

Multivariate Newton Interpolation

Michael Hecht · Karl B. Hoffmann ·
Bevan L. Cheeseman · Ivo F. Sbalzarini

Abstract For $m, n \in \mathbb{N}$, $m \geq 1$ and a given function $f : \mathbb{R}^m \rightarrow \mathbb{R}$, the *polynomial interpolation problem* (PIP) is to determine a *unisolvent node set* $P_{m,n} \subseteq \mathbb{R}^m$ of $N(m, n) := |P_{m,n}| = \binom{m+n}{n}$ points and the uniquely defined polynomial $Q_{m,n,f} \in \Pi_{m,n}$ in m variables of degree $\deg(Q_{m,n,f}) \leq n \in \mathbb{N}$ that fits f on $P_{m,n}$, i.e., $Q_{m,n,f}(p) = f(p)$, $\forall p \in P_{m,n}$. For $m = 1$ the solution to the PIP is well known. In higher dimensions, however, no closed framework was available. We here present a generalization of the classic Newton interpolation from one-dimensional to arbitrary-dimensional spaces. Further we formulate an algorithm, termed PIP-SOLVER, based on a *multivariate divided difference scheme* that computes the solution $Q_{m,n,f}$ in $\mathcal{O}(N(m, n)^2)$ time using $\mathcal{O}(N(m, n))$ memory. Further, we introduce unisolvent *Newton-Chebyshev nodes* and show that these nodes avoid *Runge's phenomenon* in the sense that arbitrary periodic *Sobolev functions* $f \in H^k(\Omega, \mathbb{R}) \subsetneq C^0(\Omega, \mathbb{R})$, $\Omega = [-1, 1]^m$ of regularity $k > m/2$ can be uniformly approximated, i.e., $\lim_{n \rightarrow \infty} \|f - Q_{m,n,f}\|_{C^0(\Omega)} = 0$. Numerical experiments demonstrate the computational performance and approximation accuracy of the PIP-SOLVER in practice. We expect the presented results to be relevant for many applications, including numerical solvers, quadrature, non-linear optimization, polynomial regression, adaptive sampling, Bayesian inference, and spectral analysis.

Keywords Newton interpolation, Vandermonde matrix, matrix inversion, unisolvent nodes, multivariate interpolation, numerical stability, polynomial approximation, Runge's phenomenon.

Mathematics Subject Classification (2010) 26C99, 41A10, 65D05, 65F05, 65F99, 65L20

MOSAIC Group, Chair of Scientific Computing for Systems Biology, Faculty of Computer Science, TU Dresden & Center for Systems Biology Dresden, Max Planck Institute of Molecular Cell Biology and Genetics, Dresden
Pfortenhauerstr. 108 01307 Dresden Germany
E-mail: hecht@mpi-cbg.de, karlhoff@mpi-cbg.de, cheesema@mpi-cbg.de, ivos@mpi-cbg.de
<http://mosaic.mpi-cbg.de>

1 Introduction

In scientific computing, the problem of interpolating a function $f : \mathbb{R}^m \rightarrow \mathbb{R}$, $m \in \mathbb{N}$, is ubiquitous. Because of their simple differentiation and integration, as well as their pleasant vector space structure, polynomials $Q \in \mathbb{R}[x_1, \dots, x_m]$ in m variables of degree $\deg(Q) \leq n$, $m, n \in \mathbb{N}$, are a standard choice as interpolants and are fundamental in ordinary differential equation (ODE) and partial differential equation (PDE) solvers. For an overview, we refer to [35] and [41]. Thus, the *polynomial interpolation problem* (PIP) is one of the most fundamental problems in numerical analysis and scientific computing, formulated as:

Problem 1 (PIP) Let $m, n \in \mathbb{N}$ and $f : \mathbb{R}^m \rightarrow \mathbb{R}$ be a computable function.

- i) Choose $N(m, n)$ nodes $P_{m,n} = \{p_1, \dots, p_{N(m,n)}\} \subseteq \mathbb{R}^m$ such that $P_{m,n}$ is unisolvent, i.e., for every f there is exactly one polynomial $Q_{m,n,f} \in \Pi_{m,n}$ fitting f on $P_{m,n}$ as $Q_{m,n,f}(p) = f(p)$ for all $p \in P_{m,n}$.
- ii) Determine $Q_{m,n,f}$ once a unisolvent node set $P_{m,n}$ has been chosen.

Here, $\Pi_{m,n}$ is the vector space of polynomials $Q \in \mathbb{R}[x_1, \dots, x_m]$ in m variables of degree $\deg(Q) \leq n$. Every $Q \in \Pi_{m,n}$ has $N(m, n) := \binom{m+n}{n}$ monomials/coefficients.

The function $f : \mathbb{R}^m \rightarrow \mathbb{R}$ is assumed to be *computable* in the sense that for any $x \in \mathbb{R}^m$ the value of $f(x)$ can be evaluated in $\mathcal{O}(1)$ time, where $\mathcal{O}(\cdot)$ is the Bachmann-Landau symbol. Note that if $f \in \Pi_{m,n}$, then $Q_{m,n,f} = f$. If $f : \mathbb{R} \rightarrow \mathbb{R}$ is an arbitrary continuous function in one variable, then the *Weierstrass approximation theorem* [12] guarantees that f can be approximated by *Bernstein polynomials*, i.e., there exist polynomials $Q_n \in \Pi_{1,n}$ such that

$$\|f - Q_n\|_{C^0(\Omega)} := \sup_{x \in \Omega} |f(x) - Q_n(x)| \xrightarrow{n \rightarrow \infty} 0$$

for every fixed bounded domain $\Omega \subseteq \mathbb{R}$. However, these polynomials Q_n are not necessarily interpolating, i.e., they need not satisfy $Q_n(p) = f(p)$ for all $p \in P_{1,n}$. This additional requirement restricts the space of polynomials available to approximate f , which is the cause of Runge's phenomenon [8, 15, 34]: If the unisolvent nodes $P_{1,n}$ are chosen independently of f , the sequence of interpolants $Q_{n,f} \in \Pi_{1,n}$, $n \in \mathbb{N}$, can diverge away from f , i.e., there exist at least one $f \in C^0(\Omega)$ for which $\|f - Q_{n,f}\|_{C^0(\Omega)} \not\rightarrow 0$ for $n \rightarrow \infty$. Therefore, the approximation ability of an interpolation method depends on the choice of $P_{m,n}$ and has to be characterized. More precisely:

Question 1 Let $m, n \in \mathbb{N}$ and $\Omega \subseteq \mathbb{R}^m$ be a bounded domain. Further let $S_{m,n} : C^0(\Omega, \mathbb{R}) \rightarrow \Pi_{m,n}$ be an interpolation solver, i.e., $S_{m,n}(f) = Q_{m,n,f}$ solves the PIP for any given function $f : \Omega \subseteq \mathbb{R}^m \rightarrow \mathbb{R}$, choosing the unisolvent nodes $P_{m,n} \subseteq \Omega$ independently of f .

- i) What is the set $\mathcal{A}(S_{m,n}) \subseteq C^0(\Omega, \mathbb{R})$ of continuous functions that can be approximated by $S_{m,n}$, i.e., for which $\|f - S_{m,n}(f)\|_{C^0(\Omega)} \xrightarrow{n \rightarrow \infty} 0, \forall f \in \mathcal{A}$?

ii) How large is the absolute approximation error

$$\varepsilon(f, S_{m,n}) := \|f - S_{m,n}(f)\|_{C^0(\Omega)}?$$

iii) How large is the relative approximation error $\mu_{m,n} \geq 1$, such that

$$\|f - S_{m,n}(f)\|_{C^0(\Omega)} \leq \mu_{m,n} \|f - Q_{m,n}^*\|_{C^0(\Omega)} \quad \forall f \in C^0(\Omega, \mathbb{R})$$

where $Q_{m,n}^* \in \Pi_{m,n}$ is an optimal approximation that minimizes the C^0 -distance to f ?

The one-dimensional PIP ($m = 1$) can be solved efficiently in $\mathcal{O}(N(1, n)^2) = \mathcal{O}(n^2)$ and numerically accurately by various algorithms based on *Newton or Lagrange Interpolation* [41, 3, 19, 27]. Though, even in one dimension, there is no efficient general method for finding an optimal node set $P_{1,m}$ that minimizes Runge's phenomenon; sub-optimal node sets can be generated efficiently. For $\Omega = [-1, 1]$, a classic choice of sub-optimal nodes are the roots of the *Chebyshev polynomials*

$$\text{Cheb}_n = \left\{ p_k \in \mathbb{R} : p_k = \cos\left(\frac{2k-1}{2(n+1)}\pi\right), k = 1, \dots, n+1 \right\}, \quad (1)$$

which are optimal up to a factor depending on the $(n+1)$ -th derivative of f . Therefore, the approximation ability of *Chebyshev nodes* is characterized by:

Theorem 0 Let $S_n : C^0(\Omega, \mathbb{R}) \rightarrow \Pi_n$, $\Omega = [-1, 1]$, be an interpolation solver that uses Cheb_n as unisolvent nodes, i.e., $S_n(f) = Q_{n,f}$, $Q_{n,f}(\text{Cheb}_n) = f(\text{Cheb}_n)$.

- i) The set of approximable functions satisfies $C^1(\Omega, \mathbb{R}) \subsetneq \mathcal{A}(\Omega, \mathbb{R}) \subsetneq C^0(\Omega, \mathbb{R})$.
- ii) If $f \in C^{n+1}(\Omega, \mathbb{R})$ and $x \in \Omega$ then the absolute approximation error at x can be bounded by

$$|f - S_n(f)(x)| \leq \frac{1}{(n+1)!} f^{(n+1)}(\xi_x) \prod_{i=1}^{n+1} (x - p_i) \leq \frac{f^{(n+1)}(\xi)}{2^n (n+1)!}, \xi_x \in \Omega,$$

- iii) The relative approximation error can be bounded by the Lebesgue function $\Lambda(\text{Cheb}_n)$ as:

$$\|f - S_{m,n}(f)\|_{C^0(\Omega)} \leq (1 + \Lambda(\text{Cheb}_n)) \|f - Q_n^*\|_{C^0(\Omega)},$$

where $\Lambda(\text{Cheb}_n) = \frac{2}{\pi} (\log(n) + \gamma + \log(8/\pi)) + \mathcal{O}(1/n^2)$ and $\gamma \sim 0.5772$ is Euler's constant [6].

This provides a pleasing solution to the PIP in one dimension. We refer to [8, 19, 6, 7, 40] for further details and proofs.

However, many data sets in scientific computing are functions of more than one variable and therefore require multivariate polynomial interpolation. A solution to the multivariate PIP, complete with a computationally efficient

and numerically stable algorithm for computing it, has so far not been available. This is at least partly due to the fact that an efficiently computable characterization of unisolvent nodes in arbitrary dimensions $m \in \mathbb{N}$ was not known. In one dimension, unisolvent node sets are characterized by the simple requirement that nodes have to be pairwise different, which can obviously be asserted in $\mathcal{O}(n^2)$ time. While some unisolvent node sets have been proposed in dimensions $m = 2, 3$ [4, 14, 18, 25], generalizations to arbitrary dimensions had a complexity that prohibited their practical implementation [16, 18]. Available PIP solvers in higher dimensions therefore use randomly generated node sets and then determine the interpolation polynomial by numerically inverting the resulting *multivariate Vandermonde matrix* $V_{m,n} \in \mathbb{R}^{N(m,n) \times N(m,n)}$ in order to compute the coefficients $C_{m,n}$ of $Q_{m,n,f}$ in normal form. Using random node sets is possible due to the famous theorem of Sard, which was later generalized by Smale [39]. This theorem states that the superset $\mathcal{P}_{m,n}$ of all unisolvent node sets for $m, n \in \mathbb{N}$ is a set of second category in the sense of Baire. Therefore, any randomly generated node set is unisolvent with probability 1.

Using random nodes, however, can never guarantee numerical stability of the solver, nor can it control Runge's phenomenon. In addition, numerical inversion of the multivariate Vandermonde matrix $V_{m,n}$ in practice incurs a computational cost larger than that of Newton interpolation. In principle, inverting the Vandermonde matrix should be as complex as solving the PIP, due to the special structure of $V_{m,n}$. However, this structure depends on the choice of unisolvent nodes $P_{m,n}$. Therefore, using random node sets prevents exploiting this structure, so that solving the system of linear equations

$$V_{m,n}(P_{m,n})C_{m,n} = F, \quad F = (f(p_1), \dots, f(p_{N(m,n)}))^T \in \mathbb{R}^{N(m,n)}$$

still requires the same computational time as *general matrix inversion*. A lower complexity bound for inverting general matrices of size $N \times N$, $N \in \mathbb{N}$, is given by $\mathcal{O}(N^2 \log(N))$ [10, 32, 45]. The fastest known algorithm for general matrix inversion is the *Coppersmith-Winograd algorithm* [9], which requires runtime in $\mathcal{O}(N^{2.3728639})$ in its most efficient version [16]. However, the *Coppersmith-Winograd algorithm* is rarely used in practice, because it is only advantageous for matrices so large that memory problems become prevalent on modern hardware [33]. The algorithm that is mostly used in practice is the *Strassen algorithm* [42], which runs in $\mathcal{O}(N^{2.807355})$. Alternatively, one can perform *Gaussian elimination* in $\mathcal{O}(N^3)$. All of these approaches require $\mathcal{O}(N^2)$ memory to store the matrix. Moreover, the numerical robustness and accuracy of these approaches is limited by the condition number of the Vandermonde matrix, which again depends on the choice of $P_{m,n}$ and can therefore not be controlled when using random node sets. Hence, previous approaches to polynomial interpolation become inaccurate or intractable with increasing $N(m, n)$.

1.1 Statement of Contribution

Though the relevance of multivariate interpolation is undisputed and feasible interpolation schemes in dimension 1 are known since the 18th century, there was so far no general interpolation scheme for multivariate functions that can guarantee to solve the PIP numerically robustly and accurately, controls Runge's phenomenon, and is as computationally efficient as Newton interpolation.

We here close this gap by introducing the notion of *multivariate Newton polynomials* and a characterization of *unisolvent nodes* $P_{m,n}$ such that all of the following is true:

- R1) The unisolvent nodes $P_{m,n}$ generate a well-conditioned Vandermonde matrix such that the interpolant can be computed numerically robustly and accurately.
- R2) The unisolvent nodes $P_{m,n}$ generate a *lower triangular* Vandermonde matrix with respect to the multivariate Newton polynomials, such that the interpolant can be computed in quadratic time using a *multivariate divided difference scheme*.
- R3) The unisolvent nodes $P_{m,n}$ control Runge's phenomenon in the sense that Theorem 0 generalizes to dimension m and thereby answers Question 1.

We practically implement the solution as an algorithm, called PIP-SOLVER, which, allows to interpolate arbitrary multivariate functions $f : \mathbb{R}^m \rightarrow \mathbb{R}$ numerically accurately. The PIP-SOLVER algorithm is based on a recursive decomposition approach, which yields a *recursive generator* of *unisolvent node sets* and the associated *multivariate divided difference scheme* to determine the interpolant $Q_{m,n,f}$. We show that the PIP-SOLVER requires runtime in $\mathcal{O}(N(m, n)^2)$ and memory in $\mathcal{O}(N(m, n))$, which matches the performance of Newton interpolation for $m = 1$. Further, the multivariate Newton-form of $Q_{m,n,f}$ allows evaluating and differentiating $Q_{m,n,f}$ in linear time.

Moreover, we show that all Sobolev functions $f \in H^k(\Omega, \mathbb{R}) \subseteq C^0(\Omega, \mathbb{R})$, $\Omega = [-1, 1]^m$ for $k > m/2$ that are periodic on Ω can be approximated uniformly

$$\|f - Q_{m,n,f}\|_{C^0(\Omega)} \xrightarrow{n \rightarrow \infty} 0 \quad \text{for } \Omega = [-1, 1]^m$$

by using unisolvent nodes of *generalized Newton-Chebyshev type*. Analogous to the one-dimensional case, we provide bounds for the relative and absolute approximation errors. Note that this is probably the best result one can obtain, since uniform approximation of non-continuous functions by polynomials and extrapolating functions $f : \Omega' \supseteq \Omega \rightarrow \mathbb{R}$ with $P_{m,n} \subseteq \Omega$ are ill-posed problems. Because $H^k(\Omega, \mathbb{R})$ contains non-continuous functions for $k \leq m/2$, the Sobolev functions $f \in H^k(\Omega, \mathbb{R})$ for $k > m/2$ are the largest *Hilbert space* of continuous functions that can be approximated by polynomials in the C^0 -sense. Further, the Sobolev space, $H^k(\Omega, \mathbb{R})$, $k > m/2$, densely contains all *smooth* and thus all *analytical functions*. The periodic boundary condition can always be achieved by rescaling. For example, consider \tilde{f} periodic on $\Omega_2 =$

$[-2, 2]^m$ with $\tilde{f}|_{\Omega} = f$ and rescale Ω_2 to Ω . For these and other reasons, $H^k(\Omega, \mathbb{R})$, $k > m/2$, is the pivotal analytical choice in scientific computing [24].

1.2 Paper Outline

After stating the main results of this article in section 2, we recapitulate classic one-dimensional Newton interpolation in section 3 and previous multivariate interpolation schemes in section 4. In section 5 we provide the mathematical proofs for our results. In section 6, we present the *multivariate divided difference scheme* for generalized *multidimensional Newton nodes* as an efficient solver algorithm. In section 7, we show that multivariate Newton-Chebyshev nodes allow proving the uniform approximation result for periodic Sobolev functions $f \in H^k(\Omega, \mathbb{R})$, $k > m/2$, and the bounds on the approximation errors. The results are then demonstrated in numerical experiments in section 8. Finally, we sketch a few of the possible applications in section 9 and conclude in section 10.

2 Main Results

We summarize the main results of this article in the following three Theorems. They are based on the realization that the PIP can be decomposed into sub-problems of lesser dimension or lesser degree. Recursion then decomposes the problem along a binary tree whose leafs are associated with constants or zero-dimensional sub-problems, which only require evaluating f to be solved. The decomposition is based on the notion of multivariate Newton polynomials, which we introduce in section 6, generalizing Newton nodes to arbitrary dimensions and defining a multivariate Newton basis.

Theorem 1 (Main Result I) *Let $m, n \in \mathbb{N}$ and $f : \mathbb{R}^m \rightarrow \mathbb{R}$ be a given function. Then, there exists an algorithm with runtime complexity $\mathcal{O}(N(m, n)^2)$ requiring $\mathcal{O}(N(m, n))$ memory that computes:*

- i) *A unisolvent node set $P_{m, n} \subseteq \mathbb{R}^m$ and the coefficients of the corresponding interpolation polynomial $Q_{m, n, f} \in \Pi_{m, n}$ in normal form.*
- ii) *A unisolvent set $P_{m, n} \subseteq \mathbb{R}^m$ of multidimensional Newton nodes and the coefficients of the corresponding interpolation polynomial $Q_{m, n, f} \in \Pi_{m, n}$ in multivariate Newton form.*

We generalize classic results for Newton polynomials to show that the present multivariate Newton form is a good choice for multivariate polynomial interpolation. In particular, it enables efficient and accurate numerical computations according to the second main result:

Theorem 2 (Main Result II) *Let $m, n \in \mathbb{N}$ and a set of multidimensional Newton nodes $P_{m, n}$ be given. Let further $Q \in \Pi_{m, n}$ be a polynomial given in multivariate Newton form. Then, there exist algorithms that compute:*

- i) The value of $Q(x_0)$ for any $x_0 \in \mathbb{R}^m$ in $\mathcal{O}(N(m, n))$;
- ii) The partial derivative $\partial_{x_i} Q|_{x_0}$ for any $i \in \{1, \dots, m\}$ and $x_0 \in \mathbb{R}^m$ in $\mathcal{O}(nN(m, n))$;
- iii) The integral $\int_{\Omega} Q(x) dx$ for any hypercube $\Omega \subseteq \mathbb{R}^m$ with runtime complexity in $\mathcal{O}(nN(m, n))$.

