the domains $\widetilde{D}_i$, we consider only the discrete values for the parameters $(\mathbf{d}_i, \mathbf{d}_{i+1})$, corresponding to the vertices of the considered boxes. This is equivalent to working with discrete domains, which we denote by $\overline{D}_i$. We then select the points of the grid which are vertices of a $2q$-box contained in D_i. At the end of this process we obtain a sequence of domains $\overline{D}_i$ such that again approximate D_i and $\overline{D}_i \subseteq D_i$.

As in the continuous case, we may select an optimal solution by optimizing a suitable functional, using dynamic programming. The fact that the parameters $\mathbf{d}_i$ vary in discrete domains is well suited for applying the dynamic programming in the minimization process. Regarding the convergence analysis, the following result holds (we refer to [9] for the proof).



(a) (b)

Figure 2: (a):Every domain D_i is replaced by a union of $2q$-boxes.(b) only the values at the vertices of the considered boxes are taken.

THEOREM 3. *Let the domains $D_0, D_1, \ldots, D_{N-1}$, with $D_i \subset \mathbb{R}^q \times \mathbb{R}^q$, be given. Let $\overline{D}_0, \overline{D}_1, \ldots, \overline{D}_{N-1}$ be the corresponding discrete domains obtained with a grid of step size* $\mathbf{h}$. *Let us denote now with $(\mathbf{d}_0^*, \mathbf{d}_1^*, \ldots, \mathbf{d}_N^*)$ a solution in* $\mathbf{D}$ *which maximizes also a continuous functional F, with a unique absolute maximum, and let $(\bar{\mathbf{d}}_0^*, \bar{\mathbf{d}}_1^*, \ldots, \bar{\mathbf{d}}_N^*)$ be a discrete counterpart. Then*

$$\lim_{h_{max} \to 0} (\bar{\mathbf{d}}_0^*, \bar{\mathbf{d}}_1^*, \ldots, \bar{\mathbf{d}}_N^*) = (\mathbf{d}_0^*, \mathbf{d}_1^*, \ldots, \mathbf{d}_N^*)\ , \quad h_{max} := \max(h_1, h_2, \ldots, h_q)\ .$$

We remark that as the parameters $(\mathbf{d}_i, \mathbf{d}_{i+1})$ can assume only the discrete values corresponding to non zero elements of the i-th logical matrix, the operations of intersection, projection, cartesian product, etc. are easily performed on the matrix by the logical operators AND, OR, taking only some planes, putting together more planes and so on. This way of proceeding has revealed to be very effective from the computational point of view and it can be extended straightforwardly to domains in $\mathbb{R}^q \times \mathbb{R}^q$ (the number of planes is in general given by $2q$).

3. Interpolating spline curves maximizing the bounded area

Let us suppose we are given a set of points $\mathbf{I}_i \in \mathbb{R}^2$, $i = 0, 1, \ldots, N$, along with a parameterization t_i $i = 0, 1, \ldots, N$, (for instance chord length). Let us indicate with $\mathbf{L}_i := \mathbf{I}_{i+1} - \mathbf{I}_i$ the polygonal data and set the quantities $k_i := t_{i+1} - t_i$, $\widetilde{\mathbf{L}}_i := (\mathbf{I}_{i+1} - \mathbf{I}_i)/k_i$.

Our goal is to construct an interpolating C^1 spline curve, $\mathbf{s}(t)$, which bounds maximal area. In order to obtain a curve useful for the applications we consider the additional constraint of shape preservation, i.e. we require that the resulting curve preserves the convexity and inflections as prescribed by the given data.

To express the spline curve we use the piecewise Bézier representation so that each spline segment is a Bézier curve

$$\mathbf{s}_i(t) := \sum_{j=0}^{n} \mathbf{b}_{j,i} B_{n,j}(t) , \qquad t \in [t_i, t_{i+1}] ,$$

where $\mathbf{b}_{j,i}$ are called *control points* and form the *control polygon* and $B_{n,j}(t)$ are the Bernstein polynomials of degree n, defined by $B_{n,j}(u) = \left(\begin{array}{c} n \\ j \end{array} \right) (1-u)^{n-j} u^j$, and $u = \frac{t - t_i}{t_{i+1} - t_i}$.

