Kernel-based hyperinterpolation for scattered data

Mohammad Karimnejad Esfahani[1*], Maryam Mohammadi[2], Gabriele Santin[1], Stefano De Marchi[3]

[1*] Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, Via Torino 155, Mestre, 30172, Italy.

[2]Faculty of Mathematical Sciences and Computer, Kharazmi University, Tehran, Iran.

[3]Department of Medicine, Padova University, Via Giustiniani, 2, Padova, 35128, Italy.

*Corresponding author(s). E-mail(s): mohammad.karimnejad@unive.it; Contributing authors: m.mohammadi@khu.ac.ir; gabriele.santin@unive.it; stefano.demarchi@unipd.it;

Abstract

We extend hyperinterpolation from the polynomial setting to arbitrary reproducing kernel Hilbert spaces, replacing exact interpolation with an L_2 projection onto the space $\mathcal{N}_K(X)$ spanned by the kernel translates, discretized by a positive-weight cubature formula. For a strictly positive definite kernel of smoothness $\tau > d/2$ and data sites X with fill distance h_X, the resulting operator satisfies an $L_\infty \to L_2$ stability bound with a constant independent of n, and the L_2 error estimate $O(h_X^{\tau-d/2})$ for functions in the native space, that is $O(n^{-\tau/d+1/2})$ for quasi-uniform points. We further derive an upper bound on the Lebesgue constant and show that this bound cannot fall below $O(n^{1/2})$, the growth rate of standard kernel interpolation, regardless of the cubature nodes and weights. The required cubature formulas are constructed by the Backus–Gilbert least-squares approach, for which we give an existence result on equidistributed sequences. Numerical experiments on four benchmark functions in one and two dimensions, with Basic, Linear and Quadratic Matérn kernels, are consistent with the predicted rates.

Keywords: Kernel-based hyperinterpolation, Reproducing kernel Hilbert space, Positive-weight cubature, Lebesgue constant, Scattered data approximation

MSC Classification: 65D05 , 41A63 , 65D32

1 Introduction

The approximation of scattered data using kernel-based methods has become fundamental in numerical analysis, machine learning, and scientific computing [1–4]. Given data sites $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \subset \Omega \subset \mathbb{R}^d$ and function values $\mathbf{f}_X = [f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n)]^T$, kernel interpolation constructs an approximant

$$S_{X,K}f(\mathbf{x}) = \sum_{j=1}^{n} \alpha_j K(\mathbf{x}, \mathbf{x}_j), \quad \forall \mathbf{x} \in \mathbb{R}^d, \tag{1}$$

in the native space $\mathcal{N}_K(\Omega)$ of a strictly positive definite kernel K that exactly reproduces the data. The coefficient vector $\alpha = [\alpha_1, \ldots, \alpha_n]^T$ solves the linear system

$$\mathbf{K}\alpha = \mathbf{f}_X, \tag{2}$$

where $\mathbf{K} \in \mathbb{R}^{n\times n}$ is the kernel matrix with entries $\mathbf{K}_{ij} = K(\mathbf{x}_i, \mathbf{x}_j)$.

While kernel interpolation is theoretically optimal in the native space norm, it faces critical practical obstacles, especially when the interpolation points are not sufficiently nicely distributed. Letting the fill- and separation-distances be defined as

$$h_{X,\Omega} \coloneqq \sup_{x\in\Omega} \min_{x_i\in X} \|\mathbf{x} - \mathbf{x}_i\|, \quad q_X \coloneqq \frac{1}{2} \min_{x_i \neq \mathbf{x}_j, \mathbf{x}_i, x_j \in X} \|\mathbf{x}_i - \mathbf{x}_j\|.$$

There are well-known instability phenomena which can be described as follows:

- Conditioning: The kernel matrix $\mathbf{K}$ becomes severely ill-conditioned as n increases, especially for smooth kernels. For the widely used Sobolev kernels, which will be formally introduced later, the smallest eigenvalue typically grows as $\lambda_n(\mathbf{K}) = O(q_X^{-2\tau/d})$, where τ is a parameter measuring the kernel smoothness (see [4], a precise definition will be given later). This renders direct solution methods numerically unstable even in double precision arithmetic.
- Stability-accuracy trade-off: The Lebesgue constant

$$\Lambda(X) = \sup_{\mathbf{x}\in\Omega} \sum_{j=1}^{n} |\ell_j(\mathbf{x})|, \tag{3}$$

  where ℓ_j are cardinal basis functions, quantifies the stability of interpolation when using the direct formulation (1). For kernel interpolation with Sobolev kernels and quasi-uniform points, $\Lambda(X) \leq O(h_X^{-d/2}) = O(n^{1/2})$. Even when using more sophisticated, stabilized formulation of the interpolant, this instability is not merely a computational artifact but it reflects genuine sensitivity of the interpolant to perturbations in the data, see [5].

1.1 Existing Approaches

Several strategies have been proposed to address interpolation instability:

- Regularization methods such as ridge regression add λI to $\mathbf{K}$, yielding the modified system $(\mathbf{K} + \lambda I)\alpha = \mathbf{f}_X$. This improves conditioning while sacrificing exact interpolation, but optimal choice of λ may still lead to optimal asymptotical rates of convergence for Sobolev kernels [6].
- Domain decomposition methods partition Ω into subdomains and apply local interpolation in each region, such as for Partition of Unity methods [7].
- Stable bases have been introduced to avoid directly solving (1) and instead resort to more stable alternative formulations [8, 9].

1.2 Polynomial Hyperinterpolation: A Paradigm Shift

Polynomial hyperinterpolation, pioneered by Sloan [10], offers an alternative approach: instead of imposing exact interpolation at data sites, one projects the target function onto the polynomial space using an L_2 inner product discretized by a positive-weight cubature formula. This yields a number of advantages:

1. Unconditional stability: Lebesgue constants remain bounded independently of the polynomial degree.
2. Near-optimal approximation: The L_2 error is at most a constant factor worse than the best polynomial approximation.
3. Practical computability: Implementation via weighted least squares avoids explicit basis orthogonalization.

However, polynomial hyperinterpolation is inherently restricted to specific geometric configurations (spheres, balls, simplices) where orthogonal polynomial bases and exact cubature rules are explicitly known. On general domains, and for scattered data distributions, such bases and rules are in general not available in closed form, which limits the practical applicability of the polynomial construction.

1.3 Our Contributions

This paper extends hyperinterpolation from the polynomial setting to arbitrary reproducing kernel Hilbert spaces. Specifically:

1. Theoretical framework: For a strictly positive definite kernel K and data sites X, we show that an L_2-orthonormal basis of $\mathcal{N}_K(X) = \mathrm{span}\{K(\cdot, \mathbf{x}) : \mathbf{x} \in X\}$ can be used, together with a suitable cubature formula, to define a hyperinterpolation operator with an $L_\infty \to L_2$ stability bound independent of n, an upper bound on the Lebesgue constant, and an L_2 error estimate; we also show that this upper bound cannot be improved below $O(n^{1/2})$.
2. Cubature construction: We adapt the Backus–Gilbert least-squares approach to build positive-weight cubature formulas exact on $\mathcal{N}_K(X) \times \mathcal{N}_K(X)$, with an existence result for equidistributed sequences.
3. Computation: The hyperinterpolant can be computed via a QR/SVD factorization of the weighted kernel matrix, avoiding the normal equations, at a cost of $O(Mn^2)$ for M cubature points ($O(n^3)$ when $M = O(n)$).

4. Error analysis: For kernels of smoothness $\tau > d/2$ and $f \in \mathcal{N}_K(\Omega)$, we derive the convergence rate

$$\|f - L_{K,n}f\|_{L_2} = O(n^{-\tau/d+1/2}) \quad \text{as } n \to \infty,$$

one factor of $n^{-1/2}$ short of the optimal interpolation rate $O(n^{-\tau/d})$.
5. Numerical tests: On four benchmark functions in one and two dimensions, the observed convergence and Lebesgue-constant growth are consistent with the theory, full- and reduced-exactness cubatures perform comparably, and a comparison with polynomial hyperinterpolation shows lower condition numbers for the latter but similar Lebesgue-constant growth.

1.4 Paper Organization and Notation

Section 2 reviews kernel interpolation and native space characterization. Section 3 develops hyperinterpolation theory for general Hilbert spaces. Section 4 specializes to strictly positive definite kernels with explicit error rates. Section 5 presents numerical experiments and comparisons. Section 6 discusses extensions and open problems.

2 Preliminaries

Here and in the following, $C(\Omega)$ denotes continuous functions defined on Ω and $L_2(\Omega)$ square-integrable functions with respect to a measure ω. We use $\kappa(\mathbf{K})$ for the 2-condition number of the matrix $\mathbf{K}$. Constants C may differ from line to line, but depend only on fixed problem parameters.

We also assume Ω is a bounded domain with a Lipschitz boundary, although many results extend to unbounded domains with appropriate decay conditions.

Finally, consider $K : \Omega \times \Omega \to \mathbb{R}$ be a radial strictly positive definite kernel, which means that $K(\mathbf{x}, \mathbf{y}) = \phi(\|\mathbf{x} - \mathbf{y}\|_2)$ for some function $\phi : [0, \infty) \to \mathbb{R}$, and that the kernel matrix $\mathbf{K} \in \mathbb{R}^{n \times n}$ is positive definite for any set of distinct points $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, meaning that $\alpha^T \mathbf{K} \alpha > 0$ for all nonzero $\alpha \in \mathbb{R}^n$.

Remark 1 While we restrict our attention to radial kernels for concreteness, the hyperinterpolation framework extends to non-radial kernels (e.g., tensor-product kernels, anisotropic kernels) with minor modifications. The key requirement is that K generates a reproducing kernel Hilbert space with a well-defined native space norm, as detailed in the following section.

2.1 Native Spaces and Reproducing Properties

For our purposes, $\mathcal{N}_K(\Omega)$ denotes the *native space* of K, which is a Hilbert space of functions on Ω equipped with the inner product $\langle f, g \rangle_{\mathcal{N}_K} = \sum_{i,j} \alpha_i \beta_j K(\mathbf{x}_i, \mathbf{y}_j)$ where $f = \sum_i \alpha_i K(\cdot, \mathbf{x}_i)$ and $g = \sum_j \beta_j K(\cdot, \mathbf{y}_j)$. The native space is a reproducing kernel Hilbert space with the reproducing property

$$\langle f, K(\cdot, \mathbf{y}) \rangle_{\mathcal{N}_K} = f(\mathbf{y}) \quad \text{for all } f \in \mathcal{N}_K(\Omega),\ \mathbf{y} \in \Omega.$$

The subspace $\mathcal{N}_K(X) = \mathrm{span}\{K(\cdot, \mathbf{x}_i),\ 1 \le i \le n\} \subset \mathcal{N}_K(\Omega)$ consists of all linear combinations of kernel translates centered on X.