Studying the approximation properties of multivariate Newton interpolation, we show that periodic Sobolev functions $f \in H^k(\Omega, \mathbb{R})$ for $k > m/2$ can be approximated uniformly, and we provide upper bounds for the corresponding absolute and relative approximation errors. To do so let $A_{m,n} \subseteq \mathbb{N}^m$ be the set of all multi-indices $\alpha = (\alpha_1, \dots, \alpha_m) \in \mathbb{N}^m$ with $|\alpha| = \sum_{i=1}^m \alpha_i \leq n$.

Theorem 3 (Main Result III) *Let $m, n \in \mathbb{N}$, $\Omega = [-1, 1]^m \subseteq \mathbb{R}^m$. Let $S_{m,n} : H^k(\Omega, \mathbb{R}) \rightarrow \Pi_{m,n}$, $k > m/2$, denote an interpolation operator with respect to interpolation nodes $P_{m,n}$.*

- i) If $P_{m,n}$ are of Newton type, then the Lebesgue function

$$\Lambda(P_{m,n}, H^k(\Omega)) = \sup_{f \in H^k(\Omega, \mathbb{R}), \|f\|_{H^k(\Omega)}=1} \|S_{m,n}(f)\|_{C^0(\Omega)},$$

given by the operator norm of $S_{m,n}$ is bounded by

$$\Lambda(P_{m,n}, H^k(\Omega)) \leq \prod_{i=1}^m \Lambda(P_i, H^{k-(m-1)/2}(\Omega)).$$

- ii) If $P_{m,n}$ are of Newton-Chebyshev type, then $\Lambda(P_{m,n}, H^k(\Omega)) \in \mathcal{O}(\log(n)^m)$ and every $f \in H^k(\Omega, \mathbb{R})$, $k > m/2$, being periodic on $\Omega = [-1, 1]^m$ can be approximated uniformly:

$$\|f - S_{m,n}(f)\|_{C^0(\Omega)} \xrightarrow{n \rightarrow \infty} 0.$$

- iii) If $f \in C^{n+1}(\Omega, \mathbb{R})$ and $P_{m,n}$ are of Newton-type then for every $\alpha \in A_{m,n} \setminus A_{m,n-1}$, $i \in \{1, \dots, m\}$ and every $x \in \Omega$ there is $\xi_x \in \Omega$ such that

$$|f(x) - S_{m,n}(f)(x)| \leq \frac{1}{\alpha_i!} \partial_{x_i}^{\alpha_i+1} f(\xi) |N_{\alpha}(x)|,$$

where $N_{\alpha}(x) = \prod_{i=1}^m \prod_{j=1}^{\alpha_i} (x_i - p_{i,j})$, $x = (x_1, \dots, x_m)$, $p_{i,j} \in \mathbb{P}_i$. If $P_{m,n}$ are of Newton-Chebyshev-type then we can further estimate

$$|f(x) - S_{m,n}(f)(x)| \leq \frac{1}{2^{\alpha_i} \alpha_i!} \partial_{x_i}^{\alpha_i+1} f(\xi_x).$$

- iv) For any unisolvent node set $P_{m,n}$ and every $f \in H^k(\Omega, \mathbb{R})$ the relative error is bounded by

$$\|f - S_{m,n}(f)\|_{C^0(\Omega)} \leq (1 + \Lambda(P_{m,n}, H^k(\Omega))) \|f - Q_{m,n}^*\|_{C^0(\Omega)},$$

where $Q_{m,n}^*$ is an optimal approximation that minimizes the C^0 -distance to f .

Λ is estimated in (i), but statement (iv) holds for all unisolvent node sets. For Chebyshev nodes, (iv) provides a tight estimate.

3 Newton Interpolation in Dimension 1

Since our m -dimensional generalization is a natural extension of *Newton interpolation* in 1D, we first review some classic results in the special case of dimension $m = 1$, with more detailed discussions available elsewhere [2, 13, 27, 31, 35, 40, 41, 49]. In one dimension, the Vandermonde matrix $V_n(P_n)$ takes the classical form

$$V_n(P_n) = \begin{pmatrix} 1 & p_1 & \cdots & p_1^n \\ \vdots & \vdots & \ddots & \vdots \\ 1 & p_{n+1} & \cdots & p_{n+1}^n \end{pmatrix}.$$

It is well known that for this matrix to be regular, the nodes p_1, \dots, p_{n+1} have to be pairwise distinct. As we see later, this is also a sufficient condition for the nodes P_n to be unisolvent. In light of this fact, we observe that V_n induces a vector space isomorphism $\varphi : \mathbb{R}^{n+1} \rightarrow \Pi_n$, where $\varphi(v)$ is the polynomial with normal-form coefficients $C_n \in \mathbb{R}^{n+1}$ such that $V_n C_n = v$, $v \in \mathbb{R}^{n+1}$. The polynomials

$$N_i(x) = \prod_{j=1}^i (x - p_j), \quad i = 0, \dots, n \quad (2)$$

are the *Newton Basis* (NB) of Π_n . When represented with respect to the NB, the Vandermonde matrix becomes a lower triangular matrix of the form

$$V_{\text{NB},n}(P_n) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & (p_2 - p_1) & \cdots & 0 \\ 1 & (p_3 - p_1) & (p_3 - p_1)(p_3 - p_2) & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (p_{n+1} - p_1) & \cdots & \prod_{j=1}^n (p_{n+1} - p_j) \end{pmatrix}.$$

Thus, the solution of the PIP in dimension $m = 1$ with respect to the NB can directly be obtained as:

$$Q_{n,f}(x) = \sum_{i=0}^n c_i N_i(x) = \sum_{i=0}^n c_i \prod_{j=1}^i (x - p_j). \quad (3)$$

More efficiently, the *Aitken-Neville* or *divided difference* scheme determines the coefficients c_i by setting

$$[p_1]f := f(p_1), \quad [p_i, \dots, p_j]f := \frac{[p_i, \dots, p_{j-1}]f - [p_{i+1}, \dots, p_j]f}{x_j - x_i}, \quad j \geq i$$

and proving that $c_{i-1} = [p_1, \dots, p_i]f$. Indeed one can verify by induction that $Q_{n,f}$ satisfies the fitting condition $f(P_n) = Q_{n,f}(P_n)$, which uniquely determines $Q_{n,f}$ up to its representation. The induced recursion can be illustrated

as follows:

$$\begin{array}{ccccccc}
 [p_1]f & & & & & & \\
 & \searrow & & & & & \\
 [p_2]f & \rightarrow & [p_1, p_2]f & & & & \\
 & \searrow & & \searrow & & & \\
 [p_3]f & \rightarrow & [p_2, p_3]f & \rightarrow & [p_1, p_2, p_3]f & & \\
 \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \\
 & \searrow & & \searrow & & \searrow & \\
 [p_{n+1}]f & \rightarrow & [p_n, p_{n+1}]f & \rightarrow & [p_{n-1}, p_n, p_{n+1}]f & \cdots & \rightarrow [p_1 \cdots p_{n+1}]f.
 \end{array} \tag{4}$$

We summarize some facts about the classic 1D scheme [19, 41], as they are important for the generalization to arbitrary dimensions:

Proposition 1 *Let $n \in \mathbb{N}$ and $f \in C^0(\Omega, \mathbb{R})$, $\Omega = [-1, 1]$*

- i) The divided difference scheme allows to numerically stable determine the $n+1$ coefficients $c_0, \dots, c_n \in \mathbb{R}$ of the interpolant $Q_{n,f}$ with respect to the Newton-basis in $\mathcal{O}(n^2)$.*
- ii) Given the interpolant $Q_{n,f}$ in its Newton-form, the Horner-scheme allows to compute the value of $Q_{n,f}(x)$ in $\mathcal{O}(n)$ for any $x \in \Omega$.*
- iii) Given the interpolant $Q_{n,f}$ in its Newton-form, the value of the derivative $\frac{d}{dx}Q_{n,f}(x)$ can be computed in $\mathcal{O}(n)$ for any $x \in \Omega$.*

Due to its relationship with *Taylor expansion*, *Newton interpolation* has several advantages over other interpolation schemes. For instance, it easily extends to higher degrees without recomputing the coefficients of lower-order terms, i.e., by incrementally computing higher-order terms. Furthermore, evaluation of the Newton interpolant and its derivatives is straightforward due to their simple form. However, the recursive nature of the divided difference scheme also has certain disadvantages compared to *Lagrange interpolation*, which computes the result at once. We refer to [3] and [50] for further discussions of the properties of these schemes.

In any case, due to the uniqueness of the interpolation polynomial, the question how well the interpolant approximates a given function is independent of the specific interpolation scheme, up to numerical rounding errors. Instead, the approximation quality only depends on the choice of interpolation nodes. The approximation quality in the 1D case is given in Theorem (0).

4 Multivariate Polynomials

We follow the notation of P.J. Olver [28] to extend our considerations to the general multivariate case. For $m, n \in \mathbb{N}$, we denote by $\Pi_{m,n} \subseteq \mathbb{R}[x_1, \dots, x_m]$ the vector space of all real polynomials in m variables of bounded degree n . While the normal form of a polynomial $Q \in \Pi_{m,n}$ possesses $N(m, n) := \binom{m+n}{n}$ monomials, the number of monomials of degree k is given by $M(m, k) :=$

$\binom{m+k}{m} - \binom{m+k-1}{m}$. We enumerate the coefficients $c_0, \dots, c_{N(m,n)}$ of $Q \in \Pi_{m,n}$ in its normal form as follows:

$$\begin{aligned} Q(x) = & c_0 + c_1 x_1 + \dots + c_m x_m + c_{m+1} x_1^2 + c_{m+2} x_1 x_2 \\ & + \dots + c_{2m} x_1 x_m + c_{2m+1} x_2^2 + \dots + c_{M(m,n-1)+1} x_1^n \\ & + c_{M(m,n-1)+2} x_1^{n-1} x_2 + \dots + c_{N(m,n)-2} x_{m-1} x_m^{n-1} \\ & + c_{N(m,n)-1} x_m^n. \end{aligned} \quad (5)$$

We assume that $0 \in \mathbb{N}$ and consider $A_{m,n} := \{\alpha \in \mathbb{N}^m : |\alpha| \leq n\}$ the set of all multi-indices of order $|\alpha| := \sum_{k=1}^m \alpha_k \leq n$. The multi-index is used to address the monomials of a multivariate polynomial. For a vector $x = (x_1, \dots, x_m)$ and $\alpha \in A_{m,n}$ we define

$$x^\alpha := x_1^{\alpha_1} \dots x_m^{\alpha_m}$$

and we order the $M(m,k)$ multi-indices of order k with respect to lexicographical order, i.e., $\alpha_1 = (k, 0, \dots, 0)$, $\alpha_2 = (k-1, 1, 0, \dots, 0)$, \dots , $\alpha_{M(m,k)} = (0, \dots, 0, k)$. The k -th *symmetric power* $x^{\odot k} = (x_1^{\odot k}, \dots, x_{M(m,k)}^{\odot k}) \in \mathbb{R}^{M(m,k)}$ is defined by the entries

$$x_i^{\odot k} := x^{\alpha_i}, \quad i = 1, \dots, M(m,k). \quad (6)$$

Thus, $x_i^{\odot 0} = 1$, $x_i^{\odot 1} = x_i$.

Definition 1 (multivariate Vandermonde matrix) For given $n, m \in \mathbb{N}$ and a set of nodes $P = \{p_1, \dots, p_{N(m,n)}\} \subseteq \mathbb{R}^m$ with $p_i = (p_{1,i}, \dots, p_{m,i})$, we define the *multivariate Vandermonde matrix* $V_{m,n}(P)$ by

$$V_{m,n}(P) = \begin{pmatrix} 1 & p_1 & p_1^{\odot 2} & \dots & p_1^{\odot n} \\ 1 & p_2 & p_2^{\odot 2} & \dots & p_2^{\odot n} \\ 1 & p_3 & p_3^{\odot 2} & \dots & p_3^{\odot n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & p_{N(m,n)} & p_{N(m,n)}^{\odot 2} & \dots & p_{N(m,n)}^{\odot n} \end{pmatrix}.$$

We call a set of nodes $P = \{p_1, \dots, p_{N(m,n)}\} \subseteq \mathbb{R}^m$ *unisolvent* if and only if the Vandermonde matrix $V_{m,n}(P)$ is regular. Thus, the set of all unisolvent node sets is given by $\mathcal{P}_{m,n} = \{P \subseteq \mathbb{R}^m : \det(V_{m,n}(P)) \neq 0\}$, which is an open set in \mathbb{R}^m , since $P \mapsto \det(V_{m,n}(P))$ is a continuous function.

Given a real-valued function $f : \mathbb{R}^m \rightarrow \mathbb{R}$, and assuming that unisolvent nodes $P = \{p_1, \dots, p_{N(m,n)}\} \subseteq \mathbb{R}^m$ exist, the linear system of equations $V_{m,n}(P)x = F$, with

$$x = (x_1, \dots, x_{N(m,n)})^\top \quad \text{and} \quad F = (f(p_1), \dots, f(p_{N(m,n)}))^\top,$$

has the unique solution

$$x = V_{m,n}(P)^{-1} F.$$

Thus, by using $c_i := x_i$ as the coefficients of $Q \in \Pi_{m,n}$, enumerated as in Eq. (5), we have uniquely determined the solution of Problem 1. The essential difficulty herein lies in finding a good unisolvent node set posing a well-conditioned problem and in solving $V_{m,n}(P)x = F$ accurately and efficiently. Therefore, we state a well-known result, first mentioned in [25] and then again in [28], saying:

Proposition 2 *Let $m, n \in \mathbb{N}$ and $P_{m,n} \subseteq \mathbb{R}^m$, $\#P_{m,n} = N(m, n)$. The Vandermonde matrix $V_{m,n}(P_{m,n})$ is regular if and only if the nodes $P_{m,n}$ do not belong to a common algebraic hypersurface of degree $\leq n$, i.e., if there exists no polynomial $Q \in \Pi_{m,n}$, such that $Q(p) = 0$ for all $p \in P_{m,n}$.*

Proof Indeed $P_{m,n}$ is unisolvent if and only if the homogeneous Vandermonde problem

$$V_{m,n}(P)x = 0$$

has no non-trivial solution, which is equivalent to the fact that there is no polynomial $Q \in \Pi_{m,n} \setminus \{0\}$ generating a hypersurface $W = Q^{-1}(0)$ of degree $\deg(W) \leq n$ with $P_{m,n} \subseteq W$. \square

The geometric interpretation of Proposition 2 is crucial but not constructive, i.e., it classifies node sets but yields no algorithm or scheme to construct unisolvent node sets. In the next section, we therefore further develop our understanding of unisolvent node sets in order to provide a construction algorithm, which is our first main result.

5 Multivariate Interpolation

We provide theorems that allow solving the PIP in a way that fulfills requirements (R1) and (R2) stated in the introduction. We start by considering the multivariate interpolation problem on hyperplanes and then show that the general multivariate PIP can be decomposed into such hyperplane problems.

5.1 Interpolation on Hyperplanes

We consider the PIP on certain subplanes and therefore define:

Definition 2 (affine transformation) Let $\tau : \mathbb{R}^m \rightarrow \mathbb{R}^m$ be given by $\tau(x) = Ax + b$, where $A \in \mathbb{R}^{m \times m}$ is a full-rank matrix, i.e. $\text{rank}(A) = m$, and $b \in \mathbb{R}^m$. Then we call τ an *affine transformation* on \mathbb{R}^m .

Definition 3 For every ordered tuple of integers $i_1, \dots, i_k \in \mathbb{N}$, $i_q < i_p$ if $1 \leq q < p \leq k$, we consider

$$H_{i_1, \dots, i_k} = \{(x_1, \dots, x_n) \in \mathbb{R}^m : x_j = 0 \text{ if } j \notin \{i_1, \dots, i_k\}\}$$

the k -dimensional hyperplanes spanned by the i_1, \dots, i_k -th coordinates. We denote by $\pi_{i_1, \dots, i_k} : \mathbb{R}^m \rightarrow H_{i_1, \dots, i_k}$ and $i_{i_1, \dots, i_k} : \mathbb{R}^k \hookrightarrow \mathbb{R}^m$, with $i_{i_1, \dots, i_k}(\mathbb{R}^k) =$

H_{i_1, \dots, i_k} , the natural projections and embeddings. We denote by $\pi_{i_1, \dots, i_k}^* : \mathbb{R}[x_1, \dots, x_m] \longrightarrow \mathbb{R}[x_{i_1}, \dots, x_{i_k}]$ and $i_{i_1, \dots, i_k}^* : \mathbb{R}[y_1, \dots, y_k] \hookrightarrow \mathbb{R}[x_1, \dots, x_m]$ the induced projections and embeddings on the polynomial ring.

Definition 4 (solution of the PIP) Let $m, n \in \mathbb{N}$, $\xi_1, \dots, \xi_m \in \mathbb{R}^m$ an orthonormal frame (i.e., $\langle \xi_i, \xi_j \rangle = \delta_{ij}$, $\forall 1 \leq i, j \leq m$, where δ_{ij} denotes the Kronecker symbol), and $b \in \mathbb{R}^m$. For $\mathcal{I} \subseteq \{1, \dots, m\}$ we consider the hyperplane

$$H_{\mathcal{I}, \xi, b} := \{x \in \mathbb{R}^m : \langle x - b, \xi_i \rangle = 0, \forall i \in \mathcal{I}\}.$$

Given a function $f : \mathbb{R}^m \longrightarrow \mathbb{R}$ we say that a set of nodes $P_{k, n} \subseteq H_{\mathcal{I}, \xi, b}$ and a polynomial $Q \in \Pi_{m, n}$ solve the PIP with respect to $k = \dim H_{\mathcal{I}}$, $n \in \mathbb{N}$ and f on $H_{\mathcal{I}}$ if and only if $Q(p) = f(p)$ for all $p \in P_{k, n}$ and whenever there is a $Q' \in \Pi_{m, n}$ and $Q'(p) = f(p)$ for all $p \in P_{k, n}$, then $Q'(x) = Q(x)$ for all $x \in H_{\mathcal{I}, \xi, b}$.

