The Error of Multivariate Linear Extrapolation with Applications to Derivative-Free Optimization

Liyuan Cao, Zaiwen Wen

Peking University

SIAM OP 23, June 2, 2023

Motivation

COBYLA Michael JD Powell. "A direct search optimization method that models the objective and constraint functions by linear interpolation". In: *Advances in optimization and numerical analysis.* Springer, 1994, pp. 51–67

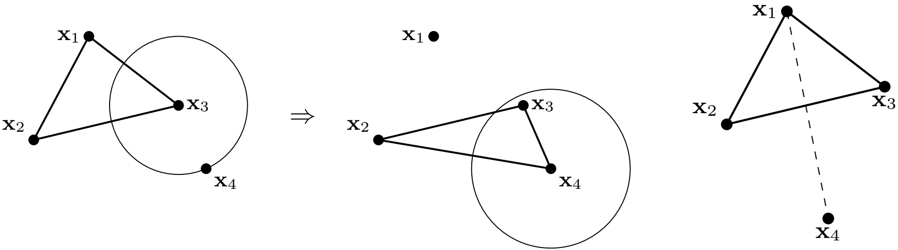
UOBYQA Michael JD Powell. "UOBYQA: unconstrained optimization by quadratic approximation". In: *Mathematical Programming* 92.3 (2002), pp. 555–582

NEWUOA Michael JD Powell. "The NEWUOA software for unconstrained optimization without derivatives". In: *Large-scale nonlinear optimization.* Springer, 2006, pp. 255–297

BOBYQA Michael JD Powell. "The BOBYQA algorithm for bound constrained optimization without derivatives". In: *Cambridge NA Report NA2009/06, University of Cambridge, Cambridge* 26 (2009)

LINCOA No paper, but please refer to Michael JD Powell. "On fast trust region methods for quadratic models with linear constraints". In: *Mathematical Programming Computation* 7.3 (2015), pp. 237–267

polynomial interpolation + trust region method



(a) linear interpolation + trust region method

(b) simplex method (Nelder-Mead)

Figure: An illustration of two DFO algorithm when minimizing a bivariate function, where $f(\mathbf{x}_1) > f(\mathbf{x}_2) > f(\mathbf{x}_3)$. The vertices of the triangles represents Θ. This figure only illustrates the algorithms' behavior when the trial point $\mathbf{x}_4$ satisfies $f(\mathbf{x}_4) < f(\mathbf{x}_3)$.

Problem Definition

For any function $f : \mathbb{R}^n \to \mathbb{R}$, given a set of $n+1$ affinely independent points $\Theta = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n+1}\} \subset \mathbb{R}^n$, a unique linear interpolation model can be constructed as $\hat{f}(\mathbf{x}) = c + \mathbf{g} \cdot \mathbf{x}$, where $c \in \mathbb{R}$ and $\mathbf{g} \in \mathbb{R}^n$ can be obtained by solving the linear system

$$\begin{bmatrix} 1 & \mathbf{x}_1^T \\ 1 & \mathbf{x}_2^T \\ & \vdots \\ 1 & \mathbf{x}_{n+1}^T \end{bmatrix} \begin{bmatrix} c \\ \mathbf{g} \end{bmatrix} = \begin{bmatrix} f(\mathbf{x}_1) \\ f(\mathbf{x}_2) \\ \vdots \\ f(\mathbf{x}_{n+1}) \end{bmatrix}.$$

Question: Assume $f \in C_\nu^{1,1}(\mathbb{R}^n)$, i.e.,

$$\|Df(\mathbf{u}) - Df(\mathbf{v})\| \le \nu \|\mathbf{u} - \mathbf{v}\| \text{ for all } \mathbf{u}, \mathbf{v} \in \mathbb{R}^n.$$

For some given Θ and $\mathbf{x}$, what is the (sharp) upper bound on the function approximation error $|\hat{f}(\mathbf{x}) - f(\mathbf{x})|$ at any given $\mathbf{x} \in \mathbb{R}^n$, particularly when $\mathbf{x} \notin \text{conv}(\Theta)$?