Now, we recall a specific class of radial kernels that is of interest for the rest of this work. The Matérn family is defined by

$$\phi(r) = \frac{2^{1-\nu}}{\Gamma(\nu)}(\varepsilon r)^{\nu} K_{\nu}(\varepsilon r),$$

where the smoothness degree (τ) of the kernel is given by $\tau = \nu + d/2$ and K_ν denotes the modified Bessel function of the second kind. When Ω is a bounded domain with a Lipschitz boundary, these kernels generate a native space that is norm equivalent to the Sobolev space $W_2^{\tau}(\Omega)$, meaning that the two are equal as sets and the corresponding norms are equivalent. We refer the reader to [4, Chap. 10] for further details, and we simply refer to these kernels as Sobolev kernels of smoothness τ in the following.

2.2 Standard Kernel Interpolation and its Stability

The kernel interpolant of f given data $\mathbf{f}_X$ is represented as (1), where the coefficients $\alpha = [\alpha_1, \ldots, \alpha_n]^T$ are obtained by solving the linear system (2). Since K is strictly positive definite, the kernel matrix $\mathbf{K}$ is invertible and the solution exists uniquely. The interpolant can equivalently be written using cardinal basis functions ℓ_j of $\mathcal{N}_K(X)$ defined by $\ell_j(\mathbf{x}_i) = \delta_{ij}$, i.e.,

$$S_{X,K}f(\mathbf{x}) = \sum_{j=1}^{n} f(\mathbf{x}_j)\ell_j(\mathbf{x}), \quad \ell_j(\mathbf{x}) = \sum_{k=1}^{n} (\mathbf{K}^{-1})_{jk} K(\mathbf{x}, \mathbf{x}_k).$$

Moreover, the stability of the interpolation as a map from $\mathbf{f}_X$ to $S_{X,K}f$ is quantified by the Lebesgue constant defined in (3) which gives the bound:

$$\|S_{X,K}f\|_{L_\infty(\Omega)} \le \Lambda(X)\|\mathbf{f}_X\|_{\ell_\infty}.$$

Consequently, small perturbations δf in the data $\mathbf{f}_X$ can be amplified by a factor $\Lambda(X)$ in the interpolant, leading to numerical instability when $\Lambda(X) \gg 1$.

For Sobolev kernels with any smoothness $\tau > d/2$ and for quasi-uniform points with fill distance h_X, it is known [5] that

$$\Lambda(X) \le C h_X^{-d/2} \le C' n^{1/2} \quad \text{as } h_X \to 0 \text{ or } n \to \infty.$$

This growth with n is a fundamental limitation of exact interpolation. Although this growth is rather limited, it is known that the computation of the interpolant is possibly ill-conditioned, as the smallest eigenvalue of the kernel matrix instead grows as $q_X^{2\tau-d}$ [11].

2.3 Native Space Characterization via Mercer's Theorem

Under quite general assumptions (see [12]), the native space admits an elegant characterization via the spectral decomposition of the integral operator [4, 13]

$$T_K(v)(\mathbf{x}) = \int_\Omega K(\mathbf{x},\mathbf{y})v(\mathbf{y})\, d\omega(\mathbf{y}), \quad v \in L_2(\Omega),\ \mathbf{x} \in \Omega,$$

where ω is a positive Borel measure on Ω (e.g., the Lebesgue measure).

Mercer's theorem [4, Thm 10.29] states that if K is continuous and

$$\int_\Omega \int_\Omega K(\mathbf{x},\mathbf{y})v(\mathbf{x})v(\mathbf{y})\, d\omega(\mathbf{x})d\omega(\mathbf{y}) \geq 0 \quad \forall v \in L_2(\Omega),$$

then T_K is a compact, self-adjoint and positive operator with spectral decomposition $T_K\varphi_k = \lambda_k\varphi_k$, where $\{\lambda_k\}$ are non-negative eigenvalues (listed with multiplicity) satisfying $\lambda_1 \geq \lambda_2 \geq \cdots \to 0$, and $\{\varphi_k\}$ are L_2-orthonormal eigenfunctions. Moreover,

$$K(\mathbf{x},\mathbf{y}) = \sum_{k=1}^{\infty} \lambda_k\varphi_k(\mathbf{x})\varphi_k(\mathbf{y}),$$

where the series converges absolutely and uniformly on $\Omega \times \Omega$. The native space is then characterized as

$$\mathcal{N}_K(\Omega) = \left\{ f \in L_2(\Omega) : \sum_{k=1}^{\infty} \frac{1}{\lambda_k} |\langle f, \varphi_k\rangle_{L_2(\Omega)}|^2 < \infty \right\}, \tag{4}$$

equipped with the norm

$$\|f\|^2_{\mathcal{N}_K} = \sum_{k=1}^{\infty} \frac{1}{\lambda_k} |\langle f, \varphi_k\rangle_{L_2}|^2.$$

Any $f \in L_2(\Omega)$ admits the generalized Fourier series

$$f(\mathbf{x}) = \sum_{k=1}^{\infty} \langle f, \varphi_k\rangle_{L_2(\Omega)}\varphi_k(\mathbf{x}), \tag{5}$$

converging absolutely and uniformly if $f \in C(\Omega)$. This representation shows $\mathcal{N}_K(\Omega) \subset L_2(\Omega)$ with continuous embedding.

Remark 2 For most kernels (except for a few cases, see, e.g. [2] for a discussion), the eigenfunctions $\{\varphi_k\}$ are not known analytically, however their numerical computation has been studied in [14]. In addition, the approximation of f as in (5) with the characterization of the native space (4), shows that recovery of f through (5) should be possible for the basis that spans the native space and is only L_2 orthonormal. In [15] and [16], the construction of this basis has been investigated for any positive definite or conditionally positive definite kernel.

3 Generalized Hyperinterpolation

In this section we develop hyperinterpolation theory for abstract Hilbert spaces embedded in $C(\Omega)$, which will be specialized to kernel native spaces in Section (4). The key innovation is replacing exact pointwise interpolation with L_2 projection discretized via positive-weight cubature formulas.

3.1 Abstract Setting and Projection

Let $\Omega \subset \mathbb{R}^d$ be bounded, ω be a given positive measure on Ω, and let $V_n \subset C(\Omega) \subset L_2(\Omega, \omega)$ be a Hilbert space of dimension n with L_2-orthonormal basis $V = \{v_i\}_{i=1}^n$. We will sometimes omit the dependence on ω if it is clear from the context. The L_2 projection of $f \in C(\Omega)$ onto V_n is

$$L_{V,n}f = \sum_{i=1}^{n} \langle f, v_i \rangle_{L_2(\Omega)} v_i, \tag{6}$$

where $\langle \cdot, \cdot \rangle_{L_2(\Omega)} = \int_\Omega (\cdot)(\cdot)\, d\omega$. We call $L_{V,n} : L_2(\Omega, \omega) \to V_n$ the hyperinterpolation operator. It is the unique minimizer of the L_2 distance, i.e.,

$$L_{V,n}f = \operatorname*{argmin}_{g \in V_n} \|f - g\|_{L_2(\Omega)}.$$

Direct evaluation of (6) requires computing integrals $\langle f, v_i \rangle_{L_2}$, which is generally infeasible. We discretize it using a cubature rule

$$\int_\Omega g(\mathbf{x})\, d\omega(\mathbf{x}) \approx \sum_{k=1}^{M} w_k g(\mathbf{z}_k), \quad g \in C(\Omega), \tag{7}$$

with cubature points $Z = \{\mathbf{z}_1, \dots, \mathbf{z}_M\} \subset \Omega$ and positive weights $w_k > 0$. We require the cubature to be exact on $V_n^2 = V_n \times V_n = \operatorname{span}\{v_i v_j : 1 \le i, j \le n\}$, i.e.,

$$\int_\Omega g(\mathbf{x})\, d\omega(\mathbf{x}) = \sum_{k=1}^{M} w_k g(\mathbf{z}_k) \quad \forall g \in V_n^2. \tag{8}$$

According to [17], there exists such a formula with $M \le \dim(V_n^2)$; however, the paper does not discuss how to construct such a cubature formula. Section (3.4) addresses constructive algorithms, while for now we simply assume that such a cubature formula exists. With exactness (8), we define the discrete hyperinterpolation operator

$$L_{V,n}f = \sum_{j=1}^{n} \langle f, v_j \rangle_{\ell_{2,w}} v_j, \tag{9}$$

where the discrete inner product is

$$\langle f, v_j \rangle_{\ell_{2,w}} = \sum_{k=1}^{M} w_k f(\mathbf{z}_k) v_j(\mathbf{z}_k).$$

Throughout, we distinguish three different inner products, defined as follows:

- $\langle \cdot, \cdot \rangle_{L_2(\Omega)}$: the continuous L_2 inner product with measure ω.
- $\langle \cdot, \cdot \rangle_{\ell_{2,w}}$: the discrete inner product using cubature weights $\{w_k\}$.
- $\langle \cdot, \cdot \rangle_{\ell_{2,r}}$: the discrete inner product with alternative weights $\{r_k\}$ (used in Section 3.4).

3.2 Fundamental Properties of Hyperinterpolation

The following lemmas establish that discrete and continuous projections coincide when exactness holds.

Lemma 1 If the cubature rule (7) is exact on V_n^2, then the basis $V = \{v_1, \ldots, v_n\}$ is $\ell_{2,w}$-orthonormal if and only if it is L_2-orthonormal.

Proof Assume V is L_2-orthonormal, i.e., $\langle v_i, v_j \rangle_{L_2} = \delta_{ij}$. Then

$$\begin{aligned} \langle v_i, v_j \rangle_{\ell_{2,w}} &= \sum_{k=1}^{M} w_k v_i(\mathbf{z}_k) v_j(\mathbf{z}_k) \\ &= \int_\Omega v_i(\mathbf{x}) v_j(\mathbf{x})\, d\omega(\mathbf{x}) \quad \text{(by exactness on } v_i v_j \in V_n^2\text{)} \\ &= \langle v_i, v_j \rangle_{L_2(\Omega)} = \delta_{i,j}. \end{aligned}$$

The reverse direction follows by reversing the steps, using that exactness allows us to compute L_2 inner products via the discrete formula. □

Lemma 2 Any cubature formula satisfying (8) (i.e., being exact on V_n^2) must have $M \geq n$.

Proof Construct the matrix $Q \in \mathbb{R}^{n \times M}$ with the entries $Q_{i,k} = v_i(\mathbf{z}_k)\sqrt{w_k}$. Since Lemma 1 implies that the L_2 orthonormal basis is also $\ell_{2,w}$ orthonormal, it follows that

$$(QQ^T)_{i,j} = \sum_{k=1}^{M} v_i(\mathbf{z}_k) w_k v_j(\mathbf{z}_k) = \langle v_i, v_j \rangle_{\ell_{2,w}} = \delta_{i,j},$$

thus $QQ^T = I_n$. This implies $\operatorname{rank}(Q) = n$, which is possible only if $M \geq n$. □

The next theorem characterizes the case in which hyperinterpolation reduces to standard interpolation.

