On hyperinterpolation and its variants

Alvise Sommariva

Università degli Studi di Padova

A two days on Approximation Theory and Applications,
Conference in honor of Paolo Emilio Ricci on his 80th Birthday,

Rome (I), May 23-24, 2024

The purpose of this talk is to

- recall the concept of the classical hyperinterpolation and its basic properties;
- outline the main variants and their features;
- show some numerical examples and discuss the advantages of the techniques mentioned above;
- conclude by mentioning some new frontiers on this topic.

The concept of hyperinterpolation has been introduced by I. Sloan in the paper

Interpolation and hyperinterpolation over general regions

published in 1995 on *Journal of Approx. Theory*.

- It is essentially an orthogonal (Fourier-like) projection on polynomial spaces, w.r.t. the discrete measure associated with a positive algebraic quadrature formula, or in other words a weighted least-squares polynomial approximation at the quadrature nodes.
- It is called *hyperinterpolation*, since under special assumptions it corresponds to interpolation.

Let $\mathbb{P}_n^d(\Omega)$ be the vector space d-variate polynomials of total-degree not exceeding n, restricted to a compact set or manifold $\Omega \subset R^d$.

Given

- an orthonormal basis $\{p_j\}$, $1 \leq j \leq N_n = \dim(\mathbb{P}_n^d(\Omega))$ of $\mathbb{P}_n^d(\Omega)$ w.r.t. an absolutely continuous measure $d\mu$ on Ω,
- a quadrature formula exact for $\mathbb{P}_{2n}^d(\Omega)$ w.r.t. μ, with nodes $X = \{x_i\} \subset \Omega$ and positive weights w_i, $1 \leq i \leq M_{2n}$ with $M_{2n} \geq N_n$,

the discretized orthogonal projection (hyperinterpolation) of $f \in C(\Omega)$ of degree n is

$$(\mathcal{L}_n f)(x) \;=\; \sum_{j=1}^{M_{2n}} \langle f, p_j \rangle_{l_{2,\mathbf{w}}(X)} p_j(x) = \sum_{i=1}^{M_{2n}} w_i f(x_i) \sum_{j=1}^{N_n} p_j(x_i) p_j(x).$$

where

$$\langle f, g \rangle_{l_{2,\mathbf{w}}(X)} := \sum_{i=1}^{M_{2n}} w_i f(x_i) g(x_i).$$

I. H. Sloan proved the following results:

1. (LS) The polynomial $\mathcal{L}_n f \in \mathbb{P}_n^d(\Omega)$ is the solution to the LS problem

$$\min_{p \in \mathbb{P}_n^d(\Omega)} \frac{1}{2}\langle f, p\rangle_{l_{2,\mathrm{w}}} = \min_{p \in \mathbb{P}_n^d(\Omega)} \frac{1}{2}\sum_{j=1}^{M_{2n}} w_j (p(x_j) - f(x_j))^2, \qquad (1)$$

2. (Operator norm) Given $f \in C(\Omega)$, $\|\mathcal{L}_n f\|_2 \leq \sqrt{\mu(\Omega)}\|f\|_\infty$,
3. (Error estimate) If $f \in C(\Omega)$ then

$$\|\mathcal{L}_n f - f\|_2 \leq 2\sqrt{\mu(\Omega)} E_n(f, \Omega)$$

where $E_n(f, \Omega) = \min_{p \in P_n^d(\Omega)} \|f - p\|_\infty$.

Definition

An algebraic quadrature rule

$$\int_\Omega f(x) d\Omega \approx \sum_{i=1}^{M_{2n}} w_i f(x_i)$$

is minimal if it has algebraic degree of precision $2n$ and $M_{2n} = \dim(\mathbb{P}_n^d) := N_n$

Examples

A. Gaussian rules in the interval;

B. Angular equispaced rules in the circle;

C. No such rules in general on the sphere ($n \geq 3$).

4. (Interpolation) The classical interpolation formula

$$\mathcal{L}_n f(x_j) = f(x_j), \;\; j = 1, \ldots, N_n$$

   holds for arbitrary f if and only if the quadrature rule is minimal.
5. (Projection) If $p \in \mathbb{P}_n^d$ then $\mathcal{L}_n p \equiv p$.