Existing Results

$\mathcal{P}_n^d$: the set of all dth-order polynomials on $\mathbb{R}^n$.
p: number of coefficients in a polynomial in $\mathcal{P}_n^d$.

1. **classical univariate interpolation error**
   When a univariate function f is interpolated by $\hat{f} \in \mathcal{P}_1^d$ on $\Theta = \{x_1, \ldots, x_{d+1}\} \subset \mathbb{R}$,
   $$f(x) - \hat{f}(x) = \frac{(x - x_1)(x - x_2) \cdots (x - x_{d+1})}{(d+1)!} D^{d+1} f(\xi) \quad \text{for any } x \in \mathbb{R}$$
   for some ξ with $\min(x, x_1, , \ldots, x_{d+1}) < \xi < \max(x, x_1, \ldots, x_{d+1})$.
2. **seminal work on interpolation error:** Philippe G Ciarlet and Pierre-Arnaud Raviart. "General Lagrange and Hermite interpolation in $\mathbb{R}^n$ with applications to finite element methods". In: *Archive for Rational Mechanics and Analysis* 46.3 (1972), pp. 177–199

Theorem (error of general Lagrange interpolation)

If f has ν_d-Lipschitz continuous derivative and is interpolated by $\hat{f} \in \mathcal{P}_n^d$, then

$$D^m \hat{f}(\mathbf{x}) - D^m f(\mathbf{x}) = \frac{1}{(d+1)!} \sum_{i=1}^{p} \left\{ D^{d+1} f(\xi_i) \cdot (\mathbf{x}_i - \mathbf{x})^{d+1} \right\} D^m \ell_i(\mathbf{x}),$$

where $\xi_i = \alpha_i \mathbf{x}_i + (1 - \alpha_i)\mathbf{x}$.

Existing Results

3. **sharp bound on linear interpolation:** Shayne Waldron. “The error in linear interpolation at the vertices of a simplex”. In: *SIAM Journal on Numerical Analysis* 35.3 (1998), pp. 1191–1200

Theorem (sharp bound on linear interpolation)

Let $\mathbf{c}$ *be the center and* R *the radius of the unique sphere containing* Θ*. Then, for each* $\mathbf{x} \in conv(\Theta)$*, there is the sharp inequality*

$$|\hat{f}(\mathbf{x}) - f(\mathbf{x})| \leq \frac{1}{2}\left(R^2 - \|\mathbf{x} - \mathbf{c}\|^2\right) |||D^2 f|||_{L_\infty(\mathrm{conv}(\Theta))}.$$

4. **bound on quadratic interpolation error:** MJD Powell. “On the Lagrange functions of quadratic models that are defined by interpolation”. In: *Optimization Methods and Software* 16.1-4 (2001), pp. 289–309

Theorem (bound on quadratic interpolation)

Assume $f \in C_M^{3,2}(\mathbb{R}^n)$*. Let* $\hat{f} \in \mathcal{P}_n^2$ *interpolate* f *on* Θ*. Then for every* $\mathbf{x} \in \mathbb{R}^n$*,*

$$|\hat{f}(\mathbf{x}) - f(\mathbf{x})| \leq \frac{M}{6} \sum_{i=1}^{p} |\ell_i(\mathbf{x})| \|\mathbf{x}_i - \mathbf{x}\|^3.$$

Definition (Lagrange Polynomial)

For a given set of $\Theta = \{\mathbf{x}_i\}_{i=1}^{n+1} \subset \mathbb{R}^n$, a set of $n+1$ linear functions $\{\ell_j\}_{j=1}^{n+1}$ is called a basis of Lagrange polynomials if

$$\ell_j(\mathbf{x}_i) = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

The sequence $\ell_1(\mathbf{x}), \ldots, \ell_{n+1}(\mathbf{x})$ also coincides with the sequence of Barycentric coordinates of $\mathbf{x}$ w.r.t. Θ. They have the following properties:

$$\sum_{i=1}^{n+1} \ell_i(\mathbf{x}) f(\mathbf{x}_i) = \hat{f}(\mathbf{x}),$$

$$\sum_{i=1}^{n+1} \ell_i(\mathbf{x}) = 1,$$

$$\text{and } \sum_{i=0}^{n+1} \ell_i(\mathbf{x})\mathbf{x}_i = \mathbf{0}.$$

Error Estimation Problem

Because

the sharp upper bound on error = the largest possible error,

the question can be formulated as

$$\max_f |\hat{f}(\mathbf{x}) - f(\mathbf{x})| \quad \text{s.t. } f \in C_\nu^{1,1}(\mathbb{R}^n).$$