Theorem 3 *Assuming $M \geq n$, the hyperinterpolant (9) satisfies the interpolation conditions*

$$(L_{V,n}f)(\mathbf{z}_k) = f(\mathbf{z}_k) \;\; \forall k = 1, \dots, M$$

for all $f \in C(\Omega)$ if and only if $M = n$.

Proof Proceeding as in the proof of Lemma 2, for $M = n$ the matrix $Q \in \mathbb{R}^{n \times M}$ with $Q_{i,k} = v_i(\mathbf{z}_k)\sqrt{w_k}$ satisfies $QQ^T = I_n$ and, since Q is square, this implies $Q^T Q = I_n$ as well (since Q is orthogonal). Writing out $(Q^T Q)_{k,i}$ we get

$$\sum_{j=1}^{n} w_k^{1/2} v_j(\mathbf{z}_k) w_i^{1/2} v_j(\mathbf{z}_i) = \delta_{k,i}.$$

Dividing by $\sqrt{w_k w_i}$:

$$\sum_{j=1}^{n} v_j(\mathbf{z}_k) v_j(\mathbf{z}_i) = \frac{1}{w_k}\delta_{k,i}. \tag{10}$$

Now evaluate the hyperinterpolant at $\mathbf{z}_k$:

$$\begin{aligned}
(L_{V,n}f)(\mathbf{z}_k) &= \sum_{j=1}^{n} \langle f, v_j \rangle_{\ell_{2,w}} v_j(\mathbf{z}_k) \\
&= \sum_{j=1}^{n} \left(\sum_{i=1}^{M} w_i f(\mathbf{z}_i) v_j(\mathbf{z}_i) \right) v_j(\mathbf{z}_k) \\
&= \sum_{i=1}^{M} w_i f(\mathbf{z}_i) \sum_{j=1}^{n} v_j(\mathbf{z}_i) v_j(\mathbf{z}_k) \\
&= \sum_{i=1}^{M} w_i f(\mathbf{z}_i) \cdot \frac{1}{w_i}\delta_{i,k} \quad \text{(by (10) with } M = n) \\
&= f(\mathbf{z}_k),
\end{aligned}$$

demonstrating that the interpolation conditions are fulfilled.

Assume instead that $(L_{V,n}f)(\mathbf{z}_k) = f(\mathbf{z}_k)$ for all k and all $f \in C(\Omega)$. For each $i \in \{1, \dots, M\}$, we take $f_i \in C(\Omega)$ to be a function with value 1 in z_i and 0 in z_j for $j \neq i$. Then by (9) we have

$$\delta_{\ell,i} = L_{V,n} f_i(\mathbf{z}_\ell) = \sum_{j=1}^{n} \left(\sum_{k=1}^{M} w_k f_i(\mathbf{z}_k) v_j(\mathbf{z}_k) \right) v_j(\mathbf{z}_\ell) = \sum_{j=1}^{n} w_i v_j(\mathbf{z}_i) v_j(\mathbf{z}_\ell),$$

forcing property (10) to hold. Combined with $QQ^T = I_n$ from Lemma 1, we need $Q^T Q = I_n$. Both conditions can only be met simultaneously if Q is square and, hence $n = M$. □

We observe in particular that equation (10) provides an explicit formula for cubature weights when $n = M$, since

$$w_k = \left(\sum_{j=1}^{n} v_j^2(\mathbf{z}_k) \right)^{-1}, \quad 1 \leq k \leq n.$$

Lemma 4 Assume the cubature formula (7) satisfies (8). Then $L_{V,n} f = f$ if $f \in V_n$.

Proof Since $f \in V_n$, write $f = \sum_{i=1}^{n} a_i v_i$. From (9) we have

$$\begin{aligned} L_{V,n} f &= \sum_{j=1}^{n} \left\langle \sum_{i=1}^{n} a_i v_i, v_j \right\rangle_{\ell_{2,w}} v_j = \sum_{j=1}^{n} \sum_{i=1}^{n} a_i \langle v_i, v_j \rangle_{\ell_{2,w}} v_j \\ &= \sum_{j=1}^{n} a_j v_j = f, \end{aligned}$$

using in the last step the orthonormality $\langle v_i, v_j \rangle_{\ell_{2,w}} = \delta_{i,j}$. □

Lemma 4 shows that hyperinterpolation is a projection operator with the property that when $M = n$, the hyperinterpolation is equivalent to standard interpolation. We conclude this section with the following remark.

Remark 3 For certain domains (e.g., sphere S^d with $d \geq 2$ and polynomial exactness degree $2n$ with $n \geq 3$), minimal cubature formulas with $M = n$ do not exist [18, 19]. In such cases, hyperinterpolation with $M > n$ is fundamentally different from interpolation.

3.3 Computation of the Hyperinterpolant

Equation (9) shows that $L_{V,n} f$ is the $\ell_{2,w}$-orthogonal projection of f onto V_n. Equivalently, $L_{V,n} f$ solves the weighted least squares problem

$$L_{V,n} f = \operatorname*{argmin}_{v \in V_n} \|f - v\|_{\ell_{2,w}}^2 .$$

Let $\{v_1, \ldots, v_n\}$ be any basis of V_n (not necessarily L_2-orthonormal). Any $v \in V_n$ can be written as

$$v = \sum_{j=1}^{n} a_j v_j,$$

so that

$$(v(\mathbf{z}_1), \ldots, v(\mathbf{z}_M))^T = V_Z a,$$

where $a = (a_1, \ldots, a_n)^T$, and $V_Z \in \mathbb{R}^{M \times n}$ is the matrix with entries $(V_Z)_{m,j} = v_j(\mathbf{z}_m)$. Setting $y = (f(\mathbf{z}_1), \ldots, f(\mathbf{z}_M))^T$ and $W = \operatorname{diag}(w_1, \ldots, w_M)$, we obtain

$$\begin{aligned} \|f - v\|_{\ell_{2,w}}^2 &= \sum_{m=1}^{M} w_m \left(f(\mathbf{z}_m) - v(\mathbf{z}_m)\right)^2 \\ &= (y - V_Z a)^T W (y - V_Z a) \\ &= y^T W y - 2a^T V_Z^T W y + a^T V_Z^T W V_Z a, \end{aligned}$$

which is minimized with respect to a by setting the gradient to zero, i.e.,

$$V_Z^T W V_Z a = V_Z^T W y. \tag{11}$$

This is the normal equation for weighted least squares. While (11) can be solved directly, the condition number may grow quadratically, i.e., $\kappa(V_Z^T W V_Z) \sim \kappa(V_Z)^2$. A more stable, standard approach that uses QR decomposition is given in Algorithm 1.

Algorithm 1 Stable Computation of Hyperinterpolant Coefficients

1: **Input:** Basis evaluations $V_Z \in \mathbb{R}^{M \times n}$, weights $W = \mathrm{diag}(w_1, \dots, w_M)$, data $y \in \mathbb{R}^M$
2: **Output:** Coefficients $a \in \mathbb{R}^n$ such that $L_{V,n} f = \sum_{j=1}^n a_j v_j$
3: Compute $\tilde{V}_Z \leftarrow W^{1/2} V_Z$ ▷ Weighted basis matrix
4: Compute $\tilde{y} \leftarrow W^{1/2} y$ ▷ Weighted data
5: Compute QR decomposition: $\tilde{V}_Z = QR$ where $Q \in \mathbb{R}^{M \times n}$, $R \in \mathbb{R}^{n \times n}$
6: Solve $Ra = Q^T \tilde{y}$ by back-substitution ▷ Avoids forming $V_Z^T W V_Z$ explicitly
7: **return** a

The QR decomposition costs $O(Mn^2)$, and back-substitution costs $O(n^2)$, implying a total cost $O(Mn^2)$ compared to $O(n^3)$ for standard interpolation. When $M = O(n)$, costs are comparable at $O(n^3)$. Moreover, letting $\kappa(A)$ be the 2-norm condition number of A, we have the following.

$$\kappa(V_Z^T W V_Z) = \kappa(\tilde{V}_Z)^2. \tag{12}$$

Consequently, Algorithm 1 avoids squaring the condition number by working with $\tilde{V}_Z$ via QR decomposition rather than forming $V_Z^T W V_Z$ explicitly. However, note that when $V_n = \mathcal{N}_K(X)$ with the kernel-translate basis, $\kappa(\tilde{V}_Z)$ is of magnitude comparable to $\kappa(\mathbf{K})^{1/2}$, so the overall condition number remains comparable to that of standard interpolation.

Remark 4 Note that when $V_n = \mathbb{P}_q^d$ (polynomials of degree $\leq q$ in dimension d), there exists a simple closed form

$$L_{V,n} f(\mathbf{x}) = \sum_{j=1}^{M} w_j G_n(\mathbf{x}, \mathbf{z}_j) f(\mathbf{z}_j),$$

where G_n is the reproducing kernel of $\mathbb{P}_q^d$ [19]. This avoids computing the basis coefficients entirely.

3.4 Construction of the Cubature Formula

Let $n' = \dim(V_n \oplus 1)$, where 1 is the constant function with value 1, so that $n' = n+1$ if $1 \notin V_n$ and $n' = n$ otherwise. Denote $V_{n'}^2 = \mathrm{span}\{g_1, \dots, g_{n'(n'+1)/2}\}$, where $g_i = v_k v_m$ for $1 \leq k \leq m \leq n'$. However, note that $\dim(V_{n'}^2) \leq n'(n'+1)/2$, with equality only when the products are linearly independent. The exactness condition (8) yields the linear system

$$AW = b, \tag{13}$$

where $A_{i,j} = g_i(z_j), b_i = \int_\Omega g_i(x)\, d\omega(x)$ and W is the vector of coefficients.

Recalling the set of cubature points $Z = \{z_1, \ldots, z_M\}$, we generally have $\mathrm{rank}(A) \leq M$ i.e. the system (13) is underdetermined and admits either no solution or infinitely many solutions. In the latter case, following [20, 21], the optimal solution is achievable based on the Backus-Gilbert method [22]. We minimize the weighted ℓ^2-norm

$$\frac{1}{2}\sum_{j=1}^{M} \frac{w_j^2}{r_j}$$

subject to the constraint (13), where $r_j > 0$ are prescribed prior weights. This constrained minimization problem admits the explicit solution

$$W = RA^\top (ARA^\top)^{-1} b. \tag{14}$$

In practice, we first solve the linear system $(ARA^\top)\lambda = b$, then recover $W = RA^\top\lambda$. This is the minimum weighted-norm solution of the underdetermined system:

$$W = \arg\min_{AW=b} \|R^{-1/2}W\|_2. \tag{15}$$

The following theorem which rephrases [20, Cor 3.6], guaranties the existence of positive-weight cubature formulas under mild conditions.