Filtered hyperinterpolation was introduced by I.H.Sloan and R.S.Womersley on the unit sphere, in the paper published in 2012

Filtered hyperinterpolation: A constructive polynomial approximation on the sphere

In the introduction they wrote:

While there are many other ways (such as radial basis functions, spline functions on a triangular mesh, wavelets) of representing scalar physical quantities on a sphere, polynomials continue to play an important role.

Of course their resolving power is limited by the degree of the polynomial, but the principal complaint about polynomials is often that irregularities or discontinuities in the quantity being approximated lead to fringes or oscillations (Gibbs' phenomenon for the case of discontinuities).

A recognised way of reducing the effect of fringes is by filtering an initial polynomial approximation, that is by modifying the amplitude of selected Fourier modes

- They introduce a filter function $h \in C([0, +\infty))$ that satisfies

$$h(x) = \begin{cases} 1, & \text{for } x \in [0, 1/2], \\ 0, & \text{for } x \in [1, \infty). \end{cases}$$

- It is immediate to observe that depending on the behaviour in $[1/2, 1]$, one can define many filters.
- Example: choose as filter $h \in C([0, +\infty))$ the function

$$h(x) = \begin{cases} 1, & x \in [0, \frac{1}{2}], \\ \sin^2(\pi x), & x \in [\frac{1}{2}, 1] \\ 0, & \text{for } x \in [1, \infty). \end{cases} \tag{2}$$

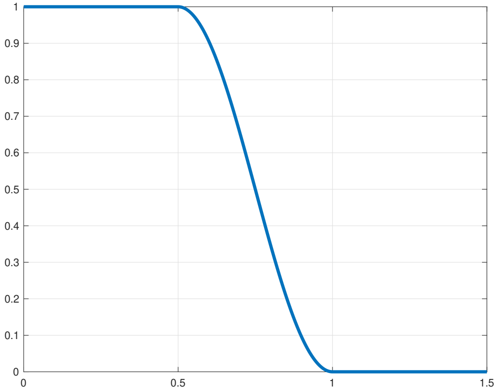

Figure: The filter h that is $\sin^2(\pi x)$, for $x \in [\frac{1}{2}, 1]$.

Filtered hyperinterpolation: definition

Definition

Consider the quadrature rule

$$\int_{\Omega} f(x) d\Omega \approx \sum_{j=1}^{M_\delta} w_j f(x_j)$$

and suppose that

- $w_j > 0$, $x_j \in \Omega$, $j = 1, \ldots, M_\delta$ (i.e. it is PI-type rule);
- it has algebraic degree of exactness $\delta = n - 1 + \lfloor n/2 \rfloor$.

Introduce the discrete scalar product determined by such formula

$$\langle f, g \rangle_{M_\delta} := \sum_{j=1}^{M_\delta} w_j f(x_j) g(x_j)$$

The filtered hyperinterpolant $\mathcal{F}_n f \in \mathbb{P}_{n-1}^d(\Omega)$ of $f \in C(\Omega)$ is defined as

$$\mathcal{F}_n f := \sum_{k=1}^{N_n} h\left(\frac{deg(p_k)}{n}\right) \langle f, p_k \rangle_{M_\delta}\, p_k \tag{3}$$

where $N_n = \dim(\mathbb{P}_n^d(\Omega))$, $\{p_k\}$ family of orthonormal polynomials*.

Remark

Since

$$\mathcal{F}_n f := \sum_{k=1}^{N_n} h\left(\frac{deg(p_k)}{n}\right) \langle f, p_k \rangle_{M_\delta} \, p_k \qquad (4)$$

and

$$h(x) = \begin{cases} 1, & \text{for } x \in [0, 1/2], \\ 0, & \text{for } x \in [1, \infty). \end{cases}$$

depending on $\overline{supp(h)} \subseteq [0,1]$ *one can achieve more sparsity in the polynomial coefficients w.r.t. classical hyperinterpolation, i.e. some less relevant discrete Fourier coefficients are dismissed.*

Remark (Projection)

Observe that $\mathcal{F}_n p = p$ *for all* $p \in \mathbb{P}^d_{\lfloor n/2 \rfloor}$.

An alternative to filtered hyperinterpolation is the so called Lasso hyperinterpolation based on the concept of Soft thresholding operator.

Definition (Soft thresholding operator)

The soft thresholding operator is defined as

$$\mathcal{S}_k(a) := \max(0, a-k) + \min(0, a+k),$$

where $k \geq 0$.