Error Estimation Problem

Because

$$\text{the sharp upper bound on error} = \text{the largest possible error},$$

the question can be formulated as

$$\max_f |\hat{f}(\mathbf{x}) - f(\mathbf{x})| \quad \text{s.t. } f \in C_\nu^{1,1}(\mathbb{R}^n).$$

Because

1. $C_\nu^{1,1}(\mathbb{R}^n)$ is symmetric
2. and $\sum_{i=1}^{n+1} \ell_i(\mathbf{x}) f(\mathbf{x}_i) = \hat{f}(\mathbf{x})$,

the problem is equivalent to

$$\max_f \sum_{i=1}^{n+1} \ell_i(\mathbf{x}) f(\mathbf{x}_i) - f(\mathbf{x}) \quad \text{s.t. } f \in C_\nu^{1,1}(\mathbb{R}^n).$$

Error Estimation Problem

$$\max_f \sum_{i=1}^{n+1} \ell_i(\mathbf{x}) f(\mathbf{x}_i) - f(\mathbf{x}) \quad \text{s.t. } f \in C_L^{1,1}(\mathbb{R}^n).$$

Notice when $\Theta = \{\mathbf{x}_1, \ldots, \mathbf{x}_{n+1}\}$ and $\mathbf{x}$ are given, all $\ell_i(\mathbf{x})$ are fixed.

First approach to handle the functional constraint:

$$\begin{aligned} \max_{\mathbf{g}_i, y_i} \quad & \sum_{i=1}^{n+1} \ell_i(\mathbf{x}) y_i - y_0 \\ \text{s.t.} \quad & y_j \leq y_i + \mathbf{g}_i \cdot (\mathbf{x}_j - \mathbf{x}_i) + \frac{\nu}{2} \|\mathbf{x}_j - \mathbf{x}_i\|^2 \ \forall i, j = 0, 1, \ldots, n+1. \end{aligned}$$

Second approach to handle the functional constraint:

$$\begin{aligned} \max_{\mathbf{g}_i, y_i} \quad & \sum_{i=1}^{n+1} \ell_i(\mathbf{x}) y_i - y_0 \\ \text{s.t.} \quad & y_j \leq y_i + \frac{1}{2}(\mathbf{g}_i + \mathbf{g}_j) \cdot (\mathbf{x}_j - \mathbf{x}_i) + \frac{\nu}{4} \|\mathbf{x}_j - \mathbf{x}_i\|^2 \\ & \qquad - \frac{1}{4\nu} \|\mathbf{g}_j - \mathbf{g}_i\|^2 \ \forall i, j = 0, 1, \ldots, n+1, \end{aligned}$$

where $(\mathbf{x}_0, \mathbf{g}_0, y_0)$ represents $(\mathbf{x}, Df(\mathbf{x}), f(\mathbf{x}))$.

Performance Estimation Problem

- Adrien B Taylor, Julien M Hendrickx, and François Glineur. "Smooth strongly convex interpolation and exact worst-case performance of first-order methods". In: *Mathematical Programming* 161 (2017), pp. 307–345
- Adrien B Taylor, Julien M Hendrickx, and François Glineur. "Exact worst-case performance of first-order methods for composite convex optimization". In: *SIAM Journal on Optimization* 27.3 (2017), pp. 1283–1313

Theorem (interpolation condition for $C_\nu^{1,1}(\mathbb{R}^n)$)