Theorem 5 *Assume $\Omega \subset \mathbb{R}^d$, $\omega : \Omega \to \mathbb{R}^+$ be a Riemann integrable weight function positive almost everywhere, and $\mathcal{F}_K(\Omega) \subset C(\Omega)$ a K-dimensional space of continuous bounded functions with $1 \in \mathcal{F}_K(\Omega)$. Let $Y = \{z_1, z_2, \ldots\}$ be an equidistributed sequence in Ω with $\omega(Y) > 0$. Then there exists $N_0 \in \mathbb{N}$ such that for all $M \geq N_0$, the Backus-Gilbert cubature formula with prior weights*

$$r_i = \frac{|\Omega|\omega(z_i)}{M}, \quad i = 1, \ldots, M,$$

and solution given by (15), has strictly positive weights $w_i > 0$ and is $\mathcal{F}_K(\Omega)$-exact.

In the following algorithm adapted from [20, Algorithm 1] we construct the cubature formula. There are two major differences in the algorithm stated here and in [20]. First, the loop is initialized with $M_0 < (n'+1)n'/2$: The reason is that the dimension of $V_{n'}^2$ is not known precisely, so starting from a smaller M_0 the true rank of A is also discovered with the lines 5–8. Second, the number of nodes increases by one at each iteration, rather than doubling. Doubling reaches a feasible M in fewer iterations, but it overshoots: the returned rule may carry considerably more nodes than are needed for exactness and positivity.

Algorithm 2 Construction of the cubature formula

$M \leftarrow M_0$, $r \leftarrow 0$, $r_{\max} \leftarrow 0$, $w_{\min} \leftarrow 0$, and $K \leftarrow n'(n'+1)/2$
while $r < K$ or $w_{\min} \le 0$ **do**
    Generate equidistributed points $Z = \{z_m\}_{m=1}^{M}$
    Form the linear system $AW = b$
    Compute the rank of A: $r \leftarrow \text{rank}(A)$
    if $r = r_{\max}$ **then**
        $K \leftarrow r$ ▷ the generators $\{g_i\}$ are dependent and $K = \dim(V_{n'} \times V_{n'})$
    end if
    $r_{\max} \leftarrow r$
    if $r = K$ **then**
        Compute the LS weights as $W = RA^\top(ARA^\top)^{-1}b$
        Determine the smallest weight: $w_{\min} = \min(W)$
    end if
    $M \leftarrow M + 1$
end while

3.5 Error and Stability Analysis

We now establish rigorous error bounds for hyperinterpolation. The following lemma characterizes the projection property.

Lemma 6 Let $f \in C(\Omega)$. Then:

(i) $\langle f - L_{V,n}f, \chi\rangle_{\ell_{2,w}} = 0$ for all $\chi \in V_n$ (orthogonality),

(ii) $\left\|L_{V,n}f\right\|_{\ell_{2,w}}^2 + \left\|f - L_{V,n}f\right\|_{\ell_{2,w}}^2 = \|f\|_{\ell_{2,w}}^2$ (Pythagorean theorem),

(iii) $\left\|L_{V,n}f\right\|_{\ell_{2,w}}^2 \le \|f\|_{\ell_{2,w}}^2$ (contraction),

(iv) $\left\|f - L_{V,n}f\right\|_{\ell_{2,w}}^2 = \min_{\chi \in V_n} \|f - \chi\|_{\ell_{2,w}}^2$ (best approximation).

Proof Property (i) follows from the definition: $L_{V,n}$ is the $\ell_{2,w}$-orthogonal projection onto V_n. Property (ii) follows by taking $\chi = L_{V,n}f$ in (i). Properties (iii) and (iv) are immediate from (i) and (ii). □

The next theorem is central to hyperinterpolation theory, providing stability and approximation guaranties.

Theorem 7 *Let* $S_M = \sum_{j=1}^{M} w_j > 0$, *and* $\mu(\Omega) = \int_\Omega d\omega$ *be the measure of* Ω. *For* $f \in C(\Omega)$ *we have the stability bound*

$$\left\|L_{V,n}f\right\|_{L_2(\Omega)} \le \sqrt{S_M}\,\|f\|_{L_\infty(\Omega)}\,, \tag{16}$$

and the error bound;

$$\left\|f - L_{V,n}f\right\|_{L_2(\Omega)} \le \left(\sqrt{\mu(\Omega)} + \sqrt{S_M}\right) \inf_{g \in V_n} \|g - f\|_{L_\infty(\Omega)}. \tag{17}$$

Furthermore, if the cubature is exact for constants (i.e. $\sum_j w_j = \mu(\Omega)$*), then*

$$\left\|f - L_{V,n}f\right\|_{L_2(\Omega)} \le 2\sqrt{\mu(\Omega)} \inf_{g \in V_n} \|g - f\|_{L_\infty(\Omega)}.$$

Proof Since $L_{V,n}f \in V_n$, we have $(L_{V,n}f)^2 \in V_n \times V_n$. By exactness of the cubature, it holds

$$\left\|L_{V,n}f\right\|^2_{L_2(\Omega)} = \left\|L_{V,n}f\right\|^2_{\ell_{2,w}} \le \|f\|^2_{\ell_{2,w}} = \sum_{j=1}^{M} w_j f(z_j)^2 \le \|f\|^2_{L_\infty(\Omega)} S_M.$$

This implies (16).

Now choose an arbitrary $g \in V_n$. By Lemma 4, $L_{V,n}(g) = g$. Using the triangle inequality:

$$\begin{aligned}
\left\|f - L_{V,n}f\right\|_{L_2(\Omega)} &\le \|f - g\|_{L_2(\Omega)} + \left\|g - L_{V,n}f\right\|_{L_2(\Omega)} \\
&= \|f - g\|_{L_2(\Omega)} + \left\|L_{V,n}(g - f)\right\|_{L_2(\Omega)} \\
&\le \sqrt{\mu(\Omega)}\, \|f - g\|_{L_\infty(\Omega)} + \sqrt{S_M}\, \|g - f\|_{L_\infty(\Omega)},
\end{aligned}$$

where we used $\|h\|_{L_2} \le \sqrt{\mu(\Omega)}\|h\|_{L_\infty}$ for any $h \in C(\Omega)$, and applied (16) to $g - f$. Since g is arbitrary, taking the infimum gives (17). □

Theorem 7 shows that the hyperinterpolant is stable with the $L_\infty \to L_2$ operator norm bounded by $\sqrt{S_M} \approx \sqrt{\mu(\Omega)}$, and the L_2 error is at most $2\sqrt{\mu(\Omega)}$ times the best L_∞ approximation from V_n. Notice that the bound (16) controls the $L_\infty \to L_2$ operator norm, with constant $\sqrt{S_M}$ independent of n. This is a different statement from the boundedness of the classical Lebesgue constant. We now analyze this Lebesgue constant directly.

Recalling the definition of hyperinterpolation (9), we write the hyperinterpolant as a linear functional of the sampled data as:

$$\mathcal{L}_{V,n}f(\mathbf{x}) = \sum_{j=1}^{n}\left(\sum_{k=1}^{M} w_k f(\mathbf{z}_k) v_j(\mathbf{z}_k)\right) v_j(\mathbf{x}) = \sum_{k=1}^{M} \underbrace{w_k\, K_n(\mathbf{x}, \mathbf{z}_k)}_{=:\, h_k(\mathbf{x})}\, f(\mathbf{z}_k),$$

where

$$K_n(\mathbf{x}, \mathbf{y}) := \sum_{j=1}^{n} v_j(\mathbf{x})\, v_j(\mathbf{y})$$

is the reproducing kernel of V_n in $L_2(\Omega, \omega)$, which means

$$\int_\Omega K_n(\mathbf{x}, \mathbf{y})\, u(y)\, d\omega(\mathbf{y}) = u(\mathbf{x}),$$

in particular,

$$\int_\Omega K_n(\mathbf{x}, \mathbf{y})\, K_n(\mathbf{z}, \mathbf{y})\, d\omega(y) = K_n(x, z).$$

We also note that $h_k(\mathbf{x}) = w_k K_n(\mathbf{x}, \mathbf{z}_k)$ are the Lagrange functions, which means that when $M = n$, the hyperinterpolation is equivalent to interpolation, $h_k(\mathbf{x})$ becomes the cardinal function. Using the Lagrange representation of $L_{V,n}f$, we have:

$$\|L_{V,n}f\|_{L_\infty(\Omega)} \leq \Lambda(Z, W)\|f\|_{\ell_\infty(Z)},$$

where the Lebesgue constant is defined as:

$$\Lambda(Z, W) := \sup_{\mathbf{x}\in\Omega} \sum_{k=1}^{M} |h_k(\mathbf{x})| = \sup_{\mathbf{x}\in\Omega} \sum_{k=1}^{M} w_k \, |K_n(\mathbf{x}, \mathbf{z}_k)|. \tag{18}$$

Proposition 8 (Lebesgue constant of hyperinterpolation) *Under the assumptions of Theorem 7 (positive weights and exactness on V_n^2), we have*

$$\Lambda(Z, W) \;\leq\; \sqrt{S_M} \, \sup_{\mathbf{x}\in\Omega} \sqrt{K_n(\mathbf{x}, \mathbf{x})} \;=\; \sqrt{S_M} \, \sup_{\mathbf{x}\in\Omega} \left(\sum_{j=1}^{n} v_j(\mathbf{x})^2 \right)^{1/2}. \tag{19}$$

Proof For a fixed $\mathbf{x} \in \Omega$, equation (18) can be written as: $\Lambda(W, Z) = \sqrt{w_k} \cdot \big(\sqrt{w_k}\, |K_n(\mathbf{x}, \mathbf{z}_k)|\big)$. Applying the Cauchy–Schwarz inequality, we have

$$\sum_{k=1}^{M} w_k \, |K_n(\mathbf{x}, \mathbf{z}_k)| \leq \Big(\underbrace{\sum_{k=1}^{M} w_k}_{=S_M} \Big)^{1/2} \Big(\sum_{k=1}^{M} w_k \, K_n(\mathbf{x}, \mathbf{z}_k)^2 \Big)^{1/2}. \tag{20}$$

For the second factor, observe that with $\mathbf{x}$ fixed the function $K_n(\mathbf{x}, \mathbf{y}) = \sum_j v_j(\mathbf{x}) v_j(\mathbf{y})$ is a (fixed) element of V_n, hence its square $K_n(\mathbf{x}, \cdot)^2 \in V_n^2$. The exactness (8) therefore applies, and using the reproducing property of the kernel K_n,

$$\sum_{k=1}^{M} w_k \, K_n(\mathbf{x}, \mathbf{z}_k)^2 = \int_\Omega K_n(\mathbf{x}, \mathbf{y})^2 \, d\omega(\mathbf{y}) = K_n(\mathbf{x}, \mathbf{x}) = \sum_{j=1}^{n} v_j(\mathbf{x})^2. \tag{21}$$