Alternatively, we can define $\mathcal{S}_k(a)$ as follows

$$\mathcal{S}_k(a) = \begin{cases} a+k, & \text{if } a < -k, \\ 0, & \text{if } -k \leq a \leq k, \\ a-k, & \text{if } a > k. \end{cases}$$

Lasso hyperinterpolation: definition

Definition

Consider the quadrature rule

$$\int_{\Omega} f(x)d\Omega \approx \sum_{j=1}^{M_{2n}} w_j f(x_j)$$

and suppose that

- $w_j > 0$, $x_j \in \Omega$, $j = 1, \ldots, M_{2n}$ (i.e. it is PI-type rule);
- it has algebraic degree of exactness $2n$.

Introduce the discrete scalar product determined by such formula

$$\langle f, g \rangle_{N_\delta} := \sum_{j=1}^{M_{2n}} w_j f(x_j) g(x_j)$$

Then the Lasso hyperinterpolation of f is defined as

$$\mathcal{L}_n^{\lambda} f := \sum_{j=1}^{N_n} \mathcal{S}_{\lambda \mu_j}(\langle f, p_j \rangle_{M_{2n}}) p_j, \tag{5}$$

where $\mathcal{S}_{\lambda \mu_j}$, $j = 1, \ldots, N_n$ are soft thresholding operators,

- $\lambda > 0$ is the regularization parameter,
- $\{\mu_j\}_{j=1,\ldots,N_n}$ is a set of positive penalty parameters.

Being

$$\mathcal{L}_n^{\lambda} f := \sum_{j=1}^{N_n} \mathcal{S}_{\lambda\mu_j}(\langle f, p_j \rangle_{N_\delta}) p_j, \tag{6}$$

the effect of the soft threshold operator

$$\mathcal{S}_k(a) = \begin{cases} a + k, & \text{if } a < -k, \\ 0, & \text{if } -k \leq a \leq k, \\ a - k, & \text{if } a > k. \end{cases}$$

is such that

$$|\langle f, p_j \rangle_{M_{2n}}| \leq \lambda\mu_j \Rightarrow \mathcal{S}_{\lambda\mu_j}(\langle f, p_j \rangle_{M_{2n}}) = 0$$

that is more sparsity in the polynomial coefficients is achieved (w.r.t. the classical hyperinterpolation).

Theorem

Let $f \in C(\Omega)$. Then

$$\min_{p \in \mathbb{P}_n(\Omega)} \left\{ \frac{1}{2} \sum_{j=1}^{M_{2n}} w_j (p(x_j) - f(x_j))^2 + \lambda \sum_{\ell=1}^{N_n} \mu_\ell |\gamma_\ell| \right\} \tag{7}$$

with

$$p(x) = \sum_{k=1}^{N_n} \gamma_k p_k(x) \in \mathbb{P}_n^d(\Omega).$$

Remark

Notice that

- *the first term in (7) is exactly that of classical hyperinterpolation,*
- *the second term in (7) is a penalization,*
- *issue: how to choose suitable λ, $\mu_1, \ldots, \mu_{N_n}$.*

Now we introduce hybrid interpolation that mixes the features of filtered and Lasso hyperinterpolation.

Definition (Hybrid hyperinterpolation)

Suppose that

- $\Omega \subset \mathbb{R}^d$ is a compact domain and $f \in \mathcal{C}(\Omega)$,
- $\langle f, g \rangle_{M_{2n}}$ is determined by an M_{2n}-point quadrature rule of PI-type in Ω with algebraic degree of exactness $2n$,
- h is a filter function,
- $\{\mathcal{S}_{\lambda\mu_\ell}\}_{\ell=1}^{N_n}$ are soft thresholding operators with $\lambda > 0$ and $\mu_1, \ldots, \mu_{N_n} > 0$

The *hybrid hyperinterpolation* of f onto $\mathbb{P}_n^d(\Omega)$ is defined as

$$\mathcal{H}_n^\lambda f := \sum_{k=1}^{N_n} h\left(\frac{\deg p_k}{n}\right) \mathcal{S}_{\lambda\mu_k}(\langle f, p_k \rangle_{M_{2n}}) p_k. \tag{8}$$