Definition 5 (induced transformation) Let $H \subseteq \mathbb{R}^m$ be a hyperplane of dimension $k \in \mathbb{N}$ and $\tau : \mathbb{R}^m \longrightarrow \mathbb{R}^m$ an affine transformation such that $\tau(H) = H_{1, \dots, k}$. Then we denote by

$$\tau^* : \mathbb{R}[x_1, \dots, x_m] \longrightarrow \mathbb{R}[x_1, \dots, x_m]$$

the induced transformation on the polynomial ring defined over the monomials as:

$$\tau^*(x_i) = \eta_1 x_1 + \dots + \eta_m x_m \quad \text{with} \quad \eta = (\eta_1, \dots, \eta_m) = \tau(e_i),$$

where e_1, \dots, e_m denotes the standard Cartesian basis of \mathbb{R}^m .

Lemma 1 Let $m, n \in \mathbb{N}$ and $P_{m, n}$ be a unisolvent node set with respect to (m, n) . Further let $\tau : \mathbb{R}^m \longrightarrow \mathbb{R}^m$, $\tau(x) = Ax + b$ be an affine transformation. Then, $\tau(P_{m, n})$ is also a unisolvent node set with respect to (m, n) .

Proof Assume there exists a polynomial $Q_0 \in \Pi_{m, n} \subseteq \mathbb{R}[x_1, \dots, x_m]$ such that $Q_0(\tau(P_{m, n})) = 0$. Then setting $Q_1 := \tau^*(Q_0)$ yields a non-zero polynomial of with $\deg(Q_1) \leq n$ and

$$Q_1(P_{m, n}) = \tau^*(Q_0)(P_{m, n}) = Q_0(\tau(P_{m, n})) = 0,$$

which is a contradiction to $P_{m, n}$ being unisolvent. Therefore, $\tau(P_{m, n})$ must be unisolvent. \square

Next we use the ingredients above to decompose the PIP.

5.2 Decomposition of the Multivariate PIP onto Hyperplanes

Section 5.1 provides the ingredients to prove our first key result. To avoid confusion, we note that a 0-dimensional hyperplane $H \subseteq \mathbb{R}^m$, $m \geq 1$, is given by a single point $H = \{\text{point}\}$, such that the PIP on H is solved by evaluating f at that point. Vice versa, the zeroth-order PIP (i.e., a constant) with respect to $f : \mathbb{R}^m \longrightarrow \mathbb{R}$ and $n = 0$ is solved by evaluating f at some point $q \in \mathbb{R}^m$.

Theorem 4 *Let $m, n \in \mathbb{N}$, $m \geq 1$, and $P \subseteq \mathbb{R}^m$, such that:*

- i) There exists a hyperplane $H \subseteq \mathbb{R}^m$ of co-dimension 1 such that $P_1 := P \cap H$ satisfies $\#P_1 = N(m-1, n)$ and is unisolvent with respect to H , i.e., by identifying $H \cong \mathbb{R}^{m-1}$ the Vandermonde matrix $V_{m-1, n}(P_1)$ is regular.*
- ii) The set $P_2 = P \setminus H$ satisfies $\#P_2 = N(m, n-1)$ and is unisolvent with respect to $(m, n-1)$, i.e., the Vandermonde matrix $V_{m, n-1}(P_2)$ is regular.*

Then P is a unisolvent node set.

Proof Due to Lemma 1, unisolvent node sets remain unisolvent under affine transformation. Thus, by choosing the appropriate transformation τ , we can assume w.l.o.g. that $H = H_{1, \dots, m-1}$. For any polynomial $Q \in \mathbb{R}[x_1, \dots, x_m]$ there holds $Q(\pi_{1, \dots, m-1}(p)) = \pi_{1, \dots, m-1}^*(Q(p))$ for all $p \in H$. Thus, by (i) we observe that whenever there is a $Q \in \mathbb{R}[x_1, \dots, x_m]$ with $Q(P_1) = 0$, then $\deg(Q) \geq \deg(\pi_{1, \dots, m-1}^*(Q)) > n$. Or, if $\deg(\pi_{1, \dots, m-1}^*(Q)) \leq \deg(Q) \leq n$, we consider the polynomial $\bar{Q}_1 := Q - i_{1, \dots, m-1}^*(\pi_{1, \dots, m-1}^*(Q))$, which consists of all monomials sharing the variable x_m and $\bar{Q}_2 := Q - \bar{Q}_1$ consisting of all monomials not sharing x_m . We claim that $\bar{Q}_2 = 0$. Certainly, $\bar{Q}_1(x) = 0$ for all $x \in H$. Since P_1 is unisolvent there are $p \in P_1$ with $\bar{Q}_2(p) = \pi_{1, \dots, m-1}^*(\bar{Q}_2(p)) \neq 0$ implying $Q(p) \neq 0$, which contradicts our assumption on Q and therefore yields $\bar{Q}_2 = 0$ as claimed. In light of this fact, Q can be decomposed into polynomials $Q = Q_1 \cdot Q_2$ where $Q_2(x_1, \dots, x_m) = x_m$. Since $\deg(Q_1) \leq n-1$ and P_2 is unisolvent we have that $P_2 \not\subseteq Q_1^{-1}(0)$. At the same time, $P_2 \cap H = \emptyset$ implies that $Q_2(p) \neq 0$ for all $p \in P_2$. Hence, there is $p \in P_2$ with $Q(p) \neq 0$, proving the theorem due to Proposition 2. \square

The question arises whether the decomposition of unisolvent nodes given in Theorem 4 allows us to decompose the PIP into smaller, and therefore simpler, sub-problems. Indeed we obtain:

Theorem 5 *Let $m, n \in \mathbb{N}$, $m \geq 1$, $f : \mathbb{R}^m \rightarrow \mathbb{R}$ be a computable function, $H \subseteq \mathbb{R}^m$ a hyperplane of co-dimension 1, $Q_H \in \Pi_{m,1}$ a polynomial satisfying $Q_H^{-1}(0) = H$, and $P_{m,n} = P_1 \cup P_2 \subseteq \mathbb{R}^m$ such that (i) and (ii) of Theorem 4 hold with respect to H . Require $Q_1, Q_2 \in \Pi_{m,n}$ to be such that:*

- i) Q_1 solves the PIP with respect to f and P_1 on H .*
- ii) $\deg(Q_2) \leq n-1$ and Q_2 solves the PIP with respect to $(m, n-1)$, $f_1 := (f - Q_1)/Q_H$, and $P_2 = P \setminus H$ on \mathbb{R}^m .*

Then, $Q_{m,n,f} = Q_1 + Q_H Q_2$ is the uniquely determined polynomial of $\deg(Q) \leq n$ that solves the PIP with respect to f and $P_{m,n}$ on \mathbb{R}^m .

Proof By our assumptions on Q_H and Q_1 we have that $Q(x) = Q_1(x) \forall x \in H$ and therefore $Q(p) = f(p) \forall p \in P_1 = P_{m,n} \cap H$. At the same time, $Q_H(x) \neq 0$ for all $x \notin H$. Therefore, f_1 is well defined on P_2 and $Q(p) = Q_1(p) + f_1(p)Q_H(p) = f(p) \forall p \in P_2 = P_{m,n} \setminus H$. Hence, $\deg(Q) \leq n$, and Q solves the PIP with respect to f and a unisolvent set of nodes $P_{m,n} \subseteq \mathbb{R}^m$. Thus, $Q_{m,n,f}$ is the unique solution of the PIP with respect to $P_{m,n}$ and f . \square

Remark 1 Note that Q_H can be constructed by choosing a (usually unit) normal vector $\nu \in \mathbb{R}^m$ onto H and a vector $b \in H$, setting

$$Q_H(x) = \nu \cdot (x - b).$$

Indeed, this guarantees that $Q_H(x) = 0$ for all $x \in H$ and $Q_H(x) \neq 0$ for all $x \in \mathbb{R}^m \setminus H$.

As a consequence of Theorems 4 and 5, we can close this section by stating the first part of Main Result I and delivering its proof.

Theorem 6 *Let $m, n \in \mathbb{N}$, and $f : \mathbb{R}^m \rightarrow \mathbb{R}$ be a computable function. Then, there exists an algorithm with runtime complexity in $\mathcal{O}(N(m, n)^2)$, requiring storage in $\mathcal{O}(N(m, n))$, that computes:*

- i) A unisolvent node set $P_{m,n} \subseteq \mathbb{R}^m$.*
- ii) The normal form coefficients $c_0, \dots, c_{N(m,n)-1}$ of the interpolation polynomial $Q_{m,n,f} \in \Pi_{m,n}$ with $Q_{m,n,f}(p) = f(p)$, $\forall p \in P_{m,n}$.*

Proof We start by proving (i) and (ii) with respect to the runtime complexity. To do so, we claim that there is a constant $C \in \mathbb{R}^+$ and an algorithm computing (i) and (ii) in less than $CN(m, n)^2$ computation steps. To prove this claim, we argue by induction on $N(m, n)$. If $N(m, n) = 1$, then $m = 0$ or $n = 0$. Thus, interpolating f is given by evaluating f at one single node, which can be done in $\mathcal{O}(1)$ by our assumption on f .

Now let $N(m, n) > 1$. Then $m > 0$ and we choose $\nu, b \in \mathbb{R}^m$ with $\|\nu\| = 1$ and consider the hyperplane $H = Q_H^{-1}(0)$, $Q_H(x) = \nu(x - b)$. By identifying $H \cong \mathbb{R}^{m-1}$, induction yields that we can determine a set of unisolvent nodes $P_1 \subseteq H$ in less than $CN(m-1, n)^2$ computation steps. Induction also yields that a set $P_2 \subseteq \mathbb{R}^m$ of unisolvent nodes can be determined with respect to $n-1$ in less than $CN(m-1, n)^2$ computation steps. By translating P_2 with $\lambda\nu$, i.e., setting $P'_2 = P_2 + \lambda\nu$, $\lambda \in \mathbb{R}$, we can guarantee that $P_2 \cap H = \emptyset$. Hence, the union $P_{m,n} = P_1 \cup P'_2$ of the corresponding unisolvent sets of nodes is also unisolvent due to Theorem 4, proving (i).

To show (ii) we have to compute the coefficients of $Q_{m,n,f}$ in normal form. By induction, a polynomial $Q_1 \in \Pi_{m,n}$ that solves the PIP with respect to f and P_1 on H can be determined in normal form in less than $CN(m-1, n)^2$ steps. We consider $f'_1 : \mathbb{R}^m \rightarrow \mathbb{R}$ with $f'_1(x) := (f(x - \lambda\nu) - Q_1(x - \lambda\nu))/Q_H(x - \lambda\nu)$. By induction, we can compute $Q_2 \in \Pi_{m,n-1}$ in less than $CN(m, n-1)^2$ computation steps, such that Q_2 solves the PIP with respect to f'_1 and P_2 , $(m, n-1)$. Thus, Q_2 solves the PIP with respect to $f_1(x) = (f(x) - Q_1(x))/Q_H(x) = f'_1(x + \lambda\nu)$, P'_2 , and $(m, n-1)$. Due to Theorem 5, we have that $Q_1 + Q_H Q_2$ solves the PIP with respect to f and $P_{m,n}$. It remains to bound the steps required for computing the normal form of $Q_1 + Q_H Q_2$. The bottleneck herein lies in the computation of $Q_H Q_2$, which requires $C_2(m+1)N(m, n-1)$, $C_2 \in \mathbb{R}^+$ computation steps. Observe that $m \leq N(m-1, n)$ for $n \geq 1$, which shows that $Q_1 + Q_H Q_2$ can be computed in less than $C_3 N(m-1, n)N(m, n-1)$, $C_3 \in \mathbb{R}^+$ computation steps. It is

$N(m, n) = N(m-1, n) + N(m, n-1)$. Hence, by assuming $2C \geq C_3$, we have that

$$C(N(m-1, n)^2 + N(m, n-1)^2) + C_3 N(m-1, n)N(m, n-1) \leq CN(m, n)^2,$$

proving (ii).

We obtain the storage complexity by using an analogous induction argument. If $N(m, n) = 1$ then $m = 0$ or $n = 0$ and $f : \{\text{point}\} \rightarrow \mathbb{R}$ is interpolated by evaluating f at one point. Therefore we have to store that point yielding storage complexity $\mathcal{O}(1)$. If $N(m, n) > 1$, using the same splitting of the problem as above, induction yields that there is $D \in \mathbb{R}^+$ such that we have to store at most $DN(m-1, n)$ and $DN(m, n-1)$ numbers for each of the two sub-problems, respectively. Altogether we need to store $D(N(m-1, n) + N(m, n-1)) = DN(m, n)$ numbers, proving the storage complexity. \square

So far, we have provided an existence result stating that the PIP can be solved efficiently in $\mathcal{O}(N(m, n)^2)$. The derivation of an actual algorithm based on the recursion implicitly used in Theorem 6 is given in the next section.

6 Recursive Decomposition of the PIP

Based on the recursion expressed in Theorems 4 and 5, we can derive an efficient algorithm to compute solutions to the PIP in a numerically robust and computationally efficient way. The resulting algorithm derived hereafter, called PIP-SOLVER is based on a binary tree $T_{m,n}$ as a straightforward data structure resulting from the recursive problem decomposition. It also uses a *multivariate divided difference scheme* to solve the sub-problems, leading to better numerical robustness and accuracy than other approaches (comparisons will be given in Section 8).

6.1 Multivariate Newton Polynomials

The essential data structure required to formulate multivariate Newton interpolation is given by the following binary tree:

Definition 6 (PIP tree) Let $m, n \in \mathbb{N}$. We define a binary tree $T_{m,n} = (\mathcal{V}_{m,n}, \mathcal{E}_{m,n})$ with vertex labeling $\sigma : \mathcal{V}_{m,n} \rightarrow \mathbb{N} \times \mathbb{N}$, $\sigma(v) = (\dim(v), \deg(v))$ as follows: We start with a root ρ and set $\dim(\rho) = m$, $\deg(\rho) = n$. For a vertex $v \in \mathcal{V}_{m,n}$ with $\dim(v) > 0$ and $\deg(v) > 0$ we introduce a left child l with label $\sigma(l) = (\dim(v) - 1, \deg(v))$ and a right child r with label $\sigma(r) = (\dim(v), \deg(v) - 1)$.

Furthermore, we consider the set $\Gamma_{m,n}$ of *leaf paths* γ starting at ρ and ending in a leaf $v_\gamma \in L_{m,n} \subset \mathcal{V}_{m,n}$ of $T_{m,n}$. For every leaf path $\gamma \in \Gamma_{m,n}$ we define

$$[\gamma] = (k_1, \dots, k_m)^\top \in \mathbb{N}^m, \quad k_i = \#\{v \in \gamma : \dim(v) = i\} \quad (7)$$

to be its *descent vector*, which uniquely identifies the leaf path γ .

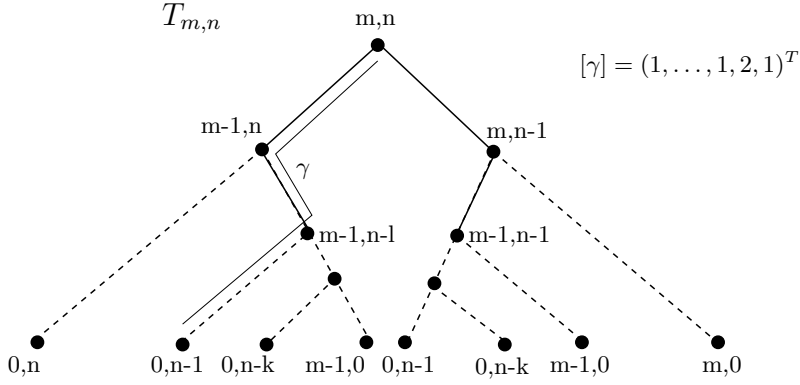


Fig. 1 The binary tree $T_{m,n}$ and a leaf path $\gamma \in \Gamma_{m,n}$ with descent vector $[\gamma]$.

Figure 6 illustrates the tree $T_{m,n}$ and the concept of a leaf path γ . As one can easily verify, the depth of $T_{m,n}$ is given by $\text{depth}(T_{m,n}) = m + n - 1$, while the total number $\#L_{m,n}$ of leaves of $T_{m,n}$ is given by $N(m, n)$.

Next we introduce the multidimensional generalization of Newton nodes, given by a *non-uniform, affine-transformed, sparse T-grid*:

Definition 7 (Multidimensional Newton nodes) Given $m, n \in \mathbb{N}$ and a set of nodes $P_{m,n} \subseteq \mathbb{R}^m$ with $\#P_{m,n} = N(m, n)$, we say $P_{m,n}$ are *multidimensional Newton nodes* generated by

$$\mathbb{P}_{m,n} := \bigoplus_{i=1}^m \mathbb{P}_i, \quad \mathbb{P}_i = \{p_{i,1}, \dots, p_{i,n+1}\} \subseteq \mathbb{R}$$

if and only if there exists an affine transformation $\tau : \mathbb{R}^m \rightarrow \mathbb{R}^m$ such that for every $p \in P_{m,n}$ there is exactly one $\gamma \in \Gamma_{m,n}$ with $[\gamma] = (k_1, \dots, k_m)^T \in \mathbb{N}^m$ such that:

$$\tau(p) = (\bar{p}_{1,k_1}, \dots, \bar{p}_{m,k_m})^T, \quad \text{where we set } \bar{p}_{k_l} = \begin{cases} p_{l,n+1} & \text{if } k_l = 0, \forall l' < l \\ p_{l,k_l} & \text{else.} \end{cases}$$

If $\tau = \text{id}_{\mathbb{R}^m}$, we call $P_{m,n}$ *canonical multidimensional Newton nodes*.

Considering the special case of canonical multidimensional Newton nodes, and assuming $\mathbb{P}_i = \mathbb{P}_{i'}$ for all occurring $1 \leq i, i' \leq m$, an intuition of the definition in terms of sparse T -grids reveals. Indeed, T -grids are based on decomposing space along a binary tree, as we do here. Furthermore, the tree $T_{m,n}$ is a sparse binary tree, i.e., the full tree depth is only reached by few of the leaves.

Note that the $P_{m,n}$ can be recursively split into subsets $P_1, P_2 \subseteq P_{m,n}$ with $P_1 \subseteq H$ for some hyperplane $H \subseteq \mathbb{R}^m$, such that the assumptions of Theorem 4 are satisfied for every recursion step. Even more, the recursion satisfies the *defining property of Newton nodes*, which is that the projection

$$\pi_H(P_2) \subseteq H \text{ of } P_2 \text{ onto } H \text{ satisfies } \pi_H(P_2) \subseteq P_1. \quad (8)$$

This fact is going to be the key ingredient for the proof of Theorem 7.

Definition 8 (Multivariate Newton polynomials) Given $m, n \in \mathbb{N}$, $T_{m,n}$, $\Gamma_{m,n}$, and multidimensional Newton nodes $P_{m,n}$ generated by $\mathbb{P}_{m,n}$. For $\gamma \in \Gamma_{m,n}$ with descent vector $[\gamma] = (k_1, \dots, k_m)^\top \in \mathbb{N}^m$ and $x = (x_1, \dots, x_m)^\top \in \mathbb{R}^m$, we call

$$N_{[\gamma]}(x) = \prod_{i=1}^m \prod_{l=1}^{k_i-1} (x_i - p_{i,l}) \quad (9)$$

the *multivariate Newton polynomials* with respect to $P_{m,n}$.