Substituting (21) into (20) gives, for each $\mathbf{x}$, $\sum_k w_k |K_n(\mathbf{x}, \mathbf{z}_k)| \leq \sqrt{S_M} \, \sqrt{K_n(\mathbf{x}, \mathbf{x})}$; taking sup over $\mathbf{x}$ concludes the proof. $\square$

Taking into account the L_2-orthonormality of the basis $\{v_j\}_1^n$, the term $\sup_{\mathbf{x}\in\Omega} \sqrt{K_n(\mathbf{x}, \mathbf{x})}$ in (19), can be bound from below by $\sqrt{n/\mu(\Omega)}$, because

$$\int_\Omega K_n(\mathbf{x}, \mathbf{x}) \, d\omega(\mathbf{x}) = \sum_{j=1}^{n} \int_\Omega v_j(\mathbf{x})^2 \, d\omega(\mathbf{x}) = n.$$

It means that

$$\sqrt{\mu(\Omega)} \cdot \sqrt{\frac{n}{\mu(\Omega)}} = \sqrt{n} \leq \sqrt{S_M} \sup_{\mathbf{x}} \sqrt{K_n(\mathbf{x}, \mathbf{x})} \tag{22}$$

Inequality (22) gives a minimal value of the upper bound, not a lower bound on Λ itself. It expresses that the upper bound (19) on the Lebesgue constant of a hyperinterpolation operator cannot grow slower than $O(n^{1/2})$. However, we highlight that this is independent of the cubature points and weights, instead it depends only on the dimension of the Hilbert space V_n.

3.6 Hyperinterpolation under Relaxed Exactness Conditions

The requirement that the cubature formula be exact on V_n^2 is restrictive, as it typically requires $M = \mathcal{O}(n^2)$ cubature points. In this section, we consider a relaxation of this condition, following [23], which allows significantly reduced computational complexity.

Let $V_m \subset V_n$ be a subspace of dimension $m \leq n$ with a basis $\{v_1, \dots, v_m\}$. Instead of requiring exactness on V_n^2, we assume that the cubature rule (Z, W) is exact on

$$V_n \times V_m = \operatorname{span}\{v_i v_j : 1 \leq i \leq n,\, 1 \leq j \leq m\}.$$

That is,

$$\int_\Omega g(x)\, d\omega(x) = \sum_{k=1}^M w_k g(z_k) \quad \forall g \in V_n \times V_m.$$

The definition of the hyperinterpolant remains unchanged:

$$L_{V,n} f = \sum_{j=1}^n \langle f, v_j \rangle_{\ell_w^2} v_j, \quad \langle f, v_j \rangle_{\ell_w^2} = \sum_{k=1}^M w_k f(z_k) v_j(z_k).$$

We remark that we adopt the same notation for $L_{V,n}$ and ℓ_w^2, even if we use a different quadrature rule. Under this relaxed exactness condition, the hyperinterpolant reproduces all functions in V_m.

Lemma 9 If $f \in V_m$, then $L_{V,n} f = f$.

Proof Since $f \in V_m \subset V_n$, write $f = \sum_{i=1}^m a_i v_i$. Then

$$L_{V,n} f = \sum_{j=1}^n \langle f, v_j \rangle_{\ell_w^2} v_j = \sum_{j=1}^n \sum_{i=1}^m a_i \langle v_i, v_j \rangle_{\ell_w^2} v_j.$$

For $1 \leq i \leq m$ and $1 \leq j \leq n$ we have $v_i v_j \in V_n \times V_m$, hence $\langle v_i, v_j \rangle_{\ell_w^2} = \langle v_i, v_j \rangle_{L_2} = \delta_{ij}$, which gives $\mathcal{L}_{V,n} f = f$. □

In contrast to the fully exact case, $\mathcal{L}_{V,n}$ is no longer an orthogonal projection onto V_n: for $g \in V_n \setminus V_m$ we generally have $\mathcal{L}_{V,n} g \neq g$. However, stability and approximation can be retained provided that the discrete and continuous L_2 norms are comparable throughout V_n. Following [23], we assume that (Z, W) satisfies a Marcinkiewicz–Zygmund (MZ) inequality [24, 25]: there exists $\eta \in [0, 1)$ such that

$$\left| \sum_{k=1}^M w_k\, \chi(z_k)^2 - \int_\Omega \chi^2\, \mathrm{d}\omega \right| \;\leq\; \eta \int_\Omega \chi^2\, \mathrm{d}\omega \qquad \forall\, \chi \in V_n. \tag{23}$$

The value $\eta = 0$ is admissible and corresponds to exactness on V_n^2, so that η measures the departure from full exactness. The following is the abstract counterpart of [23, Thm. 2.2].

Theorem 10 *Let the cubature rule (Z, W) have positive weights, be exact on $V_n \times V_m$ and on constants, so that $S_M = \sum_{k=1}^{M} w_k = \mu(\Omega)$, and satisfy (23) with $\eta \in [0, 1)$. Then, for all $f \in \mathcal{C}(\Omega)$,*

$$\|\mathcal{L}_{V,n} f\|_{L_2(\Omega)} \le \frac{\sqrt{\mu(\Omega)}}{\sqrt{1-\eta}} \|f\|_{L_\infty(\Omega)},$$

and

$$\|f - \mathcal{L}_{V,n} f\|_{L_2(\Omega)} \le \left(\frac{1}{\sqrt{1-\eta}} + 1\right) \sqrt{\mu(\Omega)} \inf_{g \in V_m} \|f - g\|_{L_\infty(\Omega)}.$$

Proof The proof of [23, Thm. 2.2] applies upon replacing $\mathbb{P}_n$, $\mathbb{P}_k$, $E_k(f)$ and V by V_n, V_m, $\inf_{g \in V_m} \|f - g\|_{L_\infty(\Omega)}$ and $\mu(\Omega)$, respectively. The only properties used there are the L_2-orthonormality of $\{v_j\}_{j=1}^{n}$, the fact that $\{v_j\}_{j=1}^{m}$ spans V_m, the exactness on $V_n \times V_m$, and (23) applied to $\chi = \mathcal{L}_{V,n} f - \mathcal{L}_{V,m} f \in V_n$. □

The relaxed exactness framework introduces a trade-off between computational efficiency and approximation error. On the one hand, the cubature requirement is significantly reduced, since the exactness is only imposed on $V_n \times V_m$, with $m \ll n$. On the other hand, the approximation error is now governed by the smaller space V_m rather than V_n.

Remark 5 In the extreme case where $V_m = \operatorname{span}\{1\}$, the cubature rule needs to be exact only in $(V_n \oplus 1) \times 1$, which drastically reduces the number of constraints. In this setting, the approximation error is controlled only by constants, i.e.,

$$\inf_{c \in \mathbb{R}} \|f - c\|_{L^\infty(\Omega)},$$

which may be a weak bound from a theoretical perspective. Nevertheless, this choice often provides a favorable balance between stability and computational cost in practical applications.

4 The Case of Strictly Positive Definite Kernels

In this section, we specialize the hyperinterpolation framework to kernel-based spaces by choosing $V_n = \mathcal{N}_K(X)$, the native space generated by a positive definite kernel $K : \Omega \times \Omega \to \mathbb{R}$ and data sites X. This allows us to derive explicit error bounds in terms of kernel smoothness and data geometry.

Recalling that requiring exactness on V_n^2 leads to a system with $\frac{(n+1)n}{2}$ constraints, i.e., quadratic growth in n, solving (11) may be computationally prohibitive for large n as the dimension of V_n is determined by the number of data sites. The relaxed condition significantly reduces this burden, making kernel-based hyperinterpolation a feasible and practical approach.

4.1 Kernel Hyperinterpolation under Full Exactness

Let $\{v_1, \ldots, v_n\}$ be an $L_2(\Omega)$-orthonormal basis of $\mathcal{N}_K(X)$, which is available for any strictly positive definite kernel [15, 16]. However, for the construction of the cubature formula, the basis need not be orthonormal, as discussed in Section 3.3, and we may work directly with the kernel translations $K(\cdot, \mathbf{x}_j)$. The products entering the exactness conditions are then, for radial kernels,

$$K(\mathbf{x}, \mathbf{x}_k)K(\mathbf{x}, \mathbf{x}_m) = \phi(\|\mathbf{x} - \mathbf{x}_k\|)\phi(\|\mathbf{x} - \mathbf{x}_m\|), \qquad k, m \in \{1, \ldots, n\}$$

which can be evaluated efficiently, while the right-hand side of (13) requires the moments

$$b_i = \int_\Omega K(\mathbf{x}, \mathbf{x}_k)K(\mathbf{x}, \mathbf{x}_m)\, d\omega(\mathbf{x}).$$

Single-kernel moments are known analytically for several kernels, but the products above are in general not, and must be computed numerically.

We now derive explicit error bounds that depend on kernel smoothness and data geometry. Let K be of smoothness order $\tau > d/2$ (ensuring $\mathcal{N}_K \subset C(\Omega)$ by Sobolev embedding), and the given data sites X are quasi-uniform with fill distance h_X. The following corollary is an immediate consequence of Theorem 7, where we simply replace the generic V_n with $\mathcal{N}_K(X)$.

Corollary 1 Let $S_M = \sum_{j=1}^{M} w_j$ and $\mu(\Omega) = \int_\Omega d\omega$. For $f \in C(\Omega)$ we have

$$\|L_{K,n} f\|_{L_2(\Omega)} \le \sqrt{S_M}\, \|f\|_{L_\infty(\Omega)},$$

and

$$\|f - L_{K,n} f\|_{L_2(\Omega)} \le \left(\sqrt{\mu(\Omega)} + \sqrt{S_M}\right) \inf_{g \in \mathcal{N}_K(X)} \|g - f\|_{L_\infty(\Omega)}.$$

If the cubature is exact for constants, then $S_M = \mu(\Omega)$ and

$$\|f - L_{K,n} f\|_{L_2(\Omega)} \le 2\sqrt{\mu(\Omega)} \inf_{g \in \mathcal{N}_K(X)} \|g - f\|_{L_\infty(\Omega)}. \tag{24}$$

Now, the key question is: how large is $\inf_{g \in \mathcal{N}_K(X)} \|g - f\|_{L_\infty}$? We answer using classical kernel approximation theory for $f \in \mathcal{N}_K(\Omega)$. We remark that here we are bounding an L_∞ optimal error with a corresponding interpolation error, which is clearly an upper bound but may be suboptimal. We come back to this fact in Remark 6.