Consider the $\ell_2^2 + \ell_1$-regularized LS problem

$$\min_{p \in \mathbb{P}_n^d(\Omega)} \frac{1}{2} \sum_{j=1}^{N_n} w_j (p(x_j) - f(x_j))^2 + \frac{1}{2} \sum_{j=1}^{N} w_j (\mathbb{R}_n p(x_j))^2 + \lambda \sum_{\ell=1}^{d} \mu_\ell |\beta_\ell|, \tag{9}$$

where

- $\lambda > 0$, $\mu_1, \ldots, \mu_d > 0$,
- the operator $\mathbb{R}_n$ is defined as

$$\mathbb{R}_n p = \sum_{\ell=1}^{N_n} b_\ell \langle p_\ell, p \rangle_{M_n} p_\ell = \sum_{\ell=1}^{N_n} b_\ell \beta_\ell p_\ell \tag{10}$$

with

$$b_\ell = \begin{cases} 0, & \frac{\deg p_\ell}{n} \in [0, \frac{1}{2}], \\ \sqrt{\frac{1}{h_\ell} - 1}, & \frac{\deg p_\ell}{n} \in [\frac{1}{2}, 1). \end{cases} \tag{11}$$

Theorem

Let $f \in C(\Omega)$. Then $\mathbb{H}_L^{\lambda} f$ is the solution to the regularized least squares approximation problem

$$\min_{p \in \mathbb{P}_n^d(\Omega)} \frac{1}{2} \sum_{j=1}^{N_n} w_j (p(x_j) - f(x_j))^2 + \frac{1}{2} \sum_{j=1}^{N} w_j (\mathbb{R}_n p(x_j))^2 + \lambda \sum_{\ell=1}^{d} \mu_\ell |\beta_\ell|, \tag{12}$$

with

$$p(x) = \sum_{k=1}^{N_n} \gamma_k p_k(x) \in \mathbb{P}_n^d(\Omega).$$

There are technical difficulties.

- Low cardinality rules: for a fixed degree of precision n, minimal rules of PI-type exist in rare cases; alternatively one wants to determine those with the lowest number of nodes M_n or not exceeding the dimension N_n of the polynomial space $\mathbb{P}_n$.
- Computation of orthonormal basis: in some cases orthonormal basis w.r.t. a suitable weight function are available (e.g. interval, n-cube, sphere), but most of the times they are not available and one must determine them numerically.
- Computation of λ: in Lasso and Hybrid interpolation the performance depends on a parameter λ and it is not trivial to determine a good one (argument of future research).

The domain Ω is the union of M disks $B(C_k, r_k)$ with centers C_k and radii r_k, i.e.

$$\Omega = \cup_{k=1}^{M} B(C_k, r_k) \subset \mathbb{R}^2.$$

- low cardinality rules are available in which $N_n \leq M_n$;
- triangular polynomial orthonormal basis are computed numerically by a QR based algorithm.

In the tests we set $\Omega = \Omega^{(r_1)} \cup \Omega^{(r_2)}$, where $\Omega_2^{(r_j)}$, $j = 1, 2$, is the union of 19 disks, with

- centers $P_k^{(r_j)} = (r_j \cos(\theta_k), r_j \sin(\theta_k))$, with $r_1 = 2$ and $r_2 = 4$.
- $\theta_k = 2k\pi/19$, $k = 0, \ldots, 18$,
- radii equal to $r_j/4$, $j = 1, 2$.

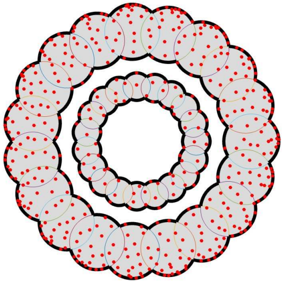

Figure: The domain $\Omega = \Omega^{(r_1)} \cup \Omega^{(r_2)}$ in which we perform our tests. We represent in red the $N = 496$ nodes of the cubature rule for algebraic degree of exactness=30, useful for classical hyperinterpolation at polynomial degree 15.

.

We consider contaminated evaluations of

$$f(\mathbf{x}) = (1 - (x^2 + y^2)) \exp(x \cos(y)),$$

contaminated by Gaussian noise ($\sigma = 0.05$).

Numerical tests show the favourable ratio between the sparsity and L_2 errors of the hybrid hyperinterpolants.