From the properties of Bézier splines we have that the resulting curve will be uniquely determined once the control polygon for every segment is constructed, see for instance [11]. By construction we have that $\mathbf{b}_{0,i} = \mathbf{I}_i$ and $\mathbf{b}_{n,i} = \mathbf{I}_{i+1}$, in this way, as Bézier curves pass through the first and the last control points, we are guaranteed the resulting curve interpolates the given data.

Let us consider now curves of degree three. In this case, for each segment i, the control points to be determined are $\mathbf{b}_{0,i}$, $\mathbf{b}_{1,i}$, $\mathbf{b}_{2,i}$, and $\mathbf{b}_{3,i}$. The first and the last points are set equal to two interpolation data points. So we have to determine the two inner control points. From cubic Bézier polynomial properties we have that the global curve is C^1 if and only if, for every segment i, the inner control points $\mathbf{b}_{1,i}$ and $\mathbf{b}_{2,i}$ are taken as follows

$$\mathbf{b}_{1,i} := \mathbf{I}_i + \frac{k_i}{3}\mathbf{T}_i, \quad \mathbf{b}_{2,i} := \mathbf{I}_{i+1} - \frac{k_i}{3}\mathbf{T}_{i+1}, \tag{6}$$

where the vectors $\mathbf{T}_i$ and $\mathbf{T}_{i+1}$, are respectively the curve tangent vectors at interpolation points $\mathbf{I}_i$ and $\mathbf{I}_{i+1}$.

If we express the tangent vectors as

$$\mathbf{T}_i \ := \ u_i \widetilde{\mathbf{L}}_{i-1} + v_i \widetilde{\mathbf{L}}_i \ ; \qquad i = 0, \ldots, N. \tag{7}$$

where we set $\mathbf{I}_{-1} = \mathbf{I}_{N-1}$, $t_{-1} = t_{N-1}$, and $\mathbf{I}_{N+1} = \mathbf{I}_1$, $t_{N+1} = t_1$, the free parameters to be determined in order to uniquely construct the resulting curve are the values of u_i and v_i, for every segment $i = 0, \ldots, N-1$. These being two dimensional vectors, the problem is called *bivariate*.

It is a standard practice to express the shape-preserving conditions through the control points. Indeed, due to the *variation diminishing* property of cubic Bézier polynomials, a sufficient condition to locally reproduce the convexity of the polygonal data is that the corresponding control polygon is convex as well. This condition can be given by the following relationships

$$(\widetilde{\mathbf{L}}_{i-1} \wedge \widetilde{\mathbf{L}}_i) \cdot ((\mathbf{b}_{1,i} - \mathbf{I}_i) \wedge (\mathbf{b}_{2,i} - \mathbf{b}_{1,i})) \geq 0$$
$$(\widetilde{\mathbf{L}}_i \wedge \widetilde{\mathbf{L}}_{i+1}) \cdot ((\mathbf{b}_{2,i} - \mathbf{b}_{1,i}) \wedge (\mathbf{I}_{i+1} - \mathbf{b}_{2,i})) \geq 0 \,. \tag{8}$$

We have to reformulate the shape-preserving conditions in terms of parameters u_i and v_i. The resulting curve should belong to the portion plane between $\mathbf{L}_i$ and the prolongations of $\mathbf{L}_{i-1}$ and $\mathbf{L}_{i+1}$; this can be assured imposing this condition to the tangent vectors giving rise to the condition $u_i, v_i \geq 0$. Moreover, a necessary condition for (8) (which can be easily deduced from the limit case of collinear data points) is given by

$$0 \leq u_i \leq 3, \quad 0 \leq v_i \leq 3 \,. \tag{9}$$

Let us then define the quantities $\rho_i := \|(\widetilde{\mathbf{L}}_{i-1} \wedge \widetilde{\mathbf{L}}_i)\|$ and $\sigma_i := \|(\widetilde{\mathbf{L}}_{i-1} \wedge \widetilde{\mathbf{L}}_{i+1})\|$, for $i = 0, \ldots, N$. Using (7) the conditions (8) can be rewritten, after straightforward computations, as

$$\begin{cases} u_i(3 - u_{i+1})\rho_i^2 - u_i v_{i+1}\, \rho_i\sigma_i - v_i v_{i+1}\rho_i\rho_{i+1} \geq 0 \\[2ex] v_{i+1}(3 - v_i)\rho_{i+1}^2 - u_i v_{i+1}\, \rho_{i+1}\sigma_i - u_i u_{i+1}\rho_i\rho_{i+1} \geq 0 \,. \end{cases} \tag{10}$$