The following error bound for the kernel approximation can be found in textbooks such as [2, 4]

Theorem 11 *There exist $h_0 > 0$ and $C = C(K, \Omega, d)$ such that, if $h_X \le h_0$, then for $f \in \mathcal{N}_K(\Omega)$ it holds*

$$\inf_{g \in \mathcal{N}_K(X)} \|f - g\|_{L_\infty(\Omega)} \le \|f - S_{X,K} f\|_{L_\infty(\Omega)} \le C h_X^{\tau - d/2} \|f\|_{\mathcal{N}_K(\Omega)}, \tag{25}$$

where $S_{X,K} : \mathcal{N}_K(\Omega) \to \mathcal{N}_K(X)$ is the standard kernel interpolant.

We remark that similar interpolation bounds, with appropriately changed exponents, are also available for certain classes of functions outside of the native space; see [26, 27]. With this bound, we can derive a concrete corollary of our main convergence theorem.

Corollary 2 According to the assumptions of Theorem 11, for $f \in \mathcal{N}_K(\Omega)$ the kernel-based hyperinterpolation operator satisfies

$$\|f - L_{K,n} f\|_{L_2(\Omega)} \le 2\sqrt{\mu(\Omega)} \cdot C h_X^{\tau - d/2} \|f\|_{\mathcal{N}_K(\Omega)}. \tag{26}$$

If the points X are quasi-uniform with $h_X = O(n^{-1/d})$, then

$$\|f - L_{K,n} f\|_{L_2(\Omega)} = O(n^{-\tau/d + 1/2}) \quad \text{as } n \to \infty. \tag{27}$$

Proof Substituting (25) into (24) gives (26). For quasi-uniform points, volume arguments give $n \asymp h_X^{-d}$, hence $h_X \asymp n^{-1/d}$ and

$$\|f - L_{K,n} f\|_{L_2} = O\left((n^{-1/d})^{\tau - d/2}\right) = O(n^{-(\tau - d/2)/d}) = O(n^{(-\tau/d + 1/2)}).$$

□

Remark 6 (Comparison with Standard Interpolation, optimality and possible improvements) The convergence rate (27) is inferior to the best known L_2 rate for plain interpolation with uniform points (see e.g. [26]), which scales asymptotically as

$$\|f - S_{X,K} f\|_{L_2} = O(n^{-\tau/d}),$$

namely, a factor $n^{-1/2}$ better than our estimate in Corollary 2. As anticipated, this discrepancy may be due to the fact that we estimate a best approximation error with an interpolation error, and this process may sacrifice exactly up to a factor $n^{-1/2}$ (see [28]). However, this is provable only for optimally chosen subspaces $V_n \subset \mathcal{N}_K(\Omega)$, and not necessarily for $V_n = \mathcal{N}_K(X)$ as discussed here. In summary, this issue and possible refinements of (27) deserve further investigation.

We also highlight that since the hyperinterpolation operator and its reproducing kernel depend only on the subspace $V_n = \mathcal{N}_K(X)$, Proposition 8 applies once K_n is expressed in the (non-orthonormal) translate basis. Writing $\mathbf{k}(\mathbf{x}) = (K(\mathbf{x}, \mathbf{x}_1), \cdots, K(\mathbf{x}, \mathbf{x}_n))^T$ and letting G be the $L_2(\Omega)$ Gram matrix of the translates,

$$G_{ij} = \left\langle K(\cdot, \mathbf{x}_i), K(\cdot, \mathbf{x}_j) \right\rangle_{L_2(\Omega)} = \int_\Omega K(\mathbf{x}, \mathbf{x}_i)\, K(\mathbf{x}, \mathbf{x}_j)\, d\omega(x)$$

the L_2-reproducing kernel of $\mathcal{N}_K(X)$ is $K_n(\mathbf{x}, \mathbf{y}) = \mathbf{k}(\mathbf{x})^T G^{-1} \mathbf{k}(\mathbf{y})$.

4.2 Kernel Hyperinterpolation under Relaxed Exactness

The results of the previous section are based on cubature formulas that are exact on $\mathcal{N}_K(X) \times \mathcal{N}_K(X)$, leading to quadratic growth in the number of constraints. We now extend the kernel-based hyperinterpolation framework to the relaxed exactness setting introduced in Section 3.6.

Let $V_n = \mathcal{N}_K(X)$ and let $V_m \subset V_n$ be a subspace of dimension $m \leq n$. Under the assumptions of Theorem 10, the hyperinterpolation operator $L_{K,n}$ satisfies the stability estimate

$$\|L_{K,n} f\|_{L^2(\Omega)} \leq \frac{\sqrt{\mu(\Omega)}}{\sqrt{1-\eta}} \|f\|_{L^\infty(\Omega)},$$

and the error bound;

$$\|f - L_{K,n} f\|_{L^2(\Omega)} \leq \left(\frac{1}{\sqrt{1-\eta}} + 1 \right) \sqrt{\mu(\Omega)} \inf_{g \in V_m} \|f - g\|_{L^\infty(\Omega)}.$$

In contrast to the fully exact case, where the approximation error is governed by space $\mathcal{N}_K(X)$, the relaxed formulation yields an error bound controlled by the smaller subspace V_m. Moreover, the flexibility in selecting V_m allows one to tune the method:

- If $V_m = \mathcal{N}_K(X)$, we recover the full exactness framework and the error bounds of Corollary 2.
- If V_m is chosen as a lower-dimensional subspace (e.g., generated by a subset of kernel translates), the cubature construction becomes significantly cheaper, while the approximation error depends on the richness of V_m.
- In the extreme case $V_m = \operatorname{span}\{1\}$, the method reduces to a stabilized least-squares-type approximation with minimal cubature constraints.

Using the kernel approximation estimate from Theorem 11, we obtain the following.

$$\inf_{g \in V_m} \|f - g\|_{L^\infty(\Omega)} \leq \inf_{g \in \mathcal{N}_K(X)} \|f - g\|_{L^\infty(\Omega)},$$

but in general the approximation rate depends on how well V_m approximates functions in $\mathcal{N}_K(\Omega)$. If V_m is chosen such that

$$\inf_{g \in V_m} \|f - g\|_{L^\infty(\Omega)} = O(h_X^{\tau - d/2}),$$

then the same convergence rate as in Corollary 2 is retained.

5 Numerical Results

In this section, we present numerical experiments designed to validate the theoretical properties of kernel-based hyperinterpolation developed in sections 3-(4). We consider various target functions:

1. $f_1 = \sin(2\pi\mathbf{x}) + 0.5\sin(4\pi\mathbf{x})$ defined on $\Omega = [0, 1]$
2. $f_2 = |(x - 0.3)|^{2.5}$ defined on $\Omega = [0, 1]$
3. The Runge function $f_3(x) = \frac{1}{1+25x^2}$ defined on $\Omega = [0, 1]$
4. The Franke function, see [29].

Throughout the experiments, we adopt the following setting.

1. **Regularization of the cubature system:** In solving (14), we perturb the matrix ARA^T by adding a small multiple of the identity:

$$ARA^T \leftarrow ARA^T + 10^{-10} \cdot \frac{\mathrm{trace}(ARA^T)}{M} I.$$

2. **SVD-based least squares:** To deal with (12), we slightly deviate from the algorithm 1 and solve the problem through the singular value decomposition of $\widetilde{V}_Z$. Writing $\widetilde{V}_Z = U\Sigma V^\top$ with singular values $\sigma_1 \geq \sigma_2 \geq \cdots$, we form the truncated pseudoinverse $\widetilde{V}_Z^+$ by discarding all singular values smaller than $10^{-12}\,\sigma_{\max}$, and set $a = \widetilde{V}_Z^+ \sqrt{W}\, y$.
3. **Computation of the Lebesgue constant.** For a fine evaluation grid X_{leb} there exists the matrix H with $(L_{V,n}f)(X_{\mathrm{leb}}) = H\, y$, and the Lebesgue constant is the discrete operator norm

$$\Lambda = \max_i \sum_k |H_{ik}| = \max_{x \in X_{\mathrm{leb}}} \sum_k |h_k(x)|.$$

We emphasize that H must be assembled by the same numerical operator used to compute the approximant. In all experiments X_{leb} is a uniform grid of 1000 points in $d = 1$ and a 50×50 tensor grid in $d = 2$.

Matérn kernels

All the numerical tests represented later are carried out using the Matérn kernels. In particular, we consider basic(B), linear(L), and quadratic(Q) Matérn as follows:

$$\begin{aligned}
\mathrm{BMatérn}(\nu = \tfrac{1}{2}): &\quad \phi(r) = \exp(-\varepsilon r),\\
\mathrm{LMatérn}(\nu = \tfrac{3}{2}): &\quad \phi(r) = (1 + \varepsilon r)\exp(-\varepsilon r),\\
\mathrm{QMatérn}(\nu = \tfrac{5}{2}): &\quad \phi(r) = \left(1 + \varepsilon r + \tfrac{(\varepsilon r)^2}{3}\right)\exp(-\varepsilon r),
\end{aligned}$$

where $r = \|\mathbf{x} - \mathbf{x}_j\|_2$ and $\varepsilon > 0$ is the shape parameter that controls the flatness or localization of the kernel.

Shape parameter selection

A common strategy, adopted throughout our numerical experiments, is to tie ε adaptively to the fill distance h_X through

$$\varepsilon = C \cdot h_X, \tag{28}$$

where $C > 0$ is a problem-dependent constant. This ensures that as the data becomes denser, the kernel does not become artificially localized relative to the point distribution, maintaining a stable approximation regime.

Numerical Tests

The numerical experiments are organized as follows:

- In Test 1 we consider 4 various methods to recover the target function, including Kernel interpolation on the given data f_X, Kernel least squares, i.e., translating kernels at the given data sites X, and sampling the target function at the cubature points Z, Kernel hyperinterpolation with CF (full) exactness on $(\mathcal{N}_X \oplus 1) \times (\mathcal{N}_X \oplus 1)$, and reduced exactness $(\mathcal{N}_X \oplus 1)$.
- In Tests 2–3 instead, we compare Polynomial hyperinterpolation of degrees $\{q = 3Q + 1; Q = 0, \dots, 16\}$ with kernel hyperinterpolation with reduced and full exactness of the cubature formula.
- In Test 4, we use the same methods as in Test 1. To investigate the behavior of the kernel hyperinterpolation when the number of data points is large, the kernel hyperinterpolation with full exactness is omitted since at large $n = 1000$ the exactness conditions amount to roughly 5×10^5 constraints which imposes high computational costs.

Polynomial hyperinterpolation is done by employing Legendre polynomials together with Gauss–Legendre quadrature rules generated by the MATLAB function **lgwt**, available on MathWorks.

5.1 Test 1

We consider Halton-distributed data sites X with cardinalities $n = [20, 30, 45, 60, 80, 105, 140, 180, 200]$ and $C = 500$ in (28). Figures 1-2 report, the root mean square error (RMSE), and the Lebesgue constant corresponding to the test function f_1.