Let $\nu > 0$ and $\mathcal{I}$ be an index set, and consider a set of triples $\{(\mathbf{x}_i, \mathbf{g}_i, y_i)\}_{i\in\mathcal{I}}$ where $\mathbf{x}_i \in \mathbb{R}^n$, $\mathbf{g} \in \mathbb{R}^n$, and $y_i \in \mathbb{R}$ for all $i \in \mathcal{I}$. There exists a function $f \in C_\nu^{1,1}(\mathbb{R}^n)$ such that both $\mathbf{g}_i = Df(\mathbf{x}_i)$ and $y_i = f(\mathbf{x}_i)$ hold for all $i \in \mathcal{I}$ if and only if the following inequality holds for all $i, j \in \mathcal{I}$:

$$y_j \leq y_i + \frac{1}{2}(\mathbf{g}_i + \mathbf{g}_j) \cdot (\mathbf{x}_j - \mathbf{x}_i) + \frac{\nu}{4}\|\mathbf{x}_j - \mathbf{x}_i\|^2 - \frac{1}{4\nu}\|\mathbf{g}_j - \mathbf{g}_i\|^2.$$

Error Estimation Problem

The infinite dimensional problem

$$\max_f |\hat{f}(\mathbf{x}) - f(\mathbf{x})| \quad \text{s.t. } f \in C_L^{1,1}(\mathbb{R}^n).$$

is equivalent to the finite dimensional problem

$$\begin{aligned} \max_{\mathbf{g}_i, y_i} \quad & \sum_{i=1}^{n+1} \ell_i(\mathbf{x}) y_i - y_0 \\ \text{s.t.} \quad & y_j \leq y_i + \frac{1}{2}(\mathbf{g}_i + \mathbf{g}_j) \cdot (\mathbf{x}_j - \mathbf{x}_i) + \frac{\nu}{4}\|\mathbf{x}_j - \mathbf{x}_i\|^2 \\ & \qquad - \frac{1}{4\nu}\|\mathbf{g}_j - \mathbf{g}_i\|^2 \ \forall i, j = 0, 1, \ldots, n+1. \end{aligned}$$

Notice how $\mathbf{x} = \mathbf{x}_0$ is nothing special except its coefficient is fixed to -1?
Define

$$\ell_0(\mathbf{x}) = -1$$

and

$$\begin{aligned} \mathcal{I}_+ &= \{i \in \{0, \ldots, n+1\} : \ell_i(\mathbf{x}) > 0\} \\ \mathcal{I}_- &= \{i \in \{0, \ldots, n+1\} : \ell_i(\mathbf{x}) < 0\}. \end{aligned}$$

Error Estimation Problem

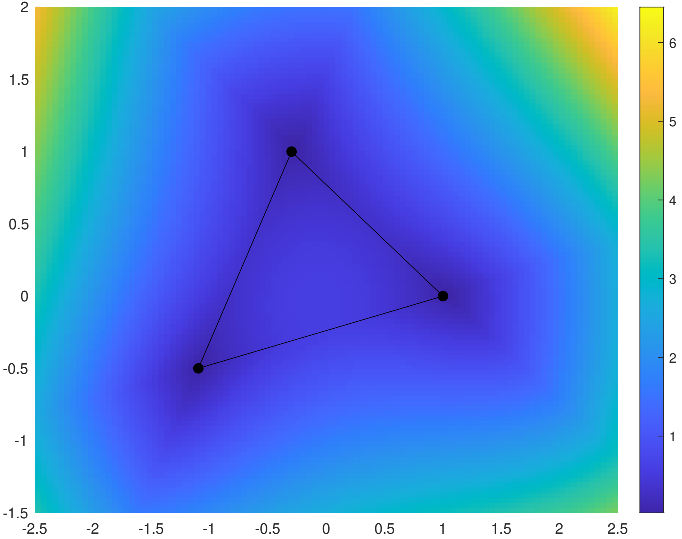


Figure: The sharp error bound on $|\hat{f}(\mathbf{x}) - f(\mathbf{x})|$ for each $\mathbf{x}$ on the 100×100 grid covering $[-2.5, 2.5] \times [-1.5, 2.5]$, where $\Theta = \{(-0.3, 1), (-1.1, -0.5), (1, 0)\}$ and $\nu = 1$.