Proposition 3 Let $m, n \in \mathbb{N}$, $f : \mathbb{R}^m \rightarrow \mathbb{R}$ be a function, and $P_{m,n}$ a set of multidimensional Newton nodes generated by $\mathbb{P}_{m,n}$. Then, there exist unique coefficients $c_{[\gamma]} \in \mathbb{R}$, $\gamma \in \Gamma_{m,n}$, such that:

$$Q(x) = \sum_{\gamma \in \Gamma(m,n)} c_{[\gamma]} N_{[\gamma]}(x)$$

satisfies $Q(p) = f(p)$ for all $p \in P_{m,n}$. Since $Q \in \Pi_{m,n}$, Q is the unique solution of the PIP with respect to $(m, n, f, P_{m,n})$, implying that $P_{m,n}$ is unisolvent.

Proof We argue by induction on $N(m, n)$. If $N(m, n) = 1$ then $P_{m,n} = \{p\}$ consists of one point. Therefore, $Q(x) = f(p)$ and the claim holds. If $N(m, n) > 1$, then we can assume w.l.o.g. that $P_{m,n}$ are canonical, i.e., that the affine transformation τ in Definition 7 is the identity, $\tau = \text{id}_{\mathbb{R}^m}$ (see Lemma 1 and Definition 7). We consider

$$P_1 = \{p = (p_1, \dots, p_m)^\top \in P_{m,n} \subseteq \mathbb{R}^m : p_m = p_{m,1} \in \mathbb{P}_m\},$$

$P_2 = P_{m,n} \setminus P_1$, and $H = \{x \in \mathbb{R}^m : x_m = p_{m,1}\}$. Now note that P_1 is a multidimensional Newton node set on H w.r.t. $(m-1, n)$ and P_2 is a multivariate Newton node set w.r.t. $(m, n-1)$ on \mathbb{R}^m . Hence, by induction, P_1, P_2 are unisolvent. Thus, P_1, P_2 , and H satisfy the assumptions of Theorem 4, which completes the proof. \square

Indeed, in dimension $m = 1$ the $N_{[\gamma]}(x)$ correspond to the classical Newton polynomials from Eq. (2). Given canonical multidimensional Newton nodes $P_{m,n}$, and ordering the paths $\Gamma_{m,n}$ from left to right with respect to the leaf ordering of $T_{m,n}$, we observe that for the multidimensional Newton nodes $p_{[\gamma]}$ corresponding to γ there holds

$$N_{[\gamma]}(p_{[\gamma']}) = 0 \quad \text{if } \gamma < \gamma'. \quad (10)$$

Thus, like in the 1D case, the associated Vandermonde matrix $V_{m,n}(P_{m,n})$ is of lower triangular form, yielding an alternative proof of Proposition 3 and a possibility of computing the coefficients $c_{[\gamma]}$ in $\mathcal{O}(N(m, n)^2)$ steps. Just like in the 1D case, a divided difference scheme can be formulated for computing the coefficients $c_{[\gamma]}$ even more efficiently.

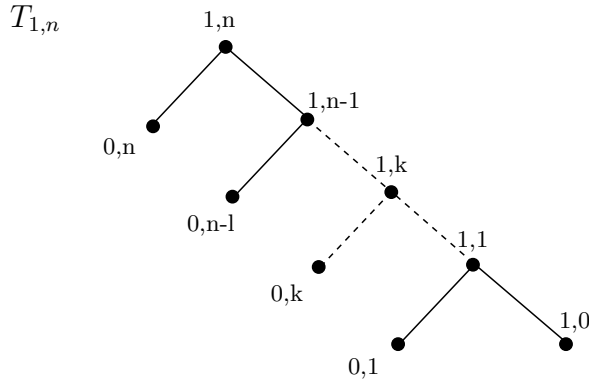


Fig. 2 The binary tree $T_{1,n}$ in the 1-dimensional case.

6.2 Alternative Formulation of the 1D Divided Differences Scheme

We consider the case of dimension $m = 1$. In this case, the tree $T_{1,n}$ has the special form illustrated in Figure 2. Choosing pairwise disjoint nodes $P_n = \{p_1, \dots, p_{n+1}\} \subseteq \mathbb{R}$ and enumerating the leaf paths $\gamma_i \in \Gamma_{1,n}$ from left to right, we observe that

$$N_{[\gamma_i]} = N_i(x) = \prod_{j=1}^i (x_j - p_j) \quad i = 0, \dots, n,$$

where the Newton polynomials N_i were defined in Eq. (2). Though it is trivial, the nodes P_n can be seen as lying on the 0-dimensional planes $H_i = \{x \in \mathbb{R}^1 : x = p_i\} = \{p_i\}$. While Theorem 4 is fulfilled anyhow in this case, this interpretation allows to recursively set up the statement of Theorem 5. That is, for a given computable function $f : \mathbb{R} \rightarrow \mathbb{R}$ and $k = 1, \dots, n$ we set:

$$\begin{aligned} f_0(x) &= f(x), & c_0 &= f_0(p_1), \\ f_k(x) &= \frac{f_{k-1}(x) - f_{k-1}(p_k)}{x - p_k}, & c_k &= f_k(p_{k+1}). \end{aligned} \quad (11)$$

Indeed the computation of the c_k is similar to the divided difference scheme illustrated in Eq. (4). However, the two schemes are not identical. In our alternative formulation, the differences and divisors appear in a different order. Still, both schemes compute the same result, as can easily be verified by hand.

Corollary 1 *Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be a computable function, and $T_{1,n}, \Gamma_{1,n}$ be given. Further let $P_n = \{p_1, \dots, p_{n+1}\} \subseteq \mathbb{R}$ be a set of pairwise disjoint nodes. Then, the polynomial*

$$Q(x) = \sum_{k=0}^n c_k N_{[\gamma_k]}(x)$$

that solves the 1-dimensional PIP with respect to (n, f, P_n) can be determined in $\mathcal{O}(n^2)$ computation steps.

Proof By induction on n , the statement follows directly from Proposition 3 and Theorem 5. The runtime complexity can be shown analogously to the proof of Theorem 6. \square

6.3 The Multivariate Divided Differences Scheme

In the general case of dimension $m \in \mathbb{N}$, we introduce the following notions:

Definition 9 (parallel paths) Let $m, n \in \mathbb{N}$, and $T_{m,n}, \Gamma_{m,n}$ be given. Let $\gamma \in \Gamma_{m,n}$ with descent vector $[\gamma] = (k_1, \dots, k_m) \in \mathbb{N}^m$. We consider $1 \leq i \leq m$ with $k_i > 1$. If the leaf l_γ where the path encoded by γ ends has degree $\deg(l_\gamma) > 0$, then we define

$$\Delta_i(\gamma) := \left\{ \delta \in \Gamma_{m,n} : [\delta] = (l_1, \dots, l_m) \text{ with } \begin{array}{l} 1 \leq l_i < k_i, \\ l_j = k_j, \quad \text{for } j \neq i \end{array} \right\}.$$

If the leaf degree $\deg(l_\gamma) = 0$ then for $i = 1$ we consider

$$\Delta_1^0(\gamma) := \{ \delta \in \Gamma_{m,n} : [\delta] = (l_1, k_2, \dots, k_m), \quad l_1 = 1, \dots, k_1 - 1 \}$$

and for $i > 1$ we define $i_0 = \min(\{j < i : k_j > 1\} \cup \{1\})$ and

$$\Delta_i^0(\gamma) := \left\{ \delta \in \Gamma_{m,n} : [\delta] = (l_1, \dots, l_m) \text{ with } \begin{array}{l} l_i < k_i, \\ l_j = k_j, \quad j \neq i, i_0 \\ l_{i_0} = k_i - l_i + k_{i_0}, \end{array} \right\}.$$

Finally, we order the paths in $\Delta_i(\gamma), \Delta_i^0(\gamma)$ with respect to the leaf ordering of $T_{m,n}$ from left to right.

Indeed, the $\Delta_i(\gamma), \Delta_i^0(\gamma)$ correspond to all multidimensional Newton nodes that are required to compute the coefficient of $N_{[\gamma]}$. For instance, the leaf l_δ of $\delta \in \Delta_i^0(\gamma)$, $i > 1$, have degree $\deg(l_\delta) = 0$ and are given by parallel translation of l_γ along the multidimensional Newton grid $P_{m,n}$ in dimension i . More precisely, we define:

Definition 10 (multivariate divided differences) Let $m, n \in \mathbb{N}$, $f : \mathbb{R}^m \rightarrow \mathbb{R}$ be a computable function, $P_{m,n}$ be a set of multidimensional Newton nodes, and $T_{m,n}, \Gamma_{m,n}$ be given. For $\gamma \in \Gamma_{m,n}$ with $[\gamma] = (k_1, \dots, k_m)$, we consider $\Delta_i(\gamma), \Delta_i(\gamma)^0 = \{\delta_{1,i}, \dots, \delta_{k_i,i}\}$, and $m_\gamma \in \mathbb{N}$ such that $J_\gamma = \{j_1, \dots, j_{m_\gamma}\} \subseteq \{1, \dots, m\}$ contains all indices with $k_i > 1$ for $i \in J_\gamma$. By reordering if necessary we assume that J_γ is ordered in reverse direction, i.e., $j_l > j_{l'}$ if $l < l'$ for all $l, l' \in \{1, \dots, m_\gamma\}$. We denote by $p_{h,i} \in P_{m,n}$ the nodes corresponding to $\delta_{h,i}$ and define

$$\begin{aligned} F_{\gamma,1}^0 &= f(p_{j_1,1}), & F_{\gamma,1}^k &:= \frac{F_{\gamma,1}^{k-1}(p_{j_1,k}) - F_{\gamma,1}^{k-1}(p_{j_1,k-1})}{(p_{j_1,k-1} - p_{j_1,k})_{j_1}}, \quad 1 \leq k \leq k_{j_1}, \\ F_{\gamma,i}^0 &:= F_{\gamma,i-1}^{k_{j_i}}, & F_{\gamma,l}^k &:= \frac{F_{\gamma,l}^{k-1}(p_{j_l,k}) - F_{\gamma,l}^{k-1}(p_{j_l,k-1})}{(p_{j_l,k-1} - p_{j_l,k})_{j_l}}, \quad 1 \leq k \leq k_{j_l}, \quad 1 \leq l \leq m_\gamma, \end{aligned}$$

where $(p_{j_l, k-1} - p_{j_l, k})_{j_l}$ denotes the j_l -th coordinate/component of $(p_{j_l, k-1} - p_{j_l, k})$.

This definition yields the alternative divided difference scheme given in Eq. (11) for the special case $m = 1$.

Theorem 7 *Let $m, n \in \mathbb{N}$, $f : \mathbb{R}^m \rightarrow \mathbb{R}$ be a computable function, $T_{m,n}$, $\Gamma_{m,n}$ be given, and $P_{m,n} \subseteq \mathbb{R}^m$ be a set of canonical multidimensional Newton nodes generated by $\mathbb{P}_{m,n}$. Then, the polynomial*

$$Q(x) = \sum_{\gamma \in \Gamma_{m,n}} c_{[\gamma]} N_{[\gamma]}(x), \quad c_{[\gamma]} = F_{\gamma, m_{\gamma}}^{k_{j_m \gamma}},$$

with $[\gamma] = (k_1, \dots, k_{j_{m_{\gamma}}}, \dots, k_{j_1}, \dots, k_m)$ solves the PIP with respect to $(m, n, f, P_{m,n})$ and can be computed in $\mathcal{O}(N(m, n)^2)$ operations.

Proof We argue by induction on (m, n) . For $m = 1, n \in \mathbb{N}$, the statement is already proven in Corollary 1. Thus, we proceed by induction on m . If $m > 1$, then we consider $\Gamma_1 = \{\gamma_1, \dots, \gamma_{N(m-1, n)}\} \subseteq \Gamma_{m,n}$ with $[\gamma_h] = (k_1, \dots, k_m) \in \mathbb{N}^m$ and $k_m = 1, h = 1, \dots, N(m-1, n)$. Furthermore, we denote by $P_1 \subseteq P_{m,n}$ the corresponding nodes and $\Gamma_2 = \Gamma_{m,n} \setminus \Gamma_1, P_2 = P_{m,n} \setminus P_1$. In addition, we consider the sub-trees $T_1 \cong T_{m-1, n}, T_2 \cong T_{m, n-1}$ spanned by Γ_1, Γ_2 , respectively. Note that the m -th coordinate is constant, i.e., for $p, p' \in P_1$ we have $p_m = p'_m = p_{m,1} \in \mathbb{P}_{m,n}$. Since the hyperplane $H = \{x \in \mathbb{R}^m : x_m = p_{m,1}\}$ can be identified with \mathbb{R}^{m-1} by shifting the last coordinate to 0, induction yields that

$$c_{[\gamma_1]} = F_{\gamma_1, m_{\gamma_1}}^{k_{j_m \gamma_1}}, \dots, c_{[\gamma_{N(m-1, n)}]} = F_{\gamma_{N(m-1, n)}, m_{\gamma_{N(m-1, n)}}}^{k_{j_m \gamma_{N(m-1, n)}}}.$$

These are the uniquely determined coefficients solving the PIP with respect to $(m-1, n, f|_H, P_1)$ on H . We further follow Theorem 5 and set $f_1(x) = (f(x) - Q_1(x))/(x_m - p_{m,1})$ with $Q_1(x) = \sum_{\gamma \in \Gamma_1} c_{[\gamma]} N_{[\gamma]}(x)$. The defining property of multidimensional Newton nodes is their parallelism, i.e., $\pi_H(P_2) \subseteq P_1$, where π_H denotes the orthogonal projection onto H . Furthermore, we observe that $Q_1(x)$ is constant in direction perpendicular to H . Therefore,

$$Q_1(p) = Q_1(\pi_H(p)) = f(\pi_H(p)) \quad \text{for all } p \in P_2.$$

Thus,

$$\begin{aligned} f_1(p) &= (f(p) - f(\pi_H(p)))/(\pi_H(p)_m - p_{m,1}) \\ &= (f(p) - f(p_1, \dots, p_{m-1}, p_{m,1}))/(\pi_H(p)_m - p_{m,1}). \end{aligned}$$

Following Definition 10, we observe that $d_{[\delta_1]} = F_{\delta_1, K_{\delta_1}}^{k_{j_m \delta_1}}, \dots, d_{[\delta_{N(m-1, n)}]}$
 $= F_{\delta_{N(m, n-1)}, K_{\delta_{N(m, n-1)}}}^{k_{j_m \delta_{N(m, n-1)}}}$ yielding

$$d_{[\delta_1]} = c_{[\gamma_{N(m-1, n)+1}]} \dots, d_{[\delta_{N(m-1, n)}]} = c_{[\gamma_{N(m, n)}]} \quad (12)$$

Hence, by induction, the $d_{[\delta]}$, $\delta \in \Gamma_{m,n-1}$, are the uniquely determined coefficients solving the PIP w.r.t. $(m, n-1, f_1, P_2)$ on \mathbb{R}^m . Moreover, due to Eq. (12), we have:

$$Q_2(x) = \sum_{\delta \in \Gamma_{m-1,n}} d_{[\delta]} N_{[\delta]}(x) = \frac{1}{(x_m - p_{m,1})} \sum_{\gamma \in \Gamma_2} c_{[\gamma]} N_{[\gamma]}(x).$$

Now Theorem 5 yields that

$$Q(x) = Q_1(x) + Q_H(x)Q_2(x) = \sum_{\gamma \in \Gamma_{m,n}} c_{[\gamma]} N_{[\gamma]}(x), \quad Q_H(x) = x_m - p_{m,1},$$

solves the PIP w.r.t. $(m, n, f, P_{m,n})$, which proves the statement. The runtime complexity follows from the proof of Theorem 6. \square

By considering $f' = f \circ \tau^{-1}$ with τ from Definition 7, Theorem 7 also extends to the case of non-canonical multidimensional Newton nodes, i.e., to affine transformations of canonical multidimensional Newton nodes. Together, Theorems 5 and 7 and Proposition 3 prove Main Result I as stated in Theorem 1 in Section 2. We next characterize some basic properties of multivariate Newton polynomials and how they can be exploited for basic computations.

6.4 Properties of Multivariate Newton Polynomials

Analogously to its 1D version, the multivariate divided difference scheme provides an efficient and numerical robust method for interpolating functions by polynomials. We generalize some classical facts of the 1D case, yielding Main Result 2, here restated with a bit more detail as:

Theorem 8 (Main Result II) *Let $m, n \in \mathbb{N}$, $T_{m,n}$, $\Gamma_{m,n}$, and a set of canonical multidimensional Newton nodes $P_{m,n}$ be given. Let further $Q \in \Pi_{m,n}$, $Q(x) = \sum_{\gamma \in \Gamma_{m,n}} c_{[\gamma]} N_{[\gamma]}(x)$ be a polynomial given in multivariate Newton form w.r.t. $T_{m,n}$, $P_{m,n}$. Then, there exist algorithms with runtime complexity in $\mathcal{O}(N(m, n))$ that compute:*

- i) *The value of $Q(x_0)$ for any $x_0 \in \mathbb{R}^m$ in $\mathcal{O}(N(m, n))$;*
- ii) *The partial derivative $\partial_{x_i} Q|_{x_0}$ for any $i \in \{1, \dots, m\}$ and $x_0 \in \mathbb{R}^m$ in $\mathcal{O}(nN(m, n))$;*
- iii) *The integral $\int_{\Omega} Q(x) dx$ for any hypercube $\Omega \subseteq \mathbb{R}^m$ with runtime complexity in $\mathcal{O}(nN(m, n))$.*

Proof We argue by induction on $m, n \in \mathbb{N}$. For $m = 1, n \in \mathbb{N}$ all statements are known to be true [41, 19, 2, 13]. If $m > 1$ then, by assumption, the affine transformation $\tau : \mathbb{R}^m \rightarrow \mathbb{R}^m$ from Definition 7 is the identity, i.e., $\tau = \text{id}_{\mathbb{R}^m}$. Let $\mathbb{P}_{m,n} := \bigoplus_{i=1}^m \mathbb{P}_i$, $\mathbb{P}_i = \{p_{i,1}, \dots, p_{i,n+1}\}$ be the generating nodes of $P_{m,n}$, see Definition 7. We consider the hyperplane $H = \{x \in \mathbb{R}^m : x_m = p_{m,1}\}$ and,

for $\Gamma_1 = \{\gamma \in \Gamma_{m,n} : [\gamma]_m = 1\}$, we denote by $P_1 \subseteq P_{m,n}$ the multidimensional Newton nodes corresponding to the leaves of $\gamma \in \Gamma_1$. Further, we set $Q_1(x) = \sum_{\gamma \in \Gamma_1} c_{[\gamma]} N_{[\gamma]}(x)$. By identifying H with \mathbb{R}^{m-1} induction yields that (i), (ii), (iii) can be computed in $\mathcal{O}(N(m-1, n))$. Setting $Q_2(x) = Q(x) - Q_1(x)$, we have $Q_2(x) = (x_m - p_{m,1})Q_3(x)$, where $Q_3 \in \Pi_{m,n-1}$ is a polynomial in Newton form w.r.t. $T_{m,n-1}$ and $P_2 = P_{m,n} \setminus P_1$ (see Theorem 7). Thus, by induction, (i), (ii) can be computed in $\mathcal{O}(N(m, n-1))$, $\mathcal{O}(nN(m, n-1))$ for Q_3 , which proves (i), (ii) because $N(m-1, n) + N(m, n-1) = N(m, n)$. Claim (iii) follows from (i) and (ii) by applying partial integration to $Q(x) = Q_1(x) + (x_m - p_{m,1})Q_3(x)$, i.e.,

$$\begin{aligned} \int_{\Omega} Q(x) dx &= \int_{\Omega} Q_1(x) dx - \int_{\Omega} Q_3(x) dx \\ &\quad + \int_{\Omega \cap H_{1, \dots, m-1}} (x_m - p_{m,1}) \partial_{x_m} Q_3(x) \Big|_{x_m = \pm 1} dx \end{aligned}$$

and using induction. \square

The recursive subdivision of the problem into sub-problems of lower dimension or degree, as also used in the proof of Theorem 8, can be used to implement a generalization of the classical (*inverse*) *Horner scheme* [41, 19, 2, 13]. In the case of arbitrary multidimensional Newton nodes $P_{m,n}$ with $\tau(P_{m,n}) = \bar{P}_{m,n}$ for canonical nodes $\bar{P}_{m,n}$, we can evaluate $Q \circ \tau^{-1}(x)$ with respect to $\bar{P}_{m,n}$. The runtime complexity then increases by adding the cost of inverting the affine transformation τ from Definition 7, which is given by the cost of inverting an $m \times m$ matrix.