The RMSE curves (Figure 1) show that all four methods achieve comparable accuracy and the same empirical convergence order, with the full and reduced-exactness variants practically coinciding: for these smooth, nearly flat kernels the additional full-exactness constraints are largely redundant, as anticipated by the numerical rank deficiency of the kernel product space. In line with Corollary 2, the smoother Quadratic Matérn kernel yields a steeper error decay than the Linear and Basic ones. The Lebesgue constants, Figure 2, grow for all methods at a rate compatible with the reference $O(n^{1/2})$.

5.2 Test 2

The target function f_2 is recovered with quasi-uniformly distributed data sites (Halton points) of the same cardinality as in Test 1. We select Linear and Quadratic Matérn kernels with the shape parameters constant 800, and 900 respectively. The results are reported in Figures 3-4.

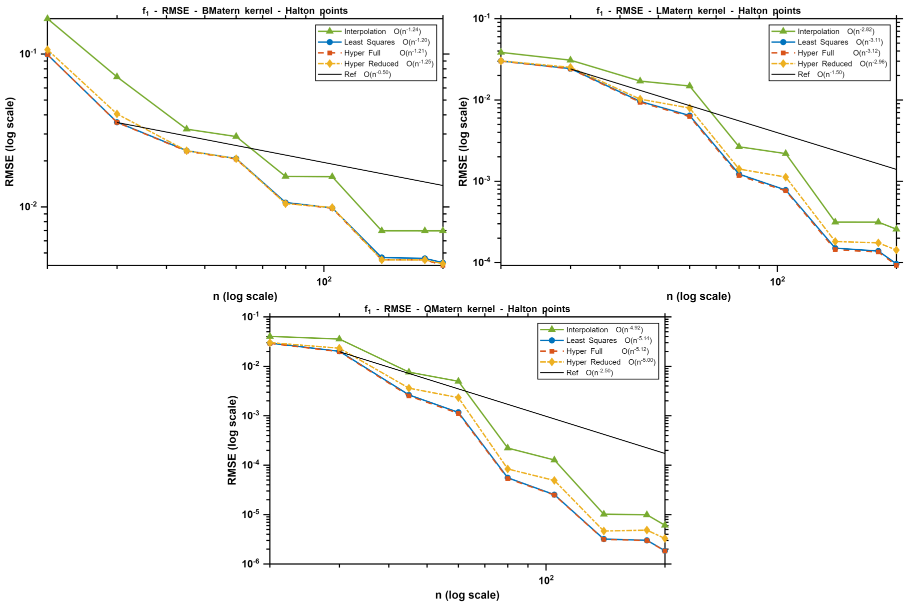

Fig. 1 RMSE comparison of various methods to recover f_1.

5.3 Test 3

To approximate f_3, we consider the data site X with a uniform distribution of the same size as Test 1. The shape parameter constant C is set to 500 and 700 for the Linear, and Quadratic Matérn kernels. The results are reported in Figures 5-6.

Figures 3 and 5 confirm that both polynomial and kernel hyperinterpolation achieve high accuracy and convergence order, the polynomial degree being capped at 50. Figures 4 and 6 show that the Lebesgue constant of polynomial hyperinterpolation grows in line with $O(n^{1/2})$, while the kernel variants grow at a comparable rate but with a larger and less regular constant; since Tests 2 and 3 differ in the node distribution, part of this irregularity is attributable to the node geometry rather than to the scheme itself.

5.4 Test 4

To recover f_4 we fix the data sites X to have cardinalities $n = [80, 105, 140, 180, 240, 300, 380, 460, 560, 680, 800, 1000]$ with Halton distribution. Similarly, linear and quadratic kernels Matérn are used with the shape parameter constant $C = 100$. The results are reported in Figures 7-8.

The RMSE curves (Figure 7) are comparable in the three methods and decay at an order consistent with the convergence rate of Corollary 2, again steeper for the smoother kernel. The Lebesgue constants, Figure 8, grow at a rate compatible with the reference $O(n^{1/2})$, confirming that the one-dimensional behavior persists at large n also in $d = 2$.

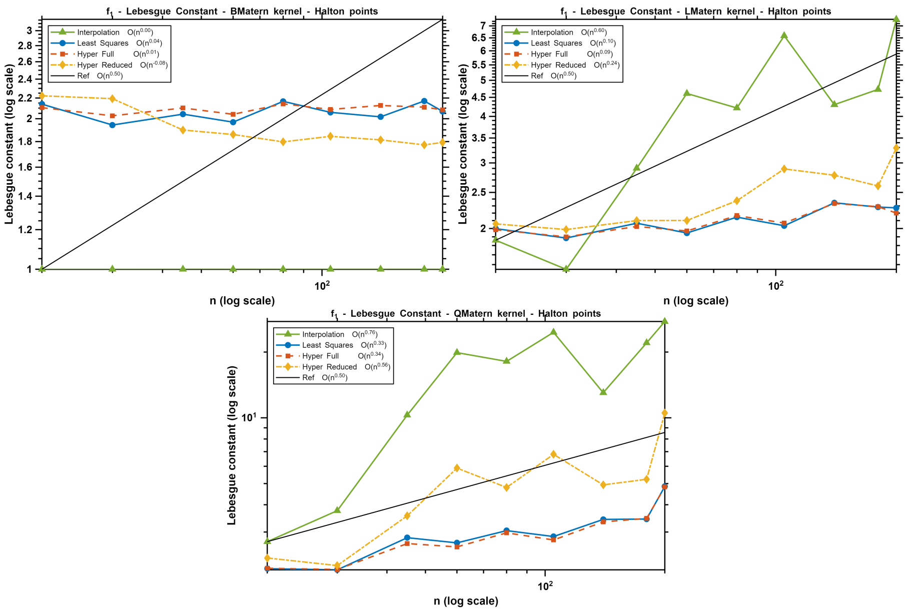

Fig. 2 Lebesgue constant comparison of various methods to recover f_1.

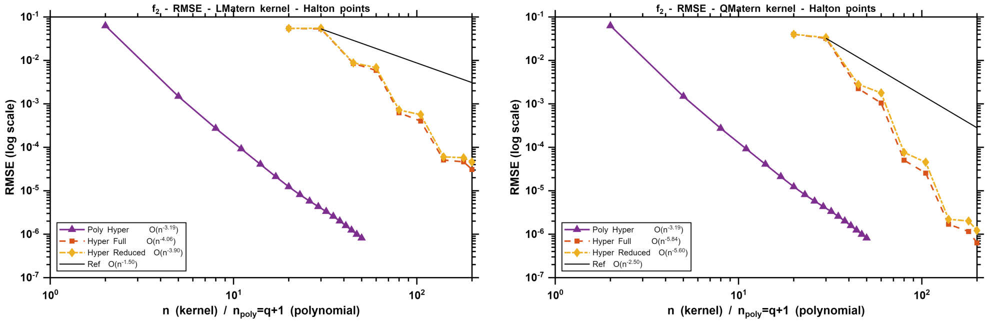

Fig. 3 RMSE comparison of various methods to recover f_2.

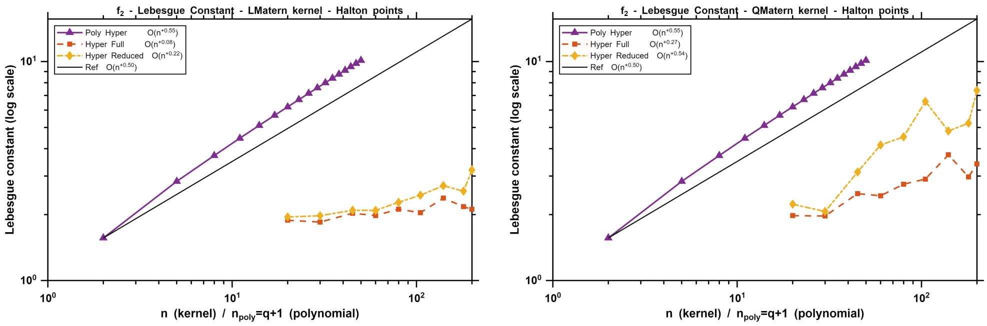

Fig. 4 Lebesgue constant comparison of various methods to recover f_2.

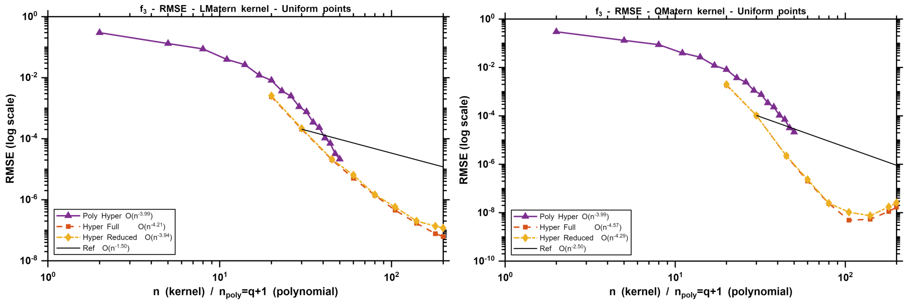

Fig. 5 RMSE comparison of various methods to recover f_3.

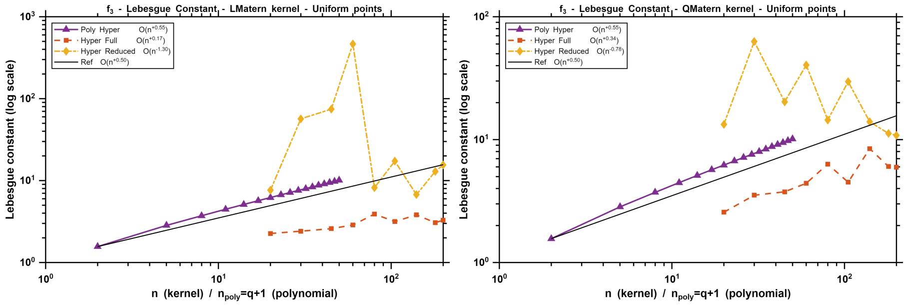

Fig. 6 Lebesgue constant comparison of various methods to recover f_3.

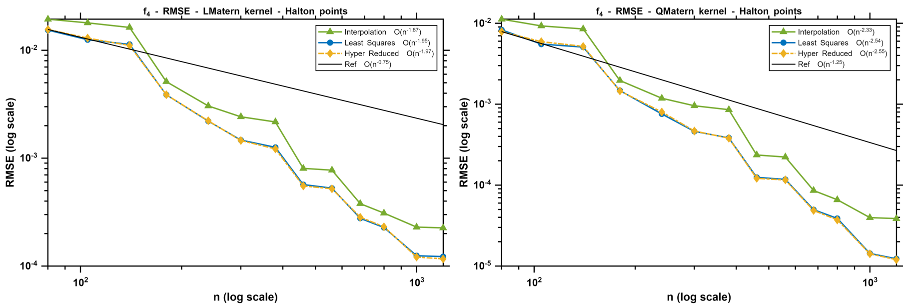

Fig. 7 RMSE comparison of various methods to recover f_4.