	λ	L_2 error	sparsity	λ	L_∞ error	sparsity
Hyperint.	–	0.045	136	–	0.1185	136
Filtered	–	0.0313	120	–	0.1043	120
Lasso	0.0059	0.0195	30	0.0072	0.0449	16
Hybrid	0.0048	0.0169	37	0.0068	0.0392	21

Table: Tests at degree 15 with λ choosen to minimize L_2 and L_∞ errors.

Some numerical experiments: union of disks

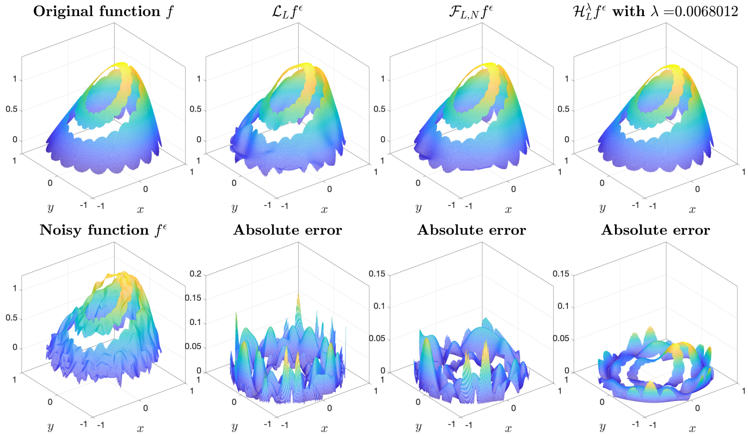

Figure: Approximate $f(x, y) = (1 - (x^2 + y^2)) \exp(x \cos(y))$, perturbed by Gaussian noise ($\sigma = 0.05$), over the union of disks, via hyperinterpolation $\mathcal{L}_L f^\epsilon$, filtered hyperinterpolation $\mathcal{F}_{L,N} f^\epsilon$, and hybrid hyperinterpolation $\mathcal{H}_L^\lambda f^\epsilon$ with $L = 15$ (different notation w.r.t. the talk).

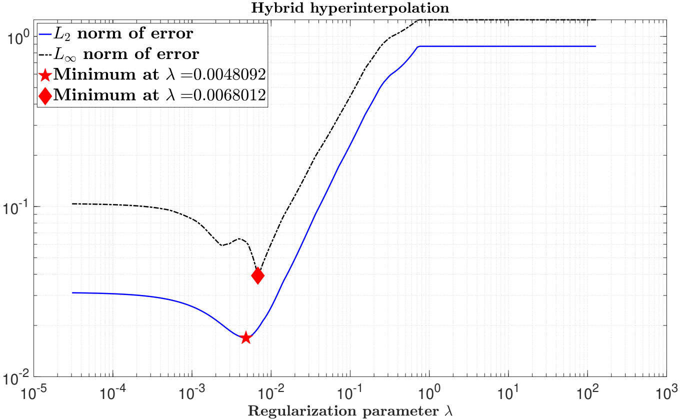

Figure: The choices of regularization parameter λ for hybrid hyperinterpolation $\mathcal{H}_n^\lambda f^\epsilon$ with $n = 15$ for approximating $f(x, y) = (1 - (x^2 + y^2)) \exp(x \cos(y))$, perturbed by Gaussian noise ($\sigma = 0.05$)).

Bibliography

- Hyperinterpolation seminal paper: I.H. Sloan, Interpolation and hyperinterpolation over general regions, J. Approx. Theory 83 (1995), 238–254.
- Filtered hyperinterpolation: I.H. Sloan, R.S. Womersley Filtered hyperinterpolation: A constructive polynomial approximation on the sphere, GEM - International Journal on Geomathematics, 3 (2012), 95–117.
- Lasso hyperinterpolation: C. An, H.-N. Wu, Lasso hyperinterpolation over general regions, SIAM J. Sci. Comput., 6 (2021), A3967–A3991.
- Hybrid hyperinterpolation: C. An, J. Ran, A. Sommariva, Hybrid hyperinterpolation over general regions, submitted,
- Details on hyperinterpolation implementation: A. Sommariva and M. Vianello, Numerical hyperinterpolation over nonstandard planar regions, Math. Comput. Simulation, 141 (2017), 110-120.
- Matlab software: A. Sommariva homepage, `https://www.math.unipd.it/~alvise/software.html`.