7 Approximation Theory

Studying how well arbitrary continuous functions can be approximated by polynomials, a fundamental observation was made. The observation is that even though the Weierstrass Theorem states that every function $f : I \subseteq \mathbb{R} \rightarrow \mathbb{R}$, $I = [a, b]$ can be approximated by Bernstein polynomials in the C^0 -sense [12], for every choice of interpolation nodes P_n there exists a continuous function for which the interpolant $Q_{n,f}$ does not converge, i.e., $\|f - Q_{n,f}\|_{C^0(\Omega)} \not\rightarrow 0$ for $n \rightarrow \infty$ [15]. This observation is known as *Runge's phenomenon*. However, when choosing *Chebyshev nodes* the class of functions that cannot be approximated (i.e., for which Runge's phenomenon occurs) seem to be pathological and extremely unlikely to occur in any "real world" application or data set. We here specify these facts and provide generalized statements for the multidimensional case.

7.1 Sobolev Theory

We start by providing the analytical setting required to formulate our results. In [1] an excellent overview of Sobolev theory is given. Here, we simply recall

that for every domain $\Omega \subseteq \mathbb{R}^m$ the functional space $C^k(\Omega, \mathbb{R})$ denotes the real vector space of all functions $f : \Omega \rightarrow \mathbb{R}$ that are k times differentiable in the interior of Ω and are continuous on the closure $\bar{\Omega} = \Omega \cup \partial\Omega$. Further, by equipping $C^k(\Omega, \mathbb{R})$ with the norm

$$\|f\|_{C^k(\Omega)} = \sum_{l=0}^k \sup_{x \in \Omega} \sum_{\alpha \in A_{m,l}} |\partial^\alpha f(x)|, \quad \partial^\alpha f(x) = \partial_{x_1}^{\alpha_1} \cdots \partial_{x_n}^{\alpha_n} f(x),$$

we obtain a Banach space. We consider the Sobolev space of L^p -functions with well-defined weak derivatives up to order k , i.e., in general for $1 \leq p < \infty$

$$W^{k,p} = \left\{ f \in L^p(\Omega, \mathbb{R}) : \|f\|_{W^{k,p}(\Omega)}^p := \sum_{\alpha \in A_{m,k}} \|\partial^\alpha f\|_{L^p(\Omega)}^p < \infty \right\}. \quad (13)$$

The *Hilbert space* $H^k(\Omega, \mathbb{R}) := W^{k,2}(\Omega, \mathbb{R})$ is of special interest. In the following, we assume that $\Omega = [-1, 1]^m$ is the standard hypercube and denote by $\mathbb{T}^m = \mathbb{R}^m / 2\mathbb{Z}^m$ the torus with fundamental domain Ω . Then, $H^k(\Omega, \mathbb{R}) \subseteq L^2(\Omega, \mathbb{R})$, $0 \leq k \leq \infty$, implies that every $f \in H^k(\Omega, \mathbb{R})$ can be written as a power series

$$f(x) = \frac{1}{|\Omega|} \sum_{\alpha \in \mathbb{N}^m} d_\alpha x^\alpha, \quad d_\alpha \in \mathbb{R}$$

almost everywhere, i.e., the identity is violated only on a set $\Omega' \subseteq \Omega$ of Lebesgue measure zero. Further, for every function $\tilde{f} : \mathbb{T}^m \rightarrow \mathbb{R}$, there is a uniquely determined function $f : \Omega \rightarrow \mathbb{R}$ such that $\tilde{f}(x + 2\mathbb{Z}^m) = f(x)$. Therefore, f can obviously be extended to a periodic function on \mathbb{R}^m and is called the *periodic representative* of \tilde{f} . Thus, we can write f as a Fourier series

$$f(x) = \frac{1}{|\Omega|} \sum_{\alpha \in \mathbb{N}^m} c_\alpha e^{\pi i \langle \alpha, x \rangle}, \quad c_\alpha \in \mathbb{C}$$

and observe that the space $H^k(\mathbb{T}^m, \mathbb{R})$ can be defined as

$$H^k(\mathbb{T}^m, \mathbb{R}) = \overline{C^\infty(\mathbb{T}^m, \mathbb{R})}^{\|\cdot\|_{H^k(\Omega)}}, \quad (14)$$

which is the completion of $C^\infty(\mathbb{T}^m, \mathbb{R})$ with respect to the H^k -norm [1]. Therefore,

$$\begin{aligned} \|f\|_{H^k(\Omega)}^2 &:= \sum_{\beta \in A_{k,m}} \langle \partial^\beta f, \partial^\beta f \rangle_{L^2(\Omega)} \\ &= \sum_{\beta \in A_{m,k}, \alpha \in \mathbb{N}^m} (\pi^{|\beta|} |\alpha^\beta| |c_\alpha|)^2, \end{aligned} \quad (15)$$

Thus, $C^\infty(\mathbb{T}^m, \mathbb{R}) \subseteq H^k(\mathbb{T}^m, \mathbb{R})$ is a dense subset and, due to the right-hand side of Eq. (15), a notion of fractal derivatives can be given, i.e., $H^k(\Omega, \mathbb{R})$ with $k \in \mathbb{R}^+$ and $\alpha \in \mathbb{R}^m$, $\alpha_i \geq 1$, $|\alpha| \leq k$, can be considered. Vice versa, due

to the *Sobolev and Rellich-Kondrachov embedding Theorem* [1], we have that whenever $k > m/2$ then $H^k(\mathbb{T}^m, \mathbb{R}) \subseteq C^0(\mathbb{T}^m, \mathbb{R})$ and the embedding

$$i : H^k(\mathbb{T}^m, \mathbb{R}) \hookrightarrow C^0(\mathbb{T}^m, \mathbb{R})$$

is well defined, continuous, and compact. Thus, for $m \in \mathbb{N}$ and all periodic representatives f of $\tilde{f} \in C^0(\mathbb{T}^m, \mathbb{R})$, there exists a constant $c = c(m, \Omega) \in \mathbb{R}^+$ such that

$$\|f\|_{C^0(\Omega)} \leq c \|f\|_{H^k(\Omega)} \quad (16)$$

and for every $B \subseteq H^k(\mathbb{T}^m, \mathbb{R})$ we have that $i(B) \subseteq C^0(\mathbb{T}^m, \mathbb{R})$ is precompact whenever B is bounded in $H^k(\mathbb{T}^m, \mathbb{R})$. By the trace Theorem [1], we observe furthermore that whenever $H \subseteq \mathbb{R}^m$ is a hyperplane of co-dimension 1, then the induced restriction

$$\varrho : H^k(\Omega, \mathbb{R}) \longrightarrow H^{k-1/2}(\Omega \cap H, \mathbb{R}) \quad (17)$$

is continuous, i.e., $\|f|_{\Omega \cap H}\|_{H^{k-1/2}(\Omega \cap H)} \leq d \|f\|_{H^k(\Omega)}$ for some $d = d(m, \Omega) \in \mathbb{R}^+$. We consider

$$\Theta_{m,n} = \left\{ \tilde{f} \in H^k(\mathbb{T}^m, \mathbb{R}) : f(x) = \frac{1}{|\Omega|} \sum_{\alpha \in A_{m,n}} c_\alpha e^{\pi i \langle \alpha, x \rangle}, c_\alpha \in \mathbb{C} \right\}$$

the space of all finite Fourier series of bounded frequencies and denote by

$$\begin{aligned} \theta_n : H^k(\mathbb{T}^m, \mathbb{R}) &\longrightarrow \Theta_{m,n} \subseteq C^0(\mathbb{T}^m, \mathbb{R}), \\ \tau_n : H^k(\Omega, \mathbb{R}) &\longrightarrow \Pi_{m,n} \subseteq C^0(\Omega, \mathbb{R}) \end{aligned} \quad (18)$$

the corresponding projections onto $\Theta_{m,n}$ or onto the space of polynomials $\Pi_{m,n}$. Further, we denote by $\theta_n^\perp = I - \theta_n$, $\tau_n^\perp = I - \tau_n$ the complementary projections. Then, we have:

Lemma 2 *Let $k, m \in \mathbb{N}$, $\Omega = [-1, 1]^m \subseteq \mathbb{R}^m$ be the standard hypercube, $\tilde{f} \in H^k(\mathbb{T}^m, \mathbb{R})$, and f its periodic representative. Then:*

- i) $\|f\|_{C^0(\Omega)} \leq \|f\|_{H^k(\Omega)}$ for all $\tilde{f} \in H^k(\mathbb{T}^m, \mathbb{R})$.
- ii) $\|\theta_n^\perp(f)\|_{C^0(\Omega)} \in o((m/n)^k)$, i.e.

$$(n/m)^k \|\tau_n^\perp(f)\|_{C^0(\Omega)} \xrightarrow{n \rightarrow \infty} 0 \quad \text{for every } m \in \mathbb{N}.$$

- iii) For $n, l \in \mathbb{N}$, the operator norm of $\tau_n^\perp \theta_l : H^k(\mathbb{T}^m, \mathbb{R}) \longrightarrow C^0(\mathbb{T}^m, \mathbb{R})$ is bounded by $\|\tau_n^\perp \theta_l\| \leq N(m, l)^{1/2} \frac{(\pi l^2)^{n+1}}{(n+1)!}$.

Proof To show (i), we approximate f by a finite Fourier series, i.e., we assume that $f(x) = \frac{1}{|\Omega|} \sum_{\alpha \in A_{m,n}} c_\alpha e^{\pi i \langle \alpha, x \rangle}$. The Fourier basis is orthonormal, i.e.,

$$\frac{1}{|\Omega|} \int_{\Omega} e^{\pi i \langle \alpha, x \rangle} \cdot e^{-\pi i \langle \beta, x \rangle} d\Omega = \begin{cases} 1, & \text{if } \alpha = \beta \\ 0 & \text{else.} \end{cases}$$

Thus, we compute

$$\|f\|_{C^0(\Omega)}^2 = \sup_{x \in \Omega} \left| \sum_{\alpha \in A_{m,n}} c_\alpha e^{\pi i \langle \alpha, x \rangle} \right|^2 \leq \sum_{\alpha \in A_{m,n}} |c_\alpha|^2 = \|f\|_{L^2(\Omega)}^2 \leq \|f\|_{H^k(\Omega)}^2.$$

Now (i) follows from the density of the approximation given in Eq. (14) and the continuity of the norm $\|\cdot\|_{H^k(\Omega)}$. To show (ii), let $\alpha \in \mathbb{N}^m$ and observe that

$$\partial^\beta (c_\alpha e^{\pi i \langle \alpha, x \rangle}) = (i\pi)^{|\beta|} \alpha^\beta e^{\pi i \langle \alpha, x \rangle}, \quad \beta \in A_{m,k}.$$

If $|\alpha| > m$ then at least one $\alpha_i > n/m$ for some $1 \leq i \leq m$. Thus, by choosing $\beta \in A_{m,k}$ with $\beta_i = k$, we obtain

$$|\partial^\beta (c_\alpha e^{\pi i \langle \alpha, x \rangle})| \geq (\pi n/m)^k |c_\alpha|, \quad \text{for } |\alpha| \geq m.$$

Hence, for $n > m$

$$(n/m)^{2k} \|\theta_n^\perp f\|_{C^0(\Omega)}^2 \leq \|\theta_n^\perp f\|_{H^k(\Omega)}^2.$$

Due to Eq. (13) we have $\|\theta_n^\perp f\|_{H^k(\Omega)} \xrightarrow{n \rightarrow \infty} 0$, which yields $\|\theta_n^\perp f\|_{C^0(\Omega)} \in o((m/n)^k)$ for every fixed $m \in \mathbb{N}$, proving (ii). To show (iii), we assume $\|f\|_{H^k(\Omega)} \leq 1$, write

$$\theta_l(f) = \sum_{\alpha \in A_{m,l}} c_\alpha e^{\pi i \langle \alpha, x \rangle}, \quad c_\alpha \in \mathbb{C},$$

and recall that for every $x \in \mathbb{R}^m$ there exists $\xi_x \in \Omega$ such that

$$e^{i\pi \langle \alpha, x \rangle} = \sum_{n \in \mathbb{N}} \frac{(i\pi \langle \alpha, x \rangle)^n}{n!} = \sum_{h \leq n} \frac{(i\pi \langle \alpha, x \rangle)^h}{h!} + \frac{\partial_v^{n+1} e^{i\pi \langle \alpha, \xi_x \rangle}}{(n+1)!} (i\pi \langle \alpha, x \rangle)^{n+1},$$

where ∂_v denotes the partial derivative in direction $v = \frac{x}{|x|} \in \mathbb{R}^m$. Therefore, the last term is the *Lagrange remainder*. Now

$$\partial_v^{n+1} e^{i\pi \langle \alpha, \xi_x \rangle} = \partial_v^n \langle \nabla e^{i\pi \langle \alpha, \xi_x \rangle}, v \rangle = \langle \alpha, v \rangle \partial_v^n e^{i\pi \langle \alpha, \xi_x \rangle} = \langle \alpha, v \rangle^{n+1} e^{i\pi \langle \alpha, \xi_x \rangle}$$

and $|\langle \alpha, v \rangle|, |\langle \alpha, x \rangle| \leq |\alpha|$ imply that

$$\left| \frac{\partial_v^{n+1} e^{i\pi \langle \alpha, \xi_x \rangle}}{(n+1)!} (i\pi \langle \alpha, x \rangle)^{n+1} \right| \leq \frac{(\pi |\alpha|)^{2n+2}}{(n+1)!}.$$

Hence, due to $|\alpha| \leq l, |c_\alpha| \leq 1$, we bound

$$\begin{aligned} \|\tau_n^\perp \theta_l(f)\|_{C^0(\Omega)} &\leq \left(\sum_{\alpha \in A_{m,l}} |c_\alpha|^2 \left(\frac{(\pi |\alpha|)^{2n+2}}{(n+1)!} \right)^2 \right)^{1/2} \\ &\leq \frac{\pi^{n+1}}{(n+1)!} \left(\sum_{\alpha \in A_{m,l}} |\alpha|^{4(n+1)} \right)^{1/2} \\ &\leq \frac{\pi^{n+1} l^{2n+2}}{(n+1)!} N(m, l)^{1/2}, \end{aligned}$$

which yields (iii). \square

7.2 Lebesgues Functions

Lebesgue functions measure the relative approximation error of an interpolation scheme. More precisely: Let $m, n \in \mathbb{N}$, $\Omega = [-1, 1]^m \subseteq \mathbb{R}^m$, and $S_{m,n} : C^0(\Omega, \mathbb{R}) \rightarrow \Pi_{m,n}$ denote an interpolation scheme with respect to unisolvent interpolation nodes $P_{m,n}$. That is, for $f \in C^0(\Omega, \mathbb{R})$, $S_{m,n}(f) \in \Pi_{m,n}$ solves the PIP with respect to $(m, n, f, P_{m,n})$. Since $P_{m,n}$ is unisolvent and independent of f , it is readily verified that $S_{m,n}$ is a linear operator. Therefore, the following is well defined:

Definition 11 (Operator norms and Lebesgues functions of interpolation operators) Let $m, n \in \mathbb{N}$, $\Omega = [-1, 1]^m$, and $P_{m,n} \subseteq \Omega$ be unisolvent interpolation nodes. Consider the interpolation operator $S_{m,n} : C^0(\Omega, \mathbb{R}) \rightarrow \Pi_{m,n}$ and its restriction $S_{m,n}|_{H^k(\Omega)} : H^k(\Omega, \mathbb{R}) \rightarrow \Pi_{m,n}$. Then we define by

$$\begin{aligned} \Lambda(P_{m,n}, C^0(\Omega)) &:= \sup_{f \in C^0(\Omega, \mathbb{R}), \|f\|_{C^0(\Omega)}=1} \|S_{m,n}(f)\|_{C^0(\Omega)} \\ \Lambda(P_{m,n}, H^k(\Omega)) &:= \sup_{f \in H^k(\Omega, \mathbb{R}), \|f\|_{H^k(\Omega)}=1} \|S_{m,n}|_{H^k(\Omega)}(f)\|_{C^0(\Omega)} \end{aligned}$$

the *operator norms* of $S_{m,n}$ and $S_{m,n}|_{H^k(\Omega)}$, respectively. Denote with $\mathcal{P}_{m,n}$ the set of all unisolvent node sets with respect to $m, n \in \mathbb{N}$ then

$$\begin{aligned} \Lambda_{m,n,C^0(\Omega)} : \mathcal{P}_{m,n} &\rightarrow \mathbb{R}, & P_{m,n} &\mapsto \Lambda(P_{m,n}, C^0(\Omega)), \\ \Lambda_{m,n,H^k(\Omega)} : \mathcal{P}_{m,n} &\rightarrow \mathbb{R}, & P_{m,n} &\mapsto \Lambda(P_{m,n}, H^k(\Omega)) \end{aligned}$$

are called the *Lebesgue functions* with respect to the considered norms.