From the above expressions we get immediately the form of the constraints domains,

$$\begin{aligned} D_i := \{ \quad & (u_i, v_i, u_{i+1}, v_{i+1}) \in \mathbb{R}^2 \times \mathbb{R}^2 \quad \text{such that} \\ & u_i(3 - u_{i+1})\rho_i^2 - u_i v_{i+1}\, \rho_i\sigma_i - v_i v_{i+1}\rho_i\rho_{i+1} \geq 0\,; \\ & v_{i+1}(3 - v_i)\rho_{i+1}^2 - u_i v_{i+1}\, \rho_{i+1}\sigma_i - u_i u_{i+1}\rho_i\rho_{i+1} \geq 0\,; \\ & u_i, v_i, u_{i+1}, v_{i+1} \;\geq\; 0\}. \end{aligned} \tag{11}$$

3.1. Optimization process

Our goal is to select as optimal solution the one which maximize a suitable functional related to the area bounded by the curve (applying Algorithm A2DP).

Therefore we have to compute the area of the region bounded by a parametric closed curve. We observe that such a region can be divided into two parts: the region enclosed in the polygonal line connecting the data points and the region between the polygonal line and the spline curve. As we consider *interpolating* spline curves the region bounded by the polygonal line is fixed, therefore we may restrict our attention to maximizing the area (which can be also negative) between the polygonal line and the curve.

The construction of the curve is done segment by segment (piecewisely), we may then maximize for each segment the area S_i bounded by the curve $s(t)$ for $t \in [t_i, t_{i+1}]$ and the line connecting $\mathbf{I}_i$ and $\mathbf{I}_{i+1}$. Using polar coordinates we may see that the area S_i can be expressed in the following way

$$S_i := \int_{t_i}^{t_{i+1}} (x(t)\frac{dy}{dt} - y(t)\frac{dx}{dt})dt \tag{12}$$

where $x(t)$ and $y(t)$ are respectively the x and y components of the parametric curve $s(t)$. We should remark, that in this case the data points have to be ordered anticlockwise. The global functional we maximize is then given by

$$F(\mathbf{d}_0, \mathbf{d}_1, \ldots, \mathbf{d}_N) := \sum_{i=0}^{N-1} S_i \ . \tag{13}$$

Regarding the error we make in the discretisation process, it can be proven the following result (for the proof we refer again to [8]).

THEOREM 4. *For every segment* $[i, i+1]$*, let us define the discrete and the continuous optimal solutions respectively*

$$\bar{s}_i^* := \bar{s}^*_{|[t_i, t_{i+1}]} \ ; \quad s_i^* := s^*_{|[t_i, t_{i+1}]} \ ,$$

where, the first one is computed using the discrete domains whose grid step size is $\mathbf{h}$*, then, setting* $h_{max} = \max(h_1, h_2, \ldots, h_q)$ *and* $k_i = t_{i+1} - t_i$ *we have*

$$\|\bar{s}_i^*(t) - s_i^*(t)\|_\infty \le h_{max} \, \frac{k_i}{4} \ . \tag{14}$$

4. C^2 Continuity

We consider now the problem of constructing spline curves with C^2 continuity. The idea is to start from a C^1 piecewise curve and modify it such that it still satisfies the imposed constraints, has maximal area, and it is also of the required continuity order along the pieces.

Thus, once we have obtained the cubic spline curve, interpolating the data set, convexity preserving and bounding maximal area, with C^1 continuity along the segments, we raise the degree of the resulting spline up to six and work on its Bézier control polygon in order to construct a shape-preserving C^2 interpolating spline with again bounds the maximal area. Following [8] this construction can be performed with the help of abstract schemes.

Let us consider the generic i-th interval. Taking the control polygon of the cubic Bézier curve $\mathbf{I}_i, \mathbf{b}_{1,i}, \mathbf{b}_{2,i}, \mathbf{I}_{i+1}$ and inserting a collinear point in every polygonal segment we obtain a control polygon with seven vertices, namely $\mathbf{I}_i, \mathbf{b}^*_{1,i}, \mathbf{b}^*_{2,i} = \mathbf{b}_{1,i}, \mathbf{b}^*_{3,i}, \mathbf{b}^*_{4,i} = \mathbf{b}_{2,i}, \mathbf{b}^*_{5,i}, \mathbf{I}_{i+1}$, corresponding to a Bézier curve of degree six. By construction, at each end point the curve has second derivative equal to zero, giving rise to a global C^2 spline curve. Moreover, as the shape of the control polygon is not changed the shape-preserving properties are maintained.