An Improved Upper Bound

Theorem

Assume $f \in C^{1,1}_{\nu}(\mathbb{R}^n)$. *Let* $\hat{f} \in \mathcal{P}^1_n$ *interpolate* f *at* $\Theta = \{\mathbf{x}_1, \dots, \mathbf{x}_{n+1}\} \subset \mathbb{R}^n$. *Then*

$$\hat{f}(\mathbf{x}) - f(\mathbf{x}) \leq \frac{\nu}{2} \sum_{i=0}^{n+1} |\ell_i(\mathbf{x})| \|\mathbf{x}_i - \mathbf{u}\|^2 \text{ for any } \mathbf{u} \in \mathbb{R}^n.$$

Proof.

The bound is the weighted sum of the following inequalities

$$\begin{aligned}
&\ell_i(\mathbf{x}) \qquad & f(\mathbf{x}_i) - f(\mathbf{u}) - Df(\mathbf{u}) \cdot (\mathbf{x}_i - \mathbf{u}) &\leq \frac{\nu}{2}\|\mathbf{x}_i - \mathbf{u}\|^2 \qquad && \text{for all } i \in \mathcal{I}_+,\\
&-\ell_j(\mathbf{x}) \qquad & -f(\mathbf{x}_j) + f(\mathbf{u}) + Df(\mathbf{u}) \cdot (\mathbf{x}_j - \mathbf{u}) &\leq \frac{\nu}{2}\|\mathbf{x}_j - \mathbf{u}\|^2 \qquad && \text{for all } j \in \mathcal{I}_-.
\end{aligned}$$

- In existing results from the literature, $\mathbf{u} = \mathbf{x}$.
- The point $\mathbf{u}$ can be set to the center of a trust region.
- Minimize the R.H.S. w.r.t. $\mathbf{u}$ to yield

$$\mathbf{u}^{\star} = \mathbf{w} \overset{\text{def}}{=} \frac{\sum_{i=0}^{n+1} |\ell_i(\mathbf{x})| \mathbf{x}_i}{\sum_{i=0}^{n+1} |\ell_i(\mathbf{x})|}$$

An Improved Upper Bound: Sharpness

Theorem

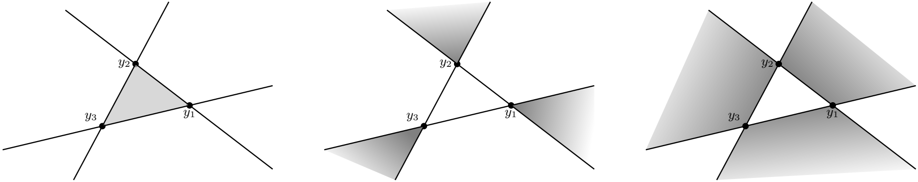
The bound $\hat{f}(\mathbf{x}) - f(\mathbf{x}) \leq \frac{\nu}{2} \sum_{i=0}^{n+1} |\ell_i(\mathbf{x})| \|\mathbf{x}_i - \mathbf{w}\|^2$ *is sharp under either of the two following conditions*

1. $\mathbf{x} \in conv(\Theta)$;
2. *there is only one positive term in* $\ell_1(\mathbf{x}), \ldots, \ell_{n+1}(\mathbf{x})$.

Proof.

This error can be achieved by the function

1. $f(\mathbf{x}) = \frac{\nu}{2}\|\mathbf{x}\|^2$ for the first case;
2. $f(\mathbf{x}) = -\frac{\nu}{2}\|\mathbf{x}\|^2$ for the second case.



(a) $\mathbf{x} \in \text{conv}(\Theta)$ **(b)** one positive ℓ **(c)** to be discussed

Maximizing Error over Quadratic Functions

Let f be a quadratic function of the form

$$f(\mathbf{u}) = c + \mathbf{g} \cdot \mathbf{u} + H\mathbf{u} \cdot \mathbf{u}/2 \text{ with } c \in \mathbb{R}, \mathbf{g} \in \mathbb{R}^n, \text{ and symmetric } H \in \mathbb{R}^{n \times n}.$$