Due to Lemma 2,

$$\{f \in H^k(\Omega, \mathbb{R}) : \|f\|_{H^k(\Omega)} = 1\} \subseteq \{f \in C^0(\Omega, \mathbb{R}) : \|f\|_{C^0(\Omega)} = 1\},$$

implying that

$$\Lambda(P_{m,n}, H^k(\Omega)) \leq \Lambda(P_{m,n}, C^0(\Omega)). \quad (19)$$

Understanding the behavior of Lebesgue functions in 1D is crucial for extending their definition to arbitrary dimensions. In particular, the following observation is key to our further considerations:

Lemma 3 Let $l, n \in \mathbb{N}$, $l, n \geq 1$, $\Omega = [-1, 1]^m$, $P_n = \{p_1, \dots, p_{n+1}\}$ be a set of $n+1$ pairwise disjoint nodes, and $P_l = \{p_1, \dots, p_{l+1}\} \subseteq P_n$. Then

$$\Lambda(P_l, C^0(\Omega)) \leq \Lambda(P_n, C^0(\Omega)), \quad \Lambda(P_l, H^k(\Omega)) \leq \Lambda(P_n, H^k(\Omega)) \text{ for } k > m/2.$$

Proof We choose small intervals $I_{h,\varepsilon} = [p_h - \varepsilon, p_h + \varepsilon]$, $I_{h,\delta} = [p_h - \delta, p_h + \delta]$, $h > l+1$ with $\varepsilon > \delta > 0$ small enough so that $I_{h,\varepsilon} \cap P_n = \{p_h\}$. Further, we consider smooth cut-off functions $\beta_h : [p_h - \varepsilon, p_h + \varepsilon] \rightarrow [0, 1]$, with

$\beta_h(p_h \pm \varepsilon) = 1$, $\beta_h^\pm(p_h \pm \delta) = 0$. We denote by $S_{1,k}$, $S_{1,n}$ the interpolation schemes with respect to P_k , P_n , respectively, and set

$$\tilde{f}(x) = \begin{cases} f(x), & \text{if } x \notin I_{h,\varepsilon}, \quad h > k + 1 \\ \beta_h^\pm(x)f(x) + (1 - \beta_h^\pm(x))f(p_{l+1}), & \text{if } x \in [p_h \pm \delta, p_h \pm \varepsilon], \quad h > l + 1 \\ f(p_{l+1}), & \text{if } x \in I_{h,\delta}, \quad h > k + 1. \end{cases}$$

Obviously, it is $\|\tilde{f}\|_{C^0(\Omega)}, \|\tilde{f}\|_{H^k(\Omega)} \leq 1$. Moreover, following the alternative divided difference scheme from Eq. (11), we find $f_j(p_j) = 0$ for all $j > l + 1$. Thus, the coefficients c_j , $j > l$, of the polynomial $S_{1,n}(\tilde{f})$ in Newton form vanish, while the c_0, \dots, c_l are given by $\tilde{f}_0(p_1) = f_0(p_1), \dots, \tilde{f}_l(p_{l+1}) = f_l(p_{l+1})$. Hence, $S_{1,n}(\tilde{f}) = S_{1,l}(f) = S_{1,l}(\tilde{f})$. Therefore,

$$\|S_{1,n}(\tilde{f})\|_{C^0(\Omega)} = \|S_{1,l}(\tilde{f})\|_{C^0(\Omega)} = \|S_{1,l}(f)\|_{C^0(\Omega)}.$$

Observing that f was arbitrarily chosen, this completes the proof. \square

7.3 Newton-Chebyshev Nodes

In order to extend the study of Runge's phenomenon to multiple dimensions, we introduce a multidimensional notion of *Chebyshev nodes* and provide the essential approximation results.

Definition 12 (Multidimensional Newton-Chebyshev nodes) Let $m, n \in \mathbb{N}$, $T_{m,n}$, and $\Gamma_{m,n}$ be given. Let $P_{m,n}$ be the canonical multidimensional Newton nodes generated by

$$\mathbb{P}_{m,n} = \bigoplus_{l=1}^m \text{Cheb}_n,$$

where Cheb_n was defined in Eq. (1). Then, we call $P_{m,n}$ *canonical multidimensional Newton-Chebyshev nodes*, and we call every affine transformation $\tau(P_{m,n})$ of $P_{m,n}$ *multidimensional Newton-Chebyshev nodes*.

Theorem 9 Let $m, n, k \in \mathbb{N}$, $k > m/2$, and $S_{m,n}$ be an interpolation operator with respect to multidimensional Newton nodes $P_{m,n}$ generated by $\mathbb{P}_{m,n} = \bigoplus_{l=1}^m P_l$, $\#P_l = n + 1$. Then

$$\Lambda(P_{m,n}, H^k(\Omega)) \leq \prod_{l=1}^m \Lambda(P_l, H^{k-(m-1)/2}([-1, 1])).$$

If $P_{m,n}$ are multidimensional Newton-Chebyshev nodes, then in particular

$$\Lambda(P_{m,n}, H^k(\Omega)) \leq \Lambda(\text{Cheb}_n, C^0([-1, 1]))^m \in \mathcal{O}(\log(n)^m). \quad (20)$$

Proof We argue by induction on m . For $m = 1$ the claim directly follows from Eq. (19). If $m > 1$ then we can assume w.l.o.g. (i.e., by changing coordinates if necessary) that $P_{m,n}$ are canonical Newton-Chebyshev nodes, and we can consider the hyperplanes H_1, \dots, H_n given by $H_i = \{x \in \mathbb{R}^m : x_m = p_{m,i} \in P_m\}$, $i = 1, \dots, n$. Denote by $Q_{H_i}(x) = x_m - p_{m,i}$ the linear polynomial defining $H_i = Q_{H_i}^{-1}(0)$ and by $\pi_{H_i} : \mathbb{R}^m \rightarrow H_i$, $\pi_{H_i}(x) = (x_1, \dots, x_{m-1}, p_{m,i})$, the corresponding projections. Then, by Theorems 4, 5, and Proposition 3, we have

$$S_{m,n}(f) = S_{m-1,n,H_1}(f) + Q_{H_1} \left(S_{m-1,n-1}(f_1) + Q_{H_2} (S_{m-1,n-2}(f_2) + \dots \right. \\ \left. Q_{H_n} (S_{m-1,1}(f_{n-1}) + S_{m,0}(f_n)) \dots \right), \quad (21)$$

where

$$f_0 = f, \quad f_k = \frac{f_{k-1}(x) - f_{k-1}(\pi_{H_k}(x))}{Q_{H_k}(x)}$$

and the $S_{m-1,n-i}$ interpolate with respect to $P_{m-1,n-i}$ generated by

$$\mathbb{P}_{m-1,n-i} = \{p_{m,i+1}\} \times \bigoplus_{l=1}^{m-1} P_l.$$

By observing that the interpolant $S_{m-1,n-i}(f_i)$ is constant along directions normal to H_i , and by identifying $H_i \cong \mathbb{R}^{m-1}$, induction and Lemma 3 yield

$$\Lambda(P_{m-1,n-i}, H^k(\Omega)) \leq \prod_{l=1}^{m-1} \Lambda(P_l, H^{k-(m-2)/2}([-1, 1]^{m-1})) =: \Lambda^*.$$

Though the f_i are multivariate functions, the remaining interpolation is done with respect to x_m in only 1 variable. Hence, recalling the 1D estimation, we get

$$\|S_{m,n}(f)\|_{C^0(\Omega)} \leq \Lambda^* \|f_0 + Q_{H_1}(f_1 + \dots + Q_{H_n}(f_{n-1} + f_n))\|_{C^0(\Omega)} \\ \leq \Lambda^* \Lambda(P_m, H^{k-(m-1)/2}([-1, 1])) \\ \leq \prod_{l=1}^m \Lambda(P_l, H^{k-(m-1)/2}([-1, 1])).$$

This proves the first statement. The proof of Eq. (20) follows from Theorem 0. \square

Theorem 10 *Let $m, n, k \in \mathbb{N}$, $k > m/2$, $\Omega = [-1, 1]^m \subseteq \mathbb{R}^m$, $\mathbb{T}^m = \mathbb{R}^m / 2\mathbb{Z}^m$, and $\tilde{f} \in H^k(\mathbb{T}^m, \mathbb{R})$ with periodic representative f . Consider an interpolation operator $S_{m,n} : H^k(\Omega, \mathbb{R}) \rightarrow \Pi_{m,n}$ with respect to multidimensional Newton-Chebyshev nodes $P_{m,n}$. Then*

$$\|f - S_{m,n}(f)\|_{C^0(\Omega)} \xrightarrow{n \rightarrow \infty} 0.$$

Proof The proof is based on balancing the approximation of f in Fourier basis and by polynomials. To do so, we let $n \in \mathbb{N}$, $D, K_1, K_0 \in \mathbb{R}^+$, $D \geq 1$, $K_1 > K_0 \geq 2$, and $l = l(n) \in \mathbb{N}$ such that

$$\frac{(Dl)^{K_1}}{n} \xrightarrow[n \rightarrow \infty]{} 0 \quad \text{and} \quad \frac{\log(n)^{K_0}}{l} \xrightarrow[n \rightarrow \infty]{} 0, \quad (22)$$

i.e., $(Dl)^{K_1} \in o(n)$ and $\log(n)^{K_0} \in o(l)$. We consider the projections $\tau_n, \tau_n^\perp, \theta_l, \theta_l^\perp$ from Eq. (18) and use the fact that $S_{m,n}$ is a linear operator in order to bound

$$\begin{aligned} \|f - S_{m,n}(f)\|_{C^0(\Omega)} &\leq \|\theta_l(f) - S_{m,n}(\theta_l(f))\|_{C^0(\Omega)} \\ &\quad + \|\theta_l^\perp(f) - S_{m,n}(\theta_l^\perp(f))\|_{C^0(\Omega)}. \end{aligned}$$

We then use Theorem 9 and Lemma 2(i) to observe that there exists a constant $c \in \mathbb{R}^+$ so that the second term on the right-hand side satisfies

$$\begin{aligned} \|\theta_l^\perp(f) - S_{m,n}(\theta_l^\perp(f))\|_{C^0(\Omega)} &\leq \|\theta_l^\perp(f) - S_{m,n}(\theta_l^\perp(f))\|_{H^k(\Omega)} \\ &\leq (1 + A(S_{m,n}, H^k(\Omega))) \|\theta_l^\perp(f)\|_{H^k(\Omega)} \\ &\leq c(1 + \log(n)^m) \|\theta_l^\perp(f)\|_{C^0(\Omega)}. \end{aligned}$$

By Lemma 2(ii), we find that $\|\theta_l^\perp(f)\|_{C^0(\Omega)} \in o((m/l)^k)$ for $k > m/2$. Since $K_0 \geq 2$, we obtain

$$\|\theta_l^\perp(f) - S_{m,n}(\theta_l^\perp(f))\|_{C^0(\Omega)} \leq \frac{1 + \log(n)^m}{l^{m/2}} \|f\|_{C^0(\Omega)} \xrightarrow[n \rightarrow \infty]{} 0.$$

Similarly, we bound the remaining term by

$$\begin{aligned} \|\theta_l(f) - S_{m,n}(\theta_l(f))\|_{C^0(\Omega)} &\leq \|\pi_n \theta_l(f) - S_{m,n}(\pi_n \theta_l(f))\|_{C^0(\Omega)} \\ &\quad + \|\pi_n^\perp \theta_l(f) - S_{m,n}(\pi_n^\perp \theta_l(f))\|_{C^0(\Omega)}. \end{aligned}$$

Since $\pi_n \theta_l(f) \in \Pi_{m,n}$, we have $S_{m,n}(\pi_n \theta_l(f)) = \pi_n \theta_l(f)$, implying that the first term vanishes. Now we again use Theorem 9 and Lemma 2(i) and 2(iii) to observe that there exists a constant $c \in \mathbb{R}^+$ so that

$$\begin{aligned} \|\pi_n^\perp \theta_l(f) - S_{m,n}(\pi_n^\perp \theta_l(f))\|_{C^0(\Omega)} &\leq \|\pi_n^\perp \theta_l(f) - S_{m,n}(\pi_n^\perp \theta_l(f))\|_{H^k(\Omega)} \\ &\leq c(1 + \log(n)^m) \|\pi_n^\perp \theta_l(f)\|_{H^k(\Omega)} \\ &\leq c(1 + \log(n)^m) N(m, l)^{1/2} \frac{(\pi l^2)^{n+1}}{(n+1)!} \|f\|_{C^0(\Omega)}. \end{aligned}$$

Since $N(m, l) = \frac{(m+l)!}{n!l!} = \frac{(m+1)\cdots(m+l)}{l!} \in \mathcal{O}(l^m)$ and because $n! > (n/2)^{n/2}$ for all $n \in \mathbb{N}$, there exists a constant $d \in \mathbb{R}^+$ so that

$$c(1 + \log(n)^m) N(m, l)^{1/2} \frac{(\pi l^2)^{n+1}}{(n+1)!} \leq d(1 + l^m) l^{m/2} \left(\frac{(\pi l^2)}{n+1} \right)^{n+1}.$$

Choosing $D, K_1 \in \mathbb{R}^+$ large enough we can further use Eq. (22) to bound

$$\begin{aligned} \|\pi_n^\perp \theta_l(f) - S_{m,n}(\pi_n^\perp \theta_l(f))\|_{C^0(\Omega)} &\leq d(1+l^m)l^{m/2} \frac{(\pi l^2)^{n+1}}{n^{(n+1)/2}} \|f\|_{C^0(\Omega)} \\ &\leq \left(\frac{(Dl)^{K_1}}{n+1}\right)^{(n+1)/2} \|f\|_{C^0(\Omega)} \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

Hence, all terms converge to zero as $n \rightarrow \infty$, proving the theorem. \square

7.4 Approximation Errors

We generalize the classic estimates of approximation errors in 1D to arbitrary dimensions $m \in \mathbb{N}$.

Theorem 11 *Let $m, n \in \mathbb{N}$ and $\Omega = [-1, 1]^m \subseteq \mathbb{R}^m$. Let $S_{m,n} : H^k(\Omega, \mathbb{R}) \rightarrow \Pi_{m,n}$, $k > m/2$, denote an interpolation operator with respect to canonical multidimensional Newton nodes $P_{m,n}$ generated by $\mathbb{P}_{m,n} = \bigoplus_{i=1}^m P_i$, $P_i = \{p_{i,1}, \dots, p_{i,n+1}\}$.*

i) If $f \in C^{n+1}(\Omega, \mathbb{R})$ then for every $\alpha \in A_{m,n} \setminus A_{m,n-1}$, $i \in \{1, \dots, m\}$ and every $x \in \Omega$ there is $\xi_x \in \Omega$ such that

$$|f(x) - S_{m,n}(f)(x)| \leq \frac{1}{\alpha_i!} \partial_{x_i}^{\alpha_i+1} f(\xi_x) |N_\alpha(x)|, \quad (23)$$

where $N_\alpha(x) = \prod_{i=1}^m \prod_{j=1}^{\alpha_i} (x_i - p_{i,j})$, $x = (x_1, \dots, x_m)$, $p_{i,j} \in P_i$. If $P_{m,n}$ are multidimensional Newton-Chebyshev nodes, then we can further bound

$$|f(x) - S_{m,n}(f)(x)| \leq \frac{1}{2^{\alpha_i} \alpha_i!} \partial_{x_i}^{\alpha_i+1} f(\xi_x). \quad (24)$$

ii) For any unisolvent node set $P_{m,n}$ the relative interpolation error is

$$\|f - S_{m,n}(f)\|_{C^0(\Omega)} \leq (1 + \Lambda(P_{m,n}, H^k(\Omega))) \|f - Q_{m,n}^*\|_{C^0(\Omega)}$$

for all $f \in H^k(\Omega, \mathbb{R})$, where $Q_{m,n}^$ is an optimal approximation that minimizes the C^0 -distance to f .*

Proof Changing coordinates if necessary, we can assume w.l.o.g. that $P_{m,n}$ are canonical nodes. To show (i), we follow the argumentation of the classical proof in 1D [19]. We consider the line

$$L_{\alpha,i} = \{x \in \mathbb{R}^m : x_j = p_{j,1+\alpha_j} \text{ for } 1 \leq j < i, \quad x_j = p_{j,1+\alpha_j} \text{ for } i < j \leq m\}$$

and choose $\bar{x} \in (\Omega \cap L_{\alpha,i}) \setminus P_{m,n}$. The function $g_{\alpha,i} : \mathbb{R}^m \rightarrow \mathbb{R}$ given by

$$g_{\alpha,i}(x) = f(x) - Q_{m,n,f}(x) - G(\bar{x})N_\alpha(x), \quad G(\bar{x}) = (f(\bar{x}) - Q_{m,n,f}(\bar{x})) / N_\alpha(\bar{x})$$

is of class C^{n+1} and possesses $\alpha_i + 1$ roots, namely $\{p_{i,1}, \dots, p_{i,\alpha_i}, \bar{x}\}$, on $L_{\alpha,i}$. Recursively applying Rolle's Theorem, this implies that $\partial_{x_i}^k g_{\alpha,i}$ possesses $\alpha_i + 1 - k$ roots on $L_{\alpha,i}$, $0 \leq k \leq \alpha_i$. Hence

$$\partial_{x_i}^{\alpha_i+1} g_{\alpha}(\xi) = 0, \quad \text{for some } \xi = \xi_{\bar{x}} \in L_{\alpha,i} \cap \Omega.$$

Theorems 4 and 5 yield Eq. (21), which implies that the restriction of $Q_{m,n,f}$ to $L_{\alpha,i}$ is of degree α_i . Thus, $\partial_{x_i}^{\alpha_i+1} Q_{m,n,f}|_{L_{\alpha,i}} = 0$. Since $\partial_{x_i}^{\alpha_i} N_{\alpha}(x) = \alpha_i!$, this yields

$$G(\bar{x}) = \frac{1}{\alpha_i!} \partial_{x_i}^{\alpha_i+1} f(\xi),$$

implying Eq. (23). Combining Lemma 3 with the classic error estimation for Chebyshev nodes in 1D [19] yields Eq. (24). To show (ii), we recall that $S_{m,n}$ is a projection, i.e., $S_{m,n}(S_{m,n}) = S_{m,n}$ implying that $S_{m,n}(Q) = Q$ for any $Q \in \Pi_{m,n}$. Thus, for every $f \in H^k(\Omega, \mathbb{R})$, we bound

$$\begin{aligned} \|f - S_{m,n}(f)\|_{C^0(\Omega)} &= \|f - Q_{m,n}^* + Q_{m,n}^* - S_{m,n}(f)\|_{C^0(\Omega)} \\ &\leq \|f - Q_{m,n}^*\|_{C^0(\Omega)} + \|Q_{m,n}^* - S_{m,n}(f)\|_{C^0(\Omega)} \\ &= \|f - Q_{m,n}^*\|_{C^0(\Omega)} + \|S_{m,n}(Q_{m,n}^*) - S_{m,n}(f)\|_{C^0(\Omega)} \\ &\leq (1 + A(S_{m,n}, H^k(\Omega))) \|f - Q_{m,n}^*\|_{C^0(\Omega)}. \end{aligned}$$

□

In summary, we have established all approximation results of Main Result III as stated in Theorem 3 in Section 2, thereby extending the well-known classic results from 1D to arbitrary dimensions. This answers Question 1 set out in the problem statement.

8 Numerical Experiments

In order to illustrate our approach and demonstrate its performance in practice, we implement a prototype MATLAB (R2018a (9.4.0.813 654) version) of the PIP-SOLVER running on an Apple MacBook Pro (Retina, 15-inch, Mid 2015) with a 2.2 GHz Intel Core i7 processor and 16 GB 1600 MHz DDR3 memory under macOS Sierra (version 10.12.6). The following numerical experiments demonstrate the computational performance and approximation accuracy of our solver in comparison with the classic numerical approaches.

For given $m, n \in \mathbb{N}$ and function $f : \mathbb{R}^m \rightarrow \mathbb{R}$, we compare the following methods in terms of accuracy and runtime:

- i) The *PIP-SOLVER* generates multidimensional Newton-Chebyshev nodes and determines the coefficients $C \in \mathbb{R}^{N(m,n)}$ of the interpolant $Q_{m,n,f}$ in multivariate Newton form.