For the sake of simplicity the three additional points, $\mathbf{b}^*_{1,i}, \mathbf{b}^*_{3,i}, \mathbf{b}^*_{5,i}$, can be taken as midpoints of each polygonal segment.

The C^1 continuity is guaranteed if we keep the points $\mathbf{b}^*_{1,i}$ and $\mathbf{b}^*_{5,i}$ fixed. On the other hand the requirement of C^2 continuity $\ddot{s}_i(t_{i+1}) = \ddot{s}_{i+1}(t_{i+1})$ can be rewritten in terms of control points, taking into account that for curve of degree six we have (see for instance [11])

$$\ddot{s}_i(t_{i+1}) = \frac{30}{k_i^2}(\mathbf{b}^*_{4,i} + \mathbf{I}_{i+1} - 2\,\mathbf{b}^*_{5,i}) \,, \tag{15}$$

$$\ddot{s}_{i+1}(t_{i+1}) = \frac{30}{k_{i+1}^2}(\mathbf{b}^*_{2,i+1} + \mathbf{I}_{i+1} - 2\,\mathbf{b}^*_{1,i+1}) \,, \tag{16}$$

where, as usual, $k_i = t_{i+1} - t_i$. We recall that the C^2 curve so far constructed corresponds to the limit position $\mathbf{b}^*_{2,i} = \mathbf{b}_{1,i}$ and $\mathbf{b}^*_{4,i} = \mathbf{b}_{2,i}$, with $\mathbf{b}_{1,i}$, $\mathbf{b}_{2,i}$ control points of the cubic curve.

In general the C^2 continuity gives a linear relationship between $\mathbf{b}^*_{2,i+1}$'s and $\mathbf{b}^*_{4,i}$'s. If we keep fixed the points $\mathbf{b}^*_{3,i}$'s, we may choose the points $\mathbf{b}^*_{2,i}$'s as free parameters for $i = 1, 2, \ldots, N-1$ and express the $\mathbf{b}^*_{4,i}$'s according to continuity relations.

$$\mathbf{b}^*_{4,i} := 2\,\mathbf{b}_{5,i} - \mathbf{I}_{i+1} + \frac{k_i^2}{k_{i+1}^2}\left(\mathbf{b}^*_{2,i+1} + \mathbf{I}_{i+1} - 2\,\mathbf{b}^*_{1,i+1}\right) \tag{17}$$

In order to maintain shape-preserving requirements the resulting control polygon should have the same convexity of the initial control polygon. More precisely the point $\mathbf{b}^*_{2,i}$ should belong to the triangle whose vertices are $\mathbf{b}^*_{1,i}$, $\mathbf{b}_{1,i}$ and $\mathbf{b}^*_{3,i}$, and equivalently that $\mathbf{b}^*_{4,i}$ should belong to the triangle given by $\mathbf{b}^*_{3,i}$, $\mathbf{b}_{2,i}$ and $\mathbf{b}^*_{5,i}$.

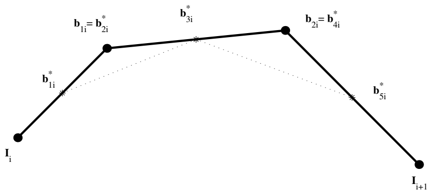


Figure 3: Starting control polygon for C^2 continuity.

Using AS formalization, these requirements lead to constraint domains belonging to $\mathbb{R}^2 \times \mathbb{R}^2$, therefore analogously to what we have done in the previous section, we approximate them with a union of hyperrectangles and take only the discrete values corresponding to the vertices.

Analogously to what we have done in the cubic case, we can now compute, for every segment i, the area bounded by the polygonal line and the curve and maximize it. This procedure lead to an *optimal* C^2 spline curve.