The error estimation problem can be formulated as

$$\begin{array}{ll} \max\limits_{H} & \hat{f}(\mathbf{x}) - f(\mathbf{x}) = G \cdot H/2 \\ \text{s.t.} & -\nu I \preceq H \preceq \nu I, \end{array}$$

where

$$G = \sum_{i=0}^{n+1} \ell_i(\mathbf{x}) \mathbf{x}_i \mathbf{x}_i^T.$$

Maximizing Error over Quadratic Functions

Let f be a quadratic function of the form

$$f(\mathbf{u}) = c + \mathbf{g} \cdot \mathbf{u} + H\mathbf{u} \cdot \mathbf{u}/2 \text{ with } c \in \mathbb{R}, \mathbf{g} \in \mathbb{R}^n, \text{ and symmetric } H \in \mathbb{R}^{n \times n}.$$

The error estimation problem can be formulated as

$$\begin{array}{ll} \max\limits_{H} & \hat{f}(\mathbf{x}) - f(\mathbf{x}) = G \cdot H/2 \\ \text{s.t.} & -\nu I \preceq H \preceq \nu I, \end{array}$$

where

$$G = \sum_{i=0}^{n+1} \ell_i(\mathbf{x}) \mathbf{x}_i \mathbf{x}_i^T.$$

Solving analytically:

1. eigendecomposition: $G = P\Lambda P^T$: $\max_H (P\Lambda P^T) \cdot H/2$ s.t. $-\nu I \preceq H \preceq \nu I$;
2. change of variable: $\max_{P^THP} \Lambda \cdot (P^THP)/2$ s.t. $-\nu I \preceq P^THP \preceq \nu I$;
3. an optimal solution: $P^T H^\star P = \nu \text{sign}(\Lambda)$ or $H^\star = \nu P \text{sign}(\Lambda) P^T$.

Largest error achievable by quadratic function in $C_L^{1,1}(\mathbb{R}^n)$ is

$$G \cdot H^\star/2 = \frac{\nu}{2} \sum_{i=1}^{n} |\lambda_i|, \text{ where } \lambda_i\text{'s are the eigenvalues of } G.$$

Maximizing Error over Quadratic Functions

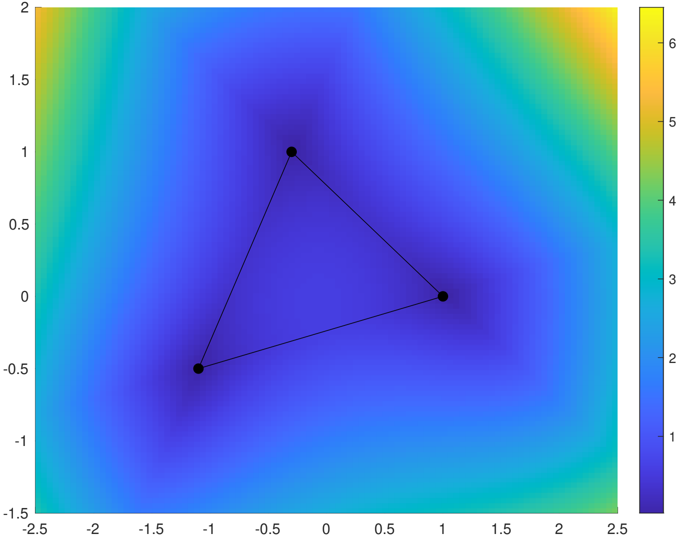
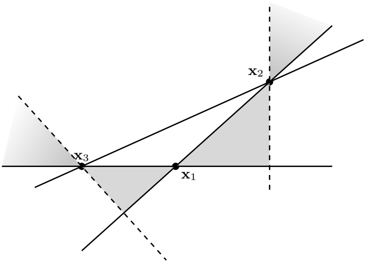


Figure: The sharp error bound on $|\hat{f}(\mathbf{x}) - f(\mathbf{x})|$ for each $\mathbf{x}$ on the 100×100 grid covering $[-2.5, 2.5] \times [-1.5, 2.5]$, where $\Theta = \{(-0.3, 1), (-1.1, -0.5), (1, 0)\}$ and $\nu = 1$.