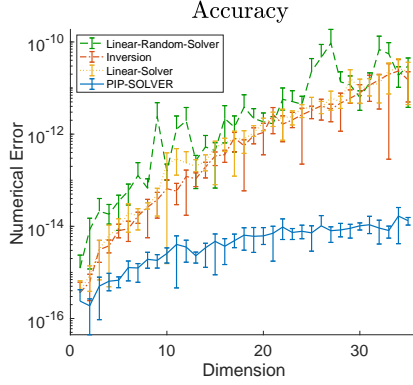


Fig. 3 Numerical error for degree $n = 3$.

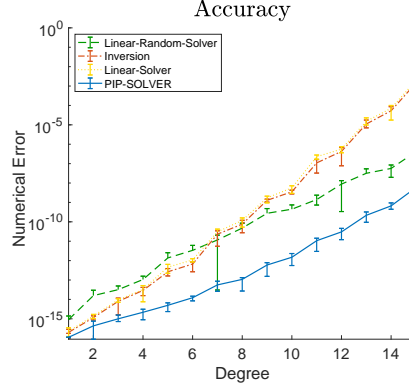


Fig. 4 Numerical error for dimension $m = 5$.

- ii) The *Linear Solver* uses the multidimensional Newton-Chebyshev nodes $P_{m,n}$ generated by the PIP-SOLVER and then applies the MATLAB linear system solver to solve

$$V_{m,n}(P_{m,n})C = F, \quad F = (f(p_1), \dots, f(p_N(m,n)))^T \in \mathbb{R}^{N(m,n)}$$

for the coefficients $C \in \mathbb{R}^{N(m,n)}$ of the interpolant $Q_{m,n,f}$ in normal form.

- iii) The *Linear Random Solver* uses nodes $P_{m,n}$ placed uniformly at random and then applies the MATLAB linear system solver to solve

$$V_{m,n}(P_{m,n})C = F, \quad F = (f(p_1), \dots, f(p_N(m,n)))^T \in \mathbb{R}^{N(m,n)}$$

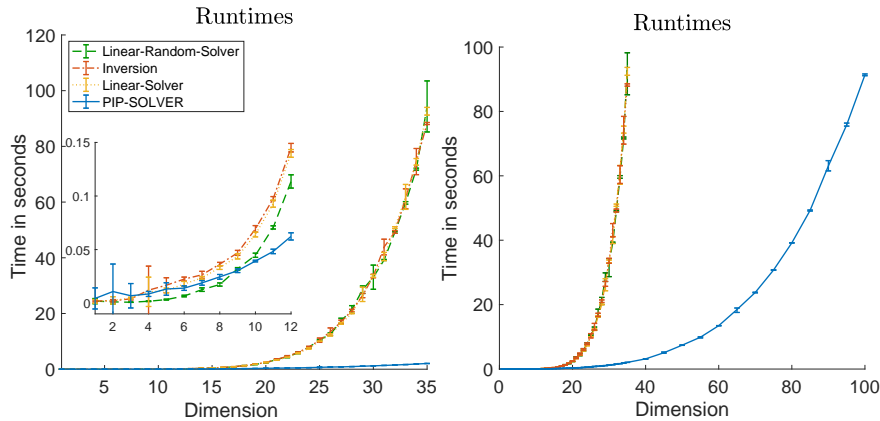
for the coefficients $C \in \mathbb{R}^{N(m,n)}$ of the interpolant $Q_{m,n,f}$ in normal form.

- iv) The *Inversion* method uses the multidimensional Newton-Chebyshev nodes $P_{m,n}$ generated by the PIP-SOLVER and then inverts the Vandermonde matrix $V_{m,n}(P_{m,n})$ using LU-decomposition to compute the coefficients $C \in \mathbb{R}^{N(m,n)}$ of the interpolant $Q_{m,n,f}$ in normal form.

Experiment 1 We first compare the accuracy of the three approaches, which also serves to validate our method. To do so, we choose uniformly-distributed random numbers $c_1, \dots, c_{N-1} \in [-1, 1]^{N(m,n)}$ to be the coefficients of a polynomial $Q \in \Pi_{m,n}$ in normal or multivariate Newton form. We then set $f = Q$ and measure the maximum absolute error in any coefficient, i.e. $\|c - \tilde{c}\|_\infty$, when recovering \tilde{Q} by solving the PIP with respect to (m, n, f) .

Figures 3 and 4 show the average and min-max span of the numerical errors (over 5 repetitions with different *i.i.d.* random coefficients; same 5 for every approach) for fixed degree $n = 3$ and dimensions $m = 2, \dots, 35$, as well as for fixed dimension $m = 5$ and degree $n = 1, \dots, 15$, with logarithmic scale in the y -axis.

The case $n = 3$ is of high practical relevance, e.g., when interpolating cubic splines. In both cases, all methods show high accuracy, which reflects the fact that Newton-Chebyshev nodes yield well-conditioned PIPs. This is

Fig. 5 Runtimes for $n = 3$, $1 \leq m \leq 35$.Fig. 6 Runtimes for $n = 3$, $1 \leq m \leq 100$.

Algorithm	Intervals	Degree	Pre-factor p	Exponent q
<i>Inversion</i>	$m = 15, \dots, 35$	$n = 3$	$p = 0.010737$	$q = 2.2982$
<i>Linear Solver</i>	$m = 15, \dots, 35$	$n = 3$	$p = 0.0076072$	$q = 2.2907$
<i>Linear Random Solver</i>	$m = 15, \dots, 35$	$n = 3$	$p = 0.012009$	$q = 2.3289$
<i>PIP-SOLVER</i>	$m = 15, \dots, 35$	$n = 3$	$p = 0.0076964$	$q = 1.2006$
<i>PIP-SOLVER</i>	$m = 15, \dots, 100$	$n = 3$	$p = 0.0034101$	$q = 1.2258$

Table 1 Scaling of the computational cost by fitting the cost model $pN(m, n)^q$.

confirmed by the *Linear Solver* and the *Inversion* method showing comparable accuracy, while the *Linear Random Solver* is less accurate. The error of the PIP-SOLVER is almost constant on the level of the machine accuracy (double-precision floating-point number types). Especially in high dimensions, the PIP-SOLVER is several orders of magnitude more accurate than the other approaches.

Experiment 2 We compare the computational runtimes of the approaches. To do so, we choose uniformly-distributed random function values $f_1, \dots, f_{N(m,n)} \in [-1, 1]$, $m, n \in \mathbb{N}$ as interpolation targets. Then, we measure the time required to generate the unisolvent interpolation nodes $P_{m,n}$ and add the time taken to solve the PIP with respect to $f : \mathbb{R}^m \rightarrow \mathbb{R}$ with $f(p_i) = f_i$, $p_i \in P_{m,n}$, $i = 1, \dots, N(m, n)$ by each approach.

The average and min-max span of the runtimes (over 10 repetitions with different *i.i.d.* random function values; same 10 for every approach) are shown in Figures 5 and 6 versus the dimension m for fixed degree $n = 3$.

While the actual problem size is $N(m, n)$, the dimension m or the degree n are more intuitive when characterizing a problem of fixed degree or fixed dimension, respectively. In low dimensions (inset figure) the *Linear Random Solver* performs best due to its low overhead for generating unisolvent nodes. However, at about $m = 9$ there is a cross-over above which the PIP-

SOLVER is much faster than the other methods. The absolute runtimes are below 0.05 seconds at the cross-over point, even though our prototype implementation of the PIP-SOLVER is not optimized. As Figure 6 shows, even our simple implementation of the PIP-SOLVER can handle instances of dimension $m = 100$ in the same time as the other methods require for $m = 35$. Since $N(100, 3)/N(35, 3) \approx 20$ the PIP-SOLVER outperforms the other approaches.

The scaling of the computational cost with respect to problem size $N(m, n)$ is reported in Table 8 for $n = 3$. We fit all measurements with the cost model $pN(m, n)^q$. All fits show an *R-square* of 1. We observe that the exponent of the cost scaling of the PIP-SOLVER is more than 1 less than the exponents of the other methods with pre-factors that are never larger. The quadratic upper bound we have proven in this paper for the PIP-SOLVER holds in all tested cases. The other approaches roughly scale with an exponent of 2.3, as expected.

In addition to having a lower time complexity, the PIP-SOLVER also requires less memory than the other approaches. Indeed, the PIP-SOLVER requires only $\mathcal{O}(N(m, n))$ storage, whereas all other approaches require $\mathcal{O}(N(m, n)^2)$ storage to hold the Vandermonde matrix. Due to this lower space complexity, we could solve the PIP for large instances, i.e., for $m > 80$, where $N(m, 3) \geq 10^5$ in less than 2 minutes, see Figure 6, while classical approaches failed to solve such large instances due to insufficient memory on the computer used for the experiment.

Experiment 3 We measure the runtime of the PIP-SOLVER for different polynomial degrees n . Again, we choose uniformly-distributed random function values $f_1, \dots, f_{N(m, n)} \in [-1, 1]$, $m, n \in \mathbb{N}$, as interpolation targets. Then, we measure the time required to generate the multidimensional Newton-Chebyshev nodes $P_{m, n}$ and add the time taken to solve the PIP with respect to $f : \mathbb{R}^m \rightarrow \mathbb{R}$ with $f(p_i) = f_i$, $p_i \in P_{m, n}$, $i = 1, \dots, N(m, n)$ for different n .

Figure 7 shows the average and min-max span of the runtimes (over 10 repetitions with different *i.i.d.* random function values) versus the dimension $m \in \mathbb{N}$ with logarithmic scale on the y -axis. We again fit the curves in the admissible intervals with the cost model $pN(m, n)^q$. Again, the goodness of fit as measured by the *R-square* is 1 in all cases. As expected, the exponent does not change much with degree. Just as in 1D, the almost linear scaling for low degrees reflects the computational power of the multivariate divided difference scheme.

Next, we test the approximation properties of the PIP-SOLVER. A classic test case in approximation theory is *Runge's function*

$$f_R(x) = \frac{1}{1 + 25x^2}.$$

One can easily verify that $\left| \frac{d^k f_R}{dx^k}(1/5) \right| \rightarrow \infty$, for $k \rightarrow \infty$. Thus, f_R is smooth but unbounded with respect to $\|\cdot\|_{C^\infty(\Omega)}$, which means that $f_R \notin (C^\infty(\Omega), \|\cdot\|_{C^\infty(\Omega)})$. This is the reason, why f_R can not be approximated

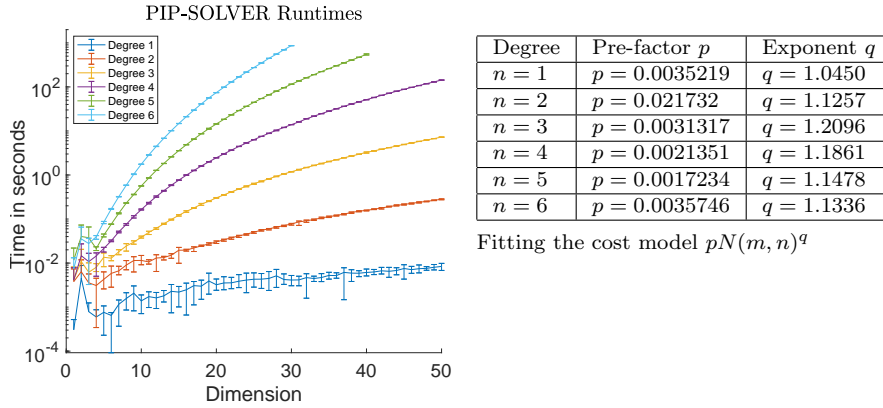


Fig. 7 Runtimes of the PIP-SOLVER for degrees $n = 1, \dots, 6$.

by interpolation with *equidistant nodes* [34]. However, as long as k remains bounded, we have $\|f_R\|_{C^k(\Omega)} \leq C_K$ for some $C_K \in \mathbb{R}^+$. Thus, for all $m \geq 1$, one verifies that the multivariate analog $f_R : \mathbb{R}^m \rightarrow \mathbb{R}$ with $f_R(x) = \frac{1}{1+25\|x\|^2}$ satisfies $f_R \in H^k(\Omega, \mathbb{R})$ for $k > m/2$. By Theorem 9, all Sobolev functions can be approximated when using Newton-Chebyshev nodes. Therefore, we consider the multidimensional f_R when testing the approximation abilities of the PIP-SOLVER.

Experiment 4 We consider $m = 5$, $\Omega = [-1, 1]^m$ and use the PIP-SOLVER to compute the interpolant $Q_{m,n,f_R,CN}$ with respect to multidimensional Newton-Chebyshev nodes and the interpolant $Q_{m,n,f_R,EN}$ with respect to multidimensional Newton nodes $P_{m,n} = \bigoplus_{i=1}^m \mathbb{P}_i$ with $\#\mathbb{P}_i = n + 1$ equidistant on $[-1, 1]$. For technical reasons we consider only even degrees $n \in 2\mathbb{N}$ allowing to choose 0 as the center of the Chebyshev nodes. To estimate the C^0 distance between the interpolants, we generate 400 uniformly random points $P \subseteq \Omega$ once and measure the relative approximation error $|Q_{m,n,f_R,CN}(p) - f_R(p)|/f_R(p)$, $|Q_{m,n,f_R,EN}(p) - f_R(p)|/f_R(p)$ for each $p \in P$ and $n = 2, 4, \dots, 24$.

Figure 8 plots the maximum and the mean of the relative distances $|Q_{m,n,f_R,CN}(p) - f_R(p)|/f_R(p)$, $|Q_{m,n,f_R,EN}(p) - f_R(p)|/f_R(p)$ over the 400 randomly chosen but fixed points, with logarithmic scale in the y -axis. Though the interpolant $Q_{m,n,f_R,EN}$ with equidistant nodes approximates f_R for low degrees, it diverges with increasing degree $n \geq 6$. In contrast, the interpolant $Q_{m,n,f_R,CN}$ continues to converge to f_R uniformly on P . Thus, we confirm that also in higher dimensions equidistant Newton nodes are infeasible for approximating Runge's function, while Newton-Chebyshev nodes result in uniform convergence.

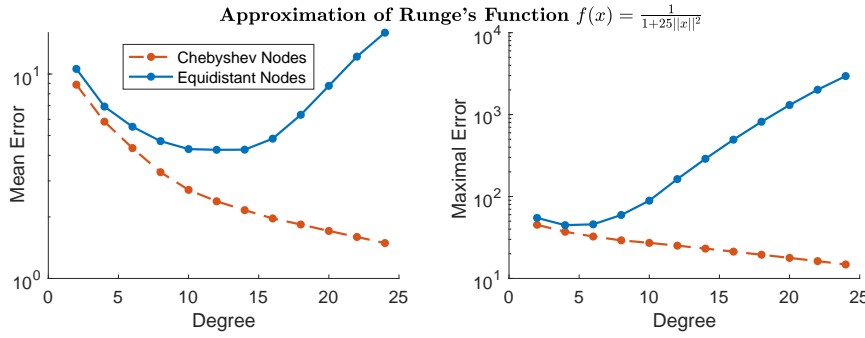


Fig. 8 Relative approximation error of Runge's function in fixed dimension $m = 5$ with degrees $n = 2, 4, \dots, 24$.

9 Potential Applications

We highlight several potential applications of the PIP-SOLVER in scientific computing and computational science. This list is by no means exhaustive, as PIPs are a fundamental component of many numerical methods. However, the following applications may not be obvious:

A1) Basic numerics: Given $m, n \in \mathbb{N}$, $m \geq 1$, and a function $f : \mathbb{R}^m \rightarrow \mathbb{R}$. It is classical in numerical analysis to determine the integral $\int_{\Omega} f \, d\Omega$, $\Omega \subseteq \mathbb{R}^m$, and the partial derivatives $\partial_{x_i} f(x)$, $1 \leq i \leq m$ of f . Upon the PIP, these desired quantities can easily be computed for the interpolation polynomial $Q_{m,n,f}$ of f . Due to our Main Result III (Theorem 3), $Q_{m,n,f} \xrightarrow{n \rightarrow \infty} f$ uniformly. Thus, $\int_{\Omega} Q_{m,n,f} \, d\Omega \xrightarrow{n \rightarrow \infty} \int_{\Omega} f \, d\Omega$ and by strengthen the conditions on f also $\partial_{x_i} Q_{m,n,f}(x) \xrightarrow{n \rightarrow \infty} \partial_{x_i} f(x)$ uniformly. A comparison with other approaches from numerical analysis is worth considering.

A2) Gradient descent over multivariate functions is often used to (locally) solve (non-convex) optimization problems, where $\Omega \subseteq \mathbb{R}^m$ models the space of possible solutions for a given problem and $f : \mathbb{R}^m \rightarrow \mathbb{R}$ is interpreted as an objective function. Thus, one wants to minimize f on Ω . Often, the function f is not explicitly known $\forall x \in \Omega$, but can be evaluated point-wise. Due to the Infeasibility of previous interpolation methods for $m \gg 1$, direct interpolation of f on Ω was often not possible or not considered. Therefore, classical numerical methods like the *Newton-Raphson iteration* could not be used if analytical gradients were not available. Instead, *discrete or stochastic gradient descent* methods were usually applied. However, these methods converge slowly and are potentially inaccurate. The PIP-SOLVER allows (locally) interpolating f even for $m \gg 1$ and enables (locally) applying classic *Newton-Raphson methods*. Consequently, local minima could be found faster and more accurately. Moreover, the analytical representation of Q potentially allows determining

the global optimum if f can be uniformly approximated [21, 29].

A3) ODE & PDE solvers: A core application of numerical analysis is the approximation of the solution of Ordinary Differential Equations (ODE) or Partial Differential Equations (PDE). This always involves a (temporal and/or spatial) discretization scheme and a solver for the resulting equations. There are three classes of methods: collocation schemes, Galerkin schemes, and spectral methods. Spectral methods are based on Fourier transforms. Since FFTs are only efficient on regular Cartesian grids, spectral methods are hard to apply in complex geometries and on adaptive-resolution discretizations. The present PIP-SOLVER, however, is not limited to polynomial bases and could enable spectral methods on arbitrary distributions of discretization points in arbitrary geometries by interpolating with respect to a Fourier basis [22]. Collocation methods can generally be understood as PIPs, as becomes obvious in the generalized formulation of finite-difference schemes and mesh-free collocation methods [37]. The inversion of the Vandermonde matrix implied in mesh-free methods, or the choice of mesh nodes in compact finite-difference schemes, could benefit from the algorithms presented here. Finally, Galerkin schemes are based on expanding the solution of the differential equation in some basis functions, which is essentially what the PIP does for the orthogonal basis of monomials. Since the theory presented here is not restricted to this particular choice of basis, it is conceivable that similar algorithms can be formulated for other bases as well, potentially even for non-orthogonal ones.

A4) Adaptive sampling methods aim to explore a domain $\Omega \subseteq \mathbb{R}^m$ such that the essential information of $f : \Omega \rightarrow \mathbb{R}$ is recovered. A classic example from statistics is multidimensional *Bayesian inference* [5], which relies on adaptive sampling methods. Mostly, these methods are based on *Markov-Chain Monte-Carlo* or *sequential Monte-Carlo* sampling. The notion of unisolvent node sets in high dimensions, as we have provided here, potentially helps design sampling proposals that explore Ω in a more controlled, complete, or more efficient way.