5. Numerical results

In this section we present the performance of the proposed scheme on three examples. In every case, in order to better comment the results obtained the curves will be depicted along with the curvature at the normal direction (*porcupine representation*),

The first example is about a symmetric set of data proposed the first time by Kaklis and Sapidis in [12]. The resulting curves are shown in Figures 4-5, where it is also displayed the polygonal line connecting the data. More precisely in Figure 4 it is shown the resulting C^1 shape-preserving curve which interpolates the given data, and maximize the bounded area.

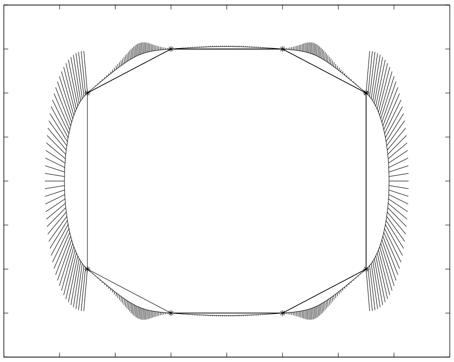


Figure 4: C^1 cubic spline curve.

In Figure 5 the spline curve with C^2 continuity is shown where, again the result is obtained though the maximization of the bounded area. We may note numerically that in this case, having performed a second optimization process the total area is increased.

The second example concerns a set of data taken from a ship hull. The results are displayed in Figure 6. Again we observe experimentally that the spline curve with C^2 continuity has a bounded area which is increased with respect to the area of the corresponding C^1 curve.

As a last example we consider a set of data with one inflection point (which is reproduced in the resulting spline curve). The two resulting curves are shown in Figure 7.

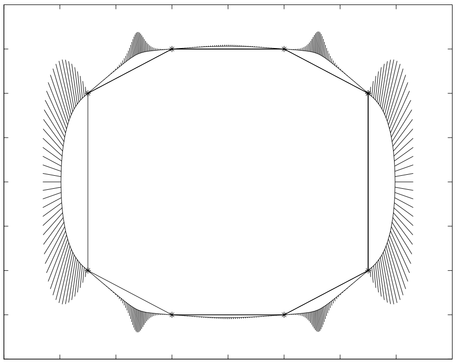
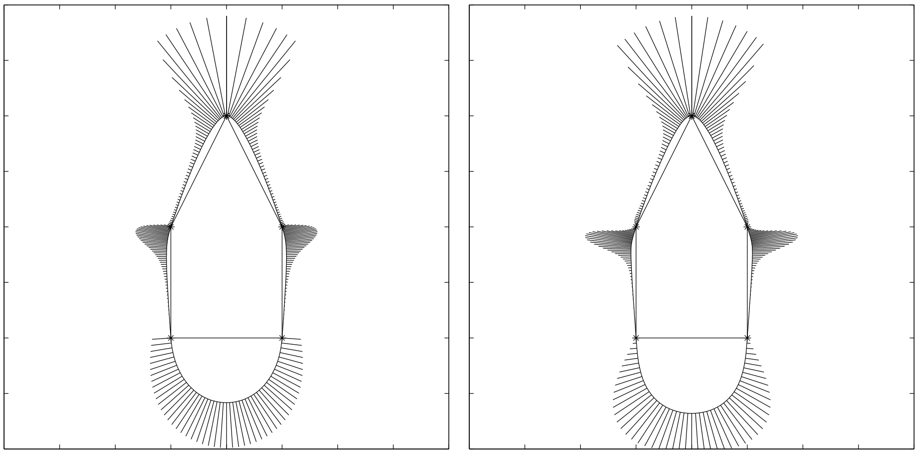
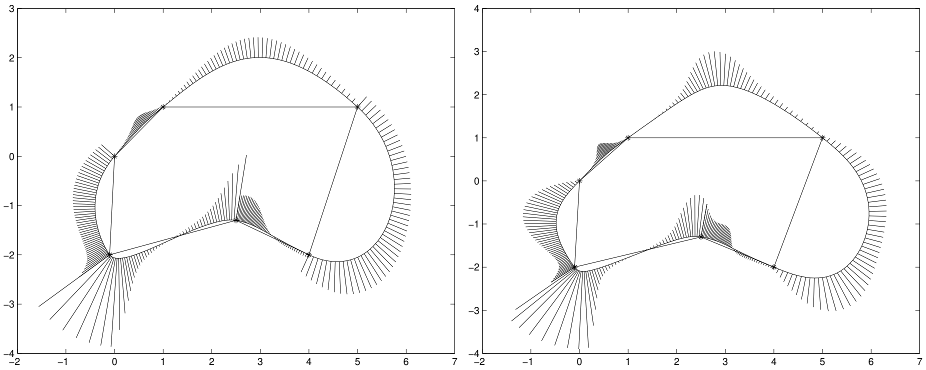


Figure 5: C^2 spline curve of degree 6.