Maximizing Error over Quadratic Functions



Mathematical description of the four open sets, from left to right:

1. $\ell_2(\mathbf{x}) > 0$ and $\ell_1(\mathbf{x})(\mathbf{x}_2 - \mathbf{x}_1) \cdot (\mathbf{x}_3 - \mathbf{x}_1) - \ell_2(\mathbf{x})(\mathbf{x}_3 - \mathbf{x}_2) \cdot (\mathbf{x}_1 - \mathbf{x}_2) > 0$
2. $\ell_3(\mathbf{x}) > 0, \ell_2(\mathbf{x}) < 0$, and
   $\ell_1(\mathbf{x})(\mathbf{x}_2 - \mathbf{x}_1) \cdot (\mathbf{x}_3 - \mathbf{x}_1) - \ell_2(\mathbf{x})(\mathbf{x}_3 - \mathbf{x}_2) \cdot (\mathbf{x}_1 - \mathbf{x}_2) < 0$
3. $\ell_2(\mathbf{x}) > 0, \ell_3(\mathbf{x}) < 0$, and
   $\ell_1(\mathbf{x})(\mathbf{x}_2 - \mathbf{x}_1) \cdot (\mathbf{x}_3 - \mathbf{x}_1) - \ell_3(\mathbf{x})(\mathbf{x}_2 - \mathbf{x}_3) \cdot (\mathbf{x}_1 - \mathbf{x}_3) < 0$
4. $\ell_3(\mathbf{x}) > 0$ and $\ell_1(\mathbf{x})(\mathbf{x}_2 - \mathbf{x}_1) \cdot (\mathbf{x}_3 - \mathbf{x}_1) - \ell_3(\mathbf{x})(\mathbf{x}_2 - \mathbf{x}_3) \cdot (\mathbf{x}_1 - \mathbf{x}_3) > 0$

There are up to 4 such open sets for bivariate extrapolation, but this number can be as large as 20 for trivariate extrapolation.

Maximizing Error over Quadratic Functions

Theorem (upper bound achieved by quadratic functions)

Assume $f \in C_L^{1,1}(\mathbb{R}^n)$. Let $\hat{f} \in \mathcal{P}_n^1$ interpolates f at any affinely independent $\Theta = \{\mathbf{x}_1, \ldots, \mathbf{x}_{n+1}\} \subset \mathbb{R}^n$. For any $\mathbf{x} \in \mathbb{R}^n$, if $\mu_{ij} \geq 0$ for all $(i,j) \in \mathcal{I}_+ \times \mathcal{I}_-$, then

$$|\hat{f}(\mathbf{x}) - f(\mathbf{x})| \leq \frac{1}{2} G \cdot H^\star = \frac{\nu}{2} \sum_{i=1}^{n} |\lambda_i|.$$

Computation of $\{\mu_{ij}\}$:

1.
$$Y_+ = \begin{bmatrix} \text{---}(\mathbf{x}_i - \mathbf{x})^T\text{---} \\ \vdots \\ \text{---}(\quad)^T\text{---} \end{bmatrix}_{i \in \mathcal{I}_+} \qquad Y_- = \begin{bmatrix} \text{---}(\mathbf{x}_j - \mathbf{x})^T\text{---} \\ \vdots \\ \text{---}(\quad)^T\text{---} \end{bmatrix}_{j \in \mathcal{I}_-}$$

$$\operatorname{diag}(\ell_+) = \begin{bmatrix} \ell_i(\mathbf{x}) & \\ & \ddots \end{bmatrix}_{i \in \mathcal{I}_+} \qquad P_- = \begin{bmatrix} \cdots & \mathbf{p}_i & \cdots \end{bmatrix}_{i:\lambda_i<0}$$

2. $$M = \operatorname{diag}(\ell_+) Y_+ P_- (Y_- P_-)^{-1} = \begin{bmatrix} & \vdots & \\ \cdots & \mu_{ij} & \cdots \\ & \vdots & \end{bmatrix}_{i \in \mathcal{I}_+, j \in \mathcal{I}_- \setminus \{0\}} \in \mathbb{R}^{|\mathcal{I}_+| \times (|\mathcal{I}_-|-1)}$$

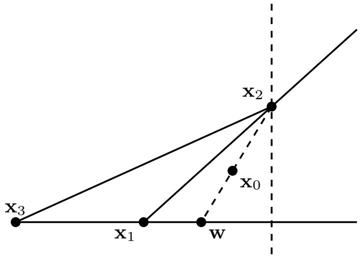
3. $\mu_{i0} = \ell_i(\mathbf{x}) - \sum_{j \in \mathcal{I}_- \setminus \{0\}} \mu_{ij}$ for all $i \in \mathcal{I}_+$.