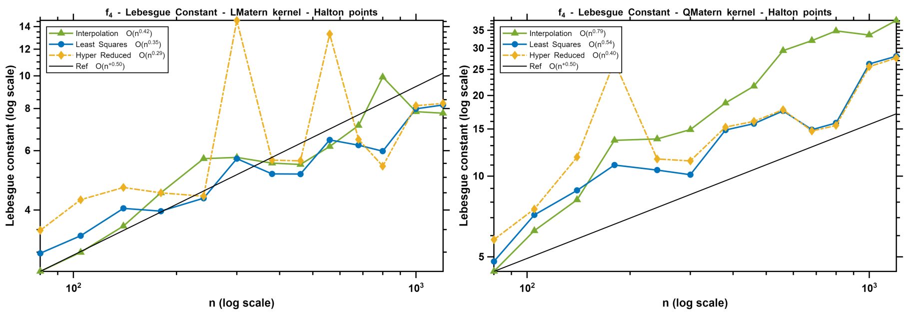

Fig. 8 Lebesgue constant comparison of various methods to recover f_4.

Remark 7 We conclude this section by emphasizing that the minimality of M sought by Algorithm 2 is not required by the theory. The constant S_M appearing in Theorem 7 equals $\mu(\Omega)$ for any exact positive-weight rule, so none of the bounds of Sections 3.5–3.6 degrades as M grows. The dominant cost is therefore not the hyperinterpolation operator itself but the rank-detection loop that locates the smallest admissible M: when evaluations of f are inexpensive, one may prescribe a large M and solve (14) once, bypassing the loop entirely. Although M and n grow together in all the tests above, preliminary experiments in the oversampled regime $M \gg n$ indicate a markedly slower growth of the Lebesgue constant and, for noisy samples, an error floor that decreases as M increases. A systematic study of this regime is left to future work.

6 Extension and Future Directions

The present framework applies directly to strictly PD kernels. A natural extension concerns conditionally positive definite (CPD) kernels, such as thin-plate splines, multiquadrics, and polyharmonic splines, for which the native space must be augmented with a polynomial space $\mathbb{P}^d_{m-1}$ and the interpolant includes a corresponding polynomial term subject to orthogonality constraints. Extending hyperinterpolation to this setting raises open questions, in particular on which product space cubature exactness should be imposed and how the polynomial constraints interact with the least-squares formulation; if exactness is required on $(\widetilde{\mathcal{N}}_K(X) + \mathbb{P}^d_{m-1}) \times (\widetilde{\mathcal{N}}_K(X) + \mathbb{P}^d_{m-1})$, Theorem (7) extends directly, but the number of required cubature points grows to $M = O((n + \dim \mathbb{P}^d_{m-1})^2)$.

Other directions include adaptive cubature schemes with computable error estimators, the use of sparse grids for high-dimensional problems, and Remark 7 discussed in Section 5.

7 Conclusion

This paper extends polynomial hyperinterpolation to arbitrary reproducing kernel Hilbert spaces. For any Hilbert subspace $V_n \subset C(\Omega)$ with an L_2-orthonormal basis and a positive-weight cubature exact on $V_n \times V_n$, the hyperinterpolation operator satisfies an $L_\infty \to L_2$ stability bound with a constant independent of n, and an L_2 error bound proportional to the best L_∞ approximation from V_n (Theorem 7); we construct such cubatures by adapting the Backus–Gilbert least-squares approach, with an existence result for equidistributed sequences. The hyperinterpolant is computed via a QR/SVD factorization of the weighted Gram matrix $\widetilde{V}_Z = \sqrt{W}\, V_Z$ in $O(Mn^2)$ operations ($O(n^3)$ when $M = O(n)$), avoiding normal equations.

In summary, the contribution of kernel-based hyperinterpolation is as follows. Polynomial hyperinterpolation requires orthogonal bases and exact cubature rules that are available only on specific geometries; the kernel construction removes this restriction and applies to arbitrary reproducing kernel Hilbert spaces and to scattered data on general domains, while retaining the projection structure and the $L_\infty \to L_2$ stability of the classical theory. On the other hand, with the kernel-translate basis, the Lebesgue growth match those of standard kernel interpolation, and the L_2 rate of Corollary 2 falls a factor $n^{-1/2}$ short of the interpolation rate. Whether this gap is intrinsic or an artifact of bounding a best-approximation error by an interpolation error, as discussed in Remark 6, remains open.

Declarations

- Funding: No funding was received for conducting this study.
- Competing interests: The authors have no competing interests to declare that are relevant to the content of this article.
- Code availability: The code used to generate the findings of this study is openly available in Github.

References

[1] Buhmann, M.D.: Radial Basis Functions: Theory and Implementations. Cambridge Monographs on Applied and Computational Mathematics, vol. 12. Cambridge University Press, Cambridge (2003)

[2] Fasshauer, G.E., McCourt, M.J.: Kernel-Based Approximation Methods Using MATLAB, 1st edn. Interdisciplinary Mathematical Sciences, vol. 19. World Scientific, Hackensack, NJ (2015)

[3] Schölkopf, B., Smola, A.J.: Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA (2002)

[4] Wendland, H.: Scattered Data Approximation. Cambridge Monographs on Applied and Computational Mathematics, vol. 17. Cambridge University Press, Cambridge (2005)

[5] De Marchi, S., Schaback, R.: Stability of kernel-based interpolation. Adv. Comput. Math. **32**, 155–161 (2010) https://doi.org/10.1007/s10444-008-9093-4

[6] Wendland, H., Rieger, C.: Approximate interpolation with applications to selecting smoothing parameters. Numer. Math. **101**(4), 729–748 (2005) https://doi.org/10.1007/s00211-005-0637-y

[7] Wendland, H.: Fast evaluation of radial basis functions: methods based on partition of unity. In: Approximation Theory, X (St. Louis, MO, 2001), pp. 473–483. Vanderbilt University Press, Nashville, TN (2002)

[8] Fasshauer, G.E., McCourt, M.J.: Stable evaluation of Gaussian radial basis function interpolants. SIAM J. Sci. Comput. **34**(2), 737–762 (2012) https://doi.org/10.1137/110824784

[9] Fornberg, B., Larsson, E., Flyer, N.: Stable computations with Gaussian radial basis functions. SIAM J. Sci. Comput. **33**(2), 869–892 (2011) https://doi.org/10.1137/09076756X

[10] Sloan, I.H.: Polynomial interpolation and hyperinterpolation over general regions. J. Approx. Theory **83**(2), 238–254 (1995) https://doi.org/10.1006/jath.1995.1119

[11] Schaback, R.: Error estimates and condition numbers for radial basis function interpolation. Adv. Comput. Math. **3**(3), 251–264 (1995) https://doi.org/10.1007/BF02432002

[12] Steinwart, I., Scovel, C.: Mercer's theorem on general domains: On the interaction between measures, kernels, and RKHSs. Constr. Approx. **35**(3), 363–417 (2012) https://doi.org/10.1007/s00365-012-9153-3

[13] Fasshauer, G.E.: Meshfree Approximation Methods with MATLAB. Interdisciplinary Mathematical Sciences, vol. 6. World Scientific, Hackensack, NJ (2007)

[14] Santin, G., Schaback, R.: Approximation of eigenfunctions in kernel-based spaces. Adv. Comput. Math. **42**(4), 973–993 (2016) https://doi.org/10.1007/s10444-015-9449-5

[15] Pazouki, M., Schaback, R.: Bases for kernel-based spaces. J. Comput. Appl. Math. **236**(4), 575–588 (2011) https://doi.org/10.1016/j.cam.2011.05.021

[16] Mohammadi, M., De Marchi, S., Esfahani, M.K.: Full-rank orthonormal bases for conditionally positive definite kernel-based spaces. J. Comput. Appl. Math. **444**, 115761 (2024) https://doi.org/10.1016/j.cam.2024.115761

[17] Tchakaloff, V.: Formules de cubatures mécaniques à coefficients non négatifs. Bull. Sci. Math **81**(2), 123–134 (1957)

[18] Bannai, E.: On tight spherical designs. J. Comb. Theory Ser. A **26**(1), 38–47 (1979)

[19] Hesse, K., Sloan, I.: Hyperinterpolation on the sphere. PURE AND APPLIED MATHEMATICS-MARCEL DEKKER INCORPORATED- **282**, 213 (2006) https://doi.org/10.1201/9781420011388.ch11

[20] Glaubitz, J.: Construction and application of provable positive and exact cubature formulas. IMA J. Numer. Anal. **43**(3), 1616–1652 (2023) https://doi.org/10.1093/imanum/drac017

[21] Glaubitz, J., Reeger, J.A.: Towards stability results for global radial basis function based quadrature formulas. BIT Numer. Math. **63**(1), 6 (2023) https://doi.org/10.1007/s10543-023-00956-0

[22] Backus, G., Gilbert, F.: The resolving power of gross earth data. Geophys. J. Int. **16**(2), 169–205 (1968) https://doi.org/10.1111/j.1365-246X.1968.tb00216.x.

[23] An, C., Wu, H.-N.: On the quadrature exactness in hyperinterpolation. BIT Numer. Math. **62**(4), 1899–1919 (2022) https://doi.org/10.1007/

s10543-022-00935-x

[24] Filbir, F., Mhaskar, H.N.: Marcinkiewicz–Zygmund measures on manifolds. J. Complexity **27**(6), 568–596 (2011) https://doi.org/10.1016/j.jco.2011.05.001

[25] Mhaskar, H., Narcowich, F., Ward, J.: Spherical Marcinkiewicz-Zygmund inequalities and positive quadrature. Math. Comput. **70**(235), 1113–1130 (2001) https://doi.org/10.1090/S0025-5718-00-01240-0

[26] Narcowich, F.J., Ward, J.D., Wendland, H.: Sobolev bounds on functions with scattered zeros, with applications to radial basis function surface fitting. Math. Comput. **74**(250), 743–763 (2005) https://doi.org/10.1090/S0025-5718-04-01708-9.

[27] Narcowich, F.J., Ward, J.D., Wendland, H.: Sobolev error estimates and a Bernstein inequality for scattered data interpolation via radial basis functions. Constr. Approx. **24**(2), 175–186 (2006) https://doi.org/10.1007/s00365-005-0624-7

[28] Steinwart, I.: A short note on the comparison of interpolation widths, entropy numbers, and Kolmogorov widths. J. Approx. Theory **215**, 13–27 (2017) https://doi.org/10.1016/j.jat.2016.11.006

[29] Franke, R.: A critical comparison of some methods for interpolation of scattered data. Technical Report NPS-53-79-003, Naval Postgraduate School, Monterey, CA (1979)