(a) (b)

Figure 6: Example 2: (a) resulting cubic spline curve with C^1 continuity.(b) resulting spline of degree six with C^2 continuity.



(a) (b)

Figure 7: Example 3:(a) resulting cubic spline curve with C^1 continuity.(b) resulting spline of degree six with C^2 continuity.

References

[1] ASATURYAN S., COSTANTINI P. AND MANNI C., *G^2 Shape-preserving parametric planar curve interpolation*, in: "Designing and Creating Shape-Preserving Curves and Surfaces", (Eds. Nowacki H. and Kaklis P.), B.G. Teubner, Stuttgart 1998, 89–98.

[2] BELLMAN R. AND DREYFUS S., *Applied dynamic programming*, Princeton University Press, New York 1962.

[3] COSTANTINI P., *An algorithm for computing shape-preserving splines of arbitrary degree*, Journal of Computational and Applied Mathematics, **22** (1988).

[4] COSTANTINI P., *A general method for constrained curves with boundary conditions*, in: "Multivariate Approximation from CAGD to Wavelets", (Eds. Jetter K. and Utreras F.I.), World Scientific Publishing Co., Singapore 1993.

[5] COSTANTINI P., *Abstract schemes for functional shape-preserving interpolation*, in: "Advanced Course on FAIRSHAPE", (Eds. Nowacki H. and Kaklis P.), B.G. Teubner, Stuttgart 1996, 185–199.

[6] COSTANTINI P., *Boundary-valued shape-preserving interpolating splines*, ACM Transactions on Mathematical Software **23** 2 (1997), 229–251.

[7] COSTANTINI P. AND SAMPOLI M. L., *Abstract schemes and constrained curve interpolation*, in: "Designing and Creating Shape-Preserving Curves and Surfaces", (Eds. Nowacki H. and Kaklis P.), B.G. Teubner, Stuttgart 1998, 121–130.

[8] COSTANTINI P. AND SAMPOLI M. L., *A general scheme for shape preserving planar interpolating curves*, BIT **40** 4 (2003)

[9] COSTANTINI P. AND SAMPOLI M. L., *Constrained interpolation in $\mathbb{R}^3$ by abstract schemes*, in: "Curve and surface design: Saint-Malo 2002", (Eds. Lyche T., Mazure M.L. and Schumaker L.L.), Nashboro Press, Nashville 2003, 93–102.

[10] GOODMAN T. N. T. AND UNSWORTH K., *Shape preserving interpolation by parametrically defined curves*, SIAM J. Numer. Anal. **25** (1988), 1451–1465.

[11] HOSCHEK J. AND LASSER D., *Fundamentals of computer aided geometric design*, AK Peters Ltd., Wellesley 1993.

[12] KAKLIS P. D. AND SAPIDIS N. S., *Convexity preserving polynomial splines of non uniform degree*, Computer Aided Geometric Design **12** (1995), 1–26.

[13] MULANSKY B. AND SCHMIDT J. W., *Convex interval interpolation using three-term staircase algorithm*, Numerische Mathematik **82** (1999), 313–337.

[14] MULANSKY B. AND SCHMIDT J. W., *Composition based staircase algorithm and constrained interpolation with boundary conditions*, Numerische Mathematik **85** (2000), 387–408.

[15] PREPARATA F. P. AND SHAMOS M. I., *Computational geometry*, Springer-Verlag, Berlin, New York 1985.

[16] SCHMIDT J. W., *On shape-preserving spline interpolation: existence theorems and determination of optimal splines*, Approximation and Function Spaces **22** PWN-Polish Scientific Publishers, Warsaw 1989.

[17] SCHMIDT J. W., *Staircase algorithm and construction of convex spline interpolants up to the continuity C^3*, Computer Math. Appl. **31** (1996), 67–79.

AMS Subject Classification: 65D05, 65D17.

Maria Lucia SAMPOLI
Dipartimento di Scienze Matematiche ed Informatiche
Università di Siena
Via del Capitano 15
53100 Siena, ITALIA
e-mail: `sampoli@unisi.it`