Theorem (upper bound when there is negative μ)

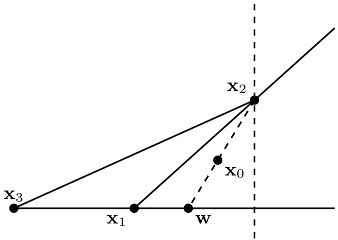
Assume $f \in C_L^{1,1}(\mathbb{R}^2)$. *Let* $\hat{f} \in \mathcal{P}_2^1$ *interpolates* f *at any affinely independent* $\Theta = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\} \subset \mathbb{R}^2$. *For any* $\mathbf{x} \in \mathbb{R}^2$ *with* $\ell_2(\mathbf{x}) > 0, \ell_3(\mathbf{x}) < 0$, *and* $\ell_1(\mathbf{x})(\mathbf{x}_2 - \mathbf{x}_1) \cdot (\mathbf{x}_3 - \mathbf{x}_1) - \ell_3(\mathbf{x})(\mathbf{x}_2 - \mathbf{x}_3) \cdot (\mathbf{x}_1 - \mathbf{x}_3) < 0$, *we have*

$$|\hat{f}(\mathbf{x}) - f(\mathbf{x})| \leq \frac{1}{2} G \cdot H^\star.$$

where $H^\star = P \begin{bmatrix} +\nu & 0 \\ 0 & -\nu \end{bmatrix} P^{-1}$ *with* $P = \begin{bmatrix} \mathbf{x}_2 - \mathbf{x}_0 & \mathbf{x}_1 - \mathbf{x}_3 \end{bmatrix}$.



The Difficult Cases in Bivariate Interpolation



Theorem

The bound in the previous theorem is sharp.

Proof.

This upper bound can be achieved by the piecewise quadratic function

$$f(\mathbf{u}) = \begin{cases} \frac{\nu}{2}\|\mathbf{u}-\mathbf{w}\|^2 - \nu\left(\frac{(\mathbf{x}_1-\mathbf{x}_3)\cdot(\mathbf{u}-\mathbf{w})}{\|\mathbf{x}_1-\mathbf{x}_3\|^2}\right)^2 & \text{if } (\mathbf{u}-\mathbf{w})\cdot(\mathbf{x}_1-\mathbf{x}_3) \le 0, \\ \frac{\nu}{2}\|\mathbf{u}-\mathbf{w}\|^2 & \text{if } (\mathbf{u}-\mathbf{w})\cdot(\mathbf{x}_1-\mathbf{x}_3) \ge 0. \end{cases}$$

When are all $\mu_{ij} \geq 0$?

Based on our observation, we believe the following statements are true.

1. When there is no obtuse angle at the vertices of the simplex conv Θ, that is, when
$$(\mathbf{x}_j - \mathbf{x}_i) \cdot (\mathbf{x}_k - \mathbf{x}_i) \geq 0 \text{ for all } i, j, k = 1, 2, \ldots, n+1,$$
the parameters $\{\mu_{ij}\}_{i \in \mathcal{I}_+, j \in \mathcal{I}_-}$ are all non-negative for any $\mathbf{x} \in \mathbb{R}^n$.
2. If there is at least one obtuse angle at the vertices of the simplex conv(Θ), then there is a non-empty subset of $\mathbb{R}^n$ to which if $\mathbf{x}$ belongs, there is at least one negative element in $\{\mu_{ij}\}_{i \in \mathcal{I}_+, j \in \mathcal{I}_-}$.

1. Motivation and Problem Definition
2. Preliminaries and Existing Results
3. Error Estimation Problem
4. An Improved Upper Bound
5. Maximizing Error over Quadratic Functions
6. The Difficult Cases in Bivariate Extrapolation