A5) Spectral analysis: The foundation of spectral analysis is to represent a function $f : \Omega \subseteq \mathbb{R}^m \rightarrow \mathbb{R}, \mathbb{C}$ with respect to some functional basis $f(x) = \sum_{i=1}^{\infty} c_i b_i(x)$, $c_i \in \mathbb{R}, \mathbb{C}$, $\text{span}\{b_i\}_{i \in \mathbb{N}} = L^2(\Omega, \mathbb{R})$, e.g., with respect to Zernike or Chebyshev polynomials, linear (Fourier) harmonics, spherical harmonics, etc. This allows analyzing and understanding the essential character $f_0 = \sum_{i \in I} c_i b_i(x)$ of f , where the finite set I is chosen such that the c_i with $i \in I$ cover the most relevant amplitudes. Interpolating f with respect to the specified basis b_i is the classic method of computing the coefficients c_i . In [22] we have already described how to extend the PIP-SOLVER to Fourier basis. An adaption to other functional bases can be done analogously, which potentially improves numerical spectral analysis.

A6) Cryptography: A maybe surprising application is found in cryptography. There, the PIP is used to “share a secret” by choosing a random polynomial $Q \in \mathbb{Z}[x]$ with integer coefficients in dimension $m = 1$. Knowing the values of Q at $n+1$ different nodes (keys) enables one to determine $Q(0) \bmod p$ for some large prime number $p \in \mathbb{N}$. However, knowing only n “keys” (nodes) prevents one from opening the “door”. Certainly, this method can be generalized to arbitrary dimensions using our approach [38]. Since the PIP-SOLVER performs with machine accuracy, it can also prevent the reconstructed message from being corrupted by numerical noise.

There are many other computational schemes that require interpolation or are closely related to the PIP *linear or polynomial regression* in machine learning. We therefore close with a qualitative discussion of the PIP-SOLVER and state open questions, which potentially yield generalizations and further improvements in the future.

10 Discussion and Conclusions

Even though *Newton interpolation* in one dimension has been known since the 18th century, this may be the first generalization of this fundamental algorithm to arbitrary dimensions. We have provided a complete characterization of the polynomial interpolation problem (PIP) in arbitrary dimensions and polynomial degrees. We have provided an algorithm called PIP-SOLVER (see Algorithm 1) that computes the solution to generalized PIPs in $\mathcal{O}(N(m, n)^2)$ time and $\mathcal{O}(N(m, n))$ space, where $N(m, n)$ is the number of unknown coefficients of the interpolation polynomial with m variables of degree n . We have shown that the algorithm generates unisolvent node sets for which the Vandermonde matrix is not only well conditioned, but has triangular form, enabling efficient and accurate numerical solution using a multivariate divided difference scheme also presented here. We have further provided the corresponding extensions of the Horner scheme, enabling evaluating the polynomial in $\mathcal{O}(N(m, n))$ and its integral and derivatives in $\mathcal{O}(nN(m, n))$ time. Lastly, we have studied the approximation properties of multivariate Newton interpolation polynomials and derived the notion of multidimensional Newton-Chebyshev nodes, showing that any Sobolev function of sufficient regularity can be uniformly approximated, and we have provided the corresponding upper bounds on the approximation errors. Taken together, these contributions solve Problem 1 and answer Question 1 for arbitrary dimensions and degrees.

The problem statement and questions considered here are not new. Nor is the idea of decomposing the PIP w.r.t. $m, n \in \mathbb{N}$ into sub-problems of dimension and degree $(m - 1, n)$ and $(m, n - 1)$, respectively, which has already been mentioned in [20, 26]. In [17], the cases $m = 2, 3$ were treated explicitly, while a generalization to arbitrary dimension was sketched and some characterizations of unisolvent nodes in arbitrary dimensions were given. However, the problem of computing the interpolation polynomial $Q_{m, n, f}$ efficiently and

Algorithm 1 PIP-SOLVER

```

1: procedure PIP-SOLVER( $f, m, n$ ) ▷  $f$  is computable
2:   Choose admissible plane  $H$  according to Theorem 4
3:   if  $m = 0$  then
4:     return  $f|_H$  ▷  $H$  is a point
5:   end if
6:   if  $n = 0$  then
7:     return  $f(0)$  ▷  $Q_{m,n,f}$  is a constant
8:   end if
9:   Choose  $Q_H \in \Pi_{m,1}$  with  $Q_H(H) = 0$ 
10:   $Q_1 = \text{PIP-SOLVER}(f|_H, m - 1, n)$  ▷ recursion over  $m$ 
11:   $\hat{f} = (f - Q_1)/Q_H$  on  $\mathbb{R}^m \setminus H$ 
12:   $Q_{m,n,f} = Q_1 + Q_H \cdot \text{PIP-SOLVER}(\hat{f}, m, n - 1)$  ▷ recursion over  $n$ 
13:  return  $Q_{m,n,f}$ 
14: end procedure

```

accurately remained unsolved. Indeed, all previous decomposition approaches were limited to relatively low dimensions and to nodes on pre-defined grids or meshes [4, 14, 16, 18, 25], not providing a general algorithm for solving the PIP for arbitrary $m, n \in \mathbb{N}$.

Here, we were able to provide such a general algorithm and prove bounds on its time and space complexity. The achievement that made this possible was the introduction of unisolvent multidimensional Newton nodes $P_{m,n}$, which, combined with the multivariate Newton basis of $\Pi_{m,n}$, yielded a form of the Vandermonde matrix $V_{m,n}(P_{m,n})$ that allows solving the system of linear equations $V_{m,n}(P_{m,n})C = F$ in less time than what is required for general matrix inversion.

We demonstrated and validated our results in a practical software implementation of the PIP-SOLVER, showing scaling to high-dimensional spaces as common in applications ranging from machine learning to computational statistics. Owing to the elegance of the theory, the implementation of the PIP-SOLVER is straightforward and results in a simple code.

Our simple reference implementation is not tuned for efficiency at the time of writing. In the future, we foresee distributed- and shared-memory parallel implementations in compiled programming languages to further reduce run-times and enable even larger problems to be solved. This is possible since the recursive decomposition yields sub-problems that can be processed in parallel, using inter-process communication to ensure correct decomposition of the problem and synthesis of the final solution.

A possible extension of the presented theory is to also cover interpolation in other bases, such as the Fourier basis [22], spherical harmonics, or Zernike polynomials. We also believe that our approach can be extended to multivariate barycentric or Lagrange Interpolation. In 1D, it is well known that Newton and Lagrange polynomials are related through the *barycentric weights* [3, 50]. Precomputing these weights, barycentric Lagrange interpolation only requires linear time $\mathcal{O}(N(1, n))$ to compute the interpolant of degree $n \in \mathbb{N}$ in 1D. The present approach to multivariate Newton polynomials could lead to multivari-

ate barycentric Lagrange interpolation schemes running in $\mathcal{O}(mN(m, n))$ for arbitrary $m, n \in \mathbb{N}$. Similarly, multivariate Hermite Interpolation could also be considered, since efficient realizations of this concept are closely related to Newton and Lagrange interpolation. In 1D, Hermite interpolation is a classic concept [19] that requires one to know the function $f : [-1, 1] \rightarrow \mathbb{R}$ and its derivatives on less than $n + 1$ nodes in order to compute the interpolant $Q_{f,n} \in \Pi_{1,n}$.

A case of special interest is spline interpolation [46]. Fast implementations of spline interpolation are available [23,47], making them a powerful and popular tool. Spline interpolation is based on decomposing the domain $\Omega = [-1, 1]^m$ into smaller, shifted hypercubes $\Omega_i = [-\varepsilon, \varepsilon]^m + p_i$, $p_i \in \Omega$, $i \in I$, $\varepsilon > 0$, and “gluing” the interpolants $Q_i : \Omega_i \rightarrow \mathbb{R}$ to a k -times differentiable global function Q , $k \in \mathbb{N}$. Therefore, $\#I$ increases exponentially with dimension m . Tensorial formulations [11,30] are available for efficient local spline interpolation. However, the exponential scaling of the number of hypercubes, $\#I$ in which this has to be done cannot be overcome. This is why spline interpolation is mostly used for lower-dimensional problems. Further, the mathematical character of f is not recovered in the spline basis and the approximation quality depends on the spline degree and on the choice of node conditions [36,44,48]. Therefore, spline interpolation is well suited to signal and image processing in low dimensions. The L^2 -bases we proposed in A5) above provide a potentially interesting choice in high dimensions. In principle, Hermite interpolation could also be used to glue spatially decomposed interpolants to a global k -times differentiable function. The notion of a globally unisolvent node set $P_{m,n}$ could then provide a way of spatially decomposing Ω such that the resulting global Hermite interpolant is of high approximation quality and can be computed efficient. We expect that this hybrid Hermite-spline interpolation method would relax some of the issues with splines in high dimensions.

The main practical limitation of our approach is that it requires the function $f : \Omega \rightarrow \mathbb{R}$ to be computable in constant time, which means that the algorithm is free to choose the interpolation nodes. In many problems, however, f is only known on a previously fixed node set $\mathbb{P} \subseteq \mathbb{R}^m$ (i.e., the data given). In the case where \mathbb{P} is a (regular) grid, we can choose multivariate Newton nodes $P_{m,n} \subseteq \mathbb{P}$ and our approach works. However, if \mathbb{P} is arbitrarily scattered, our approach does not directly apply. While resampling/reorganizing the data points can sometimes be an option, a general solution is outstanding. In particular, the optimal ordering of the nodes yielding the best numerical approximation in the sense of minimal rounding errors remains to be investigated. While this does not matter in infinite-precision arithmetic, finite-precision floating-point arithmetic accuracy, as well as algorithm speed, can be improved by appropriately ordering the points [43].

The main theoretical limitation of our approach is that we bounded the Lebesgue functions with respect to the H^k -norm for $k > m/2$. Additionally, we assumed that the considered functions $f \in C^0(\Omega, \mathbb{R})$, $\Omega = [-1, 1]^m$, $m \in \mathbb{N}$ are periodic. While these assumptions match the requirements of many

practical applications, the classic Lebesgue function estimates just require f to be continuous. Hence, a deeper study of Lebesgue functions with respect to the powerful Sobolev analysis of periodic functions might improve the bounds presented here and might provide a way of controlling the convergence rate of $Q_{m,n,f} \rightarrow f$.

Notwithstanding these open questions, we suspect that our concepts could provide a general perspective for considering multivariate interpolation problems, since, for dimension $m = 1$, our concepts include the classic Newton interpolation scheme. We thus hope that the concepts and algorithms presented here will be useful to the community across application domains.

References

1. Adams, R.A., Fournier, J.J.: Sobolev spaces, vol. 140. Academic press (2003)
2. Atkinson, K.E.: An introduction to numerical analysis. John Wiley & Sons (2008)
3. Berrut, J.P., Trefethen, L.N.: Barycentric Lagrange interpolation. *SIAM review* **46**(3), 501–517 (2004)
4. Bos, L., Marchi, S.D., Vianello, M.: Polynomial approximation on Lissajous curves in the d-cube. arXiv:1502.04114 (2015)
5. Box, G., Tiao, G.: Bayesian inference in statistical analysis. Addison-Wesley series in behavioral science: quantitative methods. Addison-Wesley Pub. Co. (1973). URL <https://books.google.de/books?id=oHOpAQAAAJ>
6. Brutman, L.: On the lebesgue function for polynomial interpolation. *SIAM Journal on Numerical Analysis* **15**(4), 694–704 (1978)
7. Burden, R.L., Faires, J.D.: Numerical analysis. PWS, Boston (1993)
8. Cheney, E.W., Light, W.A.: A course in approximation theory, vol. 101. American Mathematical Soc. (2009)
9. Coppersmith, D., Winograd, S.: Matrix multiplication via arithmetic progressions. *Journal of Symbolic Computation* **9**(3), 251 – 280 (1990). DOI [http://dx.doi.org/10.1016/S0747-7171\(08\)80013-2](http://dx.doi.org/10.1016/S0747-7171(08)80013-2). URL <http://www.sciencedirect.com/science/article/pii/S0747717108800132>
10. Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C.: Introduction to algorithms. MIT press (2009)
11. De Boor, C.: Efficient computer manipulation of tensor products. Tech. rep., WISCONSIN UNIV MADISON MATHEMATICS RESEARCH CENTER (1977)
12. De Branges, L.: The Stone-Weierstrass Theorem. *Proceedings of the American Mathematical Society* **10**(5), 822–824 (1959)
13. Endre, S., Mayers, D.: An introduction to numerical analysis. Cambridge, UK (2003)
14. Erb, W., Kaethner, C., Denker, P., Ahlborg, M.: A survey on bivariate Lagrange interpolation on Lissajous nodes. *Dolomites Research Notes on Approximation* **8**, 23–36. URL http://dx.doi.org/10.14658/pupj-drna-2015-Special_Issue-4
15. Faber, G.: Über die interpolatorische Darstellung stetiger Funktionen. *Jber. Deutsch. Math. Verein* **23**, 192–210 (1914)
16. Gall, F.L.: Powers of tensors and fast matrix multiplication. *CoRR* **abs/1401.7714** (2014). URL <http://arxiv.org/abs/1401.7714>
17. Gasca, M., Maeztu, J.I.: On Lagrange and Hermite interpolation in \mathbb{R}^k . *Numerische Mathematik* **39**(1), 1–14 (1982). DOI 10.1007/BF01399308. URL <http://dx.doi.org/10.1007/BF01399308>
18. Gasca, M., Sauer, T.: Polynomial interpolation in several variables. *Advances in Computational Mathematics* **12**(4), 377 (2000). DOI 10.1023/A:1018981505752. URL <http://dx.doi.org/10.1023/A:1018981505752>
19. Gautschi, W.: Numerical analysis. Springer Science & Business Media (2011)
20. Guenther, R.B., Roetman, E.L.: Some observations on interpolation in higher dimensions. *Math. Comp.* **24**(), 517–522 **24**, 517–522 (1970)

21. Hanzon, B., Jibeteau, D.: Global minimization of a multivariate polynomial using matrix methods. *J. of Global Optimization* **27**(1), 1–23 (2003). DOI 10.1023/A:1024664432540. URL <http://dx.doi.org/10.1023/A:1024664432540>
22. Hecht, M., Sbalzarini, I.F.: Fast interpolation and Fourier transform in high-dimensional spaces. In: K. Arai, S. Kapoor, R. Bhatia (eds.) *Intelligent Computing. Proc. 2018 IEEE Computing Conf.*, Vol. 2., *Advances in Intelligent Systems and Computing*, vol. 857, pp. 53–75. Springer Nature, London, UK (2018)
23. Huizinga, W., Klein, S., Poot, D.H.: Fast multidimensional B-spline interpolation using template metaprogramming. In: *International Workshop on Biomedical Image Registration*, pp. 11–20. Springer (2014)
24. Jost, J.: *Partial Differential Equations*. New York: Springer-Verlag (2002)
25. K. C. Chung, T.H.Y.: On lattices admitting unique Lagrange interpolations. *SIAM Journal on Numerical Analysis* **14**(4), 735–743 (1977). URL http://dx.doi.org/10.14658/pupj-drna-2015-Special_Issue-4
26. Le Méhauté, A.: On some aspects of multivariate polynomial interpolation. *Advances in Computational Mathematics* **12**(4), 311–333 (2000). DOI 10.1023/A:1018985606661. URL <http://dx.doi.org/10.1023/A:1018985606661>
27. Meijering, E.: A chronology of interpolation: From ancient astronomy to modern signal and image processing. *Proceedings of the IEEE* **90**(3), 319–342 (2002)
28. Olver, P.J.: On multivariate interpolation. *Studies in Applied Mathematics* **116**(2), 201–240 (2006)
29. Parrilo, P.A., Sturmfels, B.: Minimizing polynomial functions. *Algorithmic and quantitative real algebraic geometry*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science **60**, 83–99 (2003)
30. Pereyra, V., Scherer, G.: Efficient computer manipulation of tensor products with applications to multidimensional approximation. *Mathematics of Computation* **27**(123), 595–605 (1973)
31. Powell, M.J.D.: *Approximation theory and methods*. Cambridge university press (1981)
32. Raz, R.: On the complexity of matrix product. In: *Proceedings of the thirty-fourth annual ACM symposium on Theory of computing*, pp. 144–151. ACM (2002)
33. Robinson, S.: Toward an optimal algorithm for matrix multiplication. *SIAM news* **38**(9), 1–3 (2005)
34. Runge, C.: Über empirische Funktionen und die Interpolation zwischen äquidistanten Ordinaten. *Zeitschrift für Mathematik und Physik* **46**(224–243), 20 (1901)
35. Schatzman, M.: *Numerical Analysis: A Mathematical Introduction*. Clarendon Press, Oxford (2002)
36. Schoenberg, I.: Notes on spline functions III: On the convergence of the interpolating cardinal splines as their degree tends to infinity. *Israel Journal of Mathematics* **16**(1), 87–93 (1973)
37. Schrader, B., Reboux, S., Sbalzarini, I.F.: Discretization correction of general integral PSE operators for particle methods. *J. Comput. Phys.* **229**(11), 4159–4182 (2010). DOI 10.1016/j.jcp.2010.02.004. URL <http://dx.doi.org/10.1016/j.jcp.2010.02.004>
38. Shamir, A.: How to share a secret. *Commun. ACM* **22**(11), 612–613 (1979). DOI 10.1145/359168.359176. URL <http://doi.acm.org/10.1145/359168.359176>
39. Smale, S.: An infinite dimensional version of Sard’s theorem. *Amer. J. Math.* **87**, 861–866 (1965)
40. Stewart, G.: *Afternotes on Numerical Analysis*. Society for Industrial and Applied Mathematics (1996). DOI 10.1137/1.9781611971491. URL <http://epubs.siam.org/doi/abs/10.1137/1.9781611971491>
41. Stoer, J., Bulirsch, R., Bartels, R.H., Gautschi, W., Witzgall, C.: *Introduction to numerical analysis*. Texts in applied mathematics. Springer, New York (2002). URL <http://opac.inria.fr/record=b1098819>
42. Strassen, V.: Gaussian elimination is not optimal. *Numerische mathematik* **13**(4), 354–356 (1969)
43. Tal-Ezer, H.: High degree interpolation polynomial in Newton form. Contractor report 181677, ICASE report no. 88-39, NASA Langley Research Center (1988)
44. Tirosh, S., De Ville, D., Unser, M.: Polyharmonic smoothing splines and the multidimensional Wiener filtering of fractal-like signals. *IEEE Transactions on Image Processing* **15**(9), 2616–2630 (2006)

-
45. Tveit, A.: On the complexity of matrix inversion. *Mathematical Note* p. 1 (2003)
 46. Unser, M.: Splines: A perfect fit for signal and image processing. *IEEE Signal processing magazine* **16**(6), 22–38 (1999)
 47. Unser, M., Aldroubi, A., Eden, M.: Fast B-spline transforms for continuous image representation and interpolation. *IEEE Transactions on Pattern Analysis & Machine Intelligence* (3), 277–285 (1991)
 48. Unser, M., Blu, T.: Generalized smoothing Splines and the optimal discretization of the Wiener filter. *IEEE Trans. Signal Process.* **53**(6), 2146–2159 (2005)
 49. Walston, D.E.: *An introduction to numerical analysis* (1968)
 50. Werner, W.: Polynomial interpolation: Lagrange versus Newton. *Mathematics of computation* **43**(167), 205–217 (1984)