

Learning the Maximum of a Hölder Function from Inexact Data

Simon Foucart*— Texas A&M University

Abstract

Within the theoretical framework of Optimal Recovery, one determines in this article the *best* procedures to learn a quantity of interest depending on a Hölder function acquired via inexact point evaluations at fixed datasites. *Best* here refers to procedures minimizing worst-case errors. The elementary arguments hint at the possibility of tackling nonlinear quantities of interest, with a particular focus on the function maximum. In a local setting, i.e., for a fixed data vector, the optimal procedure (outputting the so-called Chebyshev center) is precisely described relatively to a general model of inexact evaluations. Relatively to a slightly more restricted model and in a global setting, i.e., uniformly over all data vectors, another optimal procedure is put forward, showing how to correct the natural underestimate that simply returns the data vector maximum. Jittered data are also briefly discussed as a side product of evaluating the minimal worst-case error optimized over the datasites.

Key words and phrases: Hölder functions, Lipschitz functions, optimal recovery, Chebyshev centers, information-based complexity.

AMS classification: 26A16, 41A28, 46N10, 65D15, 90C47.

1 Introduction

How can one reliably estimate the maximum $\max[f] := \max\{f(x), x \in \mathcal{X}\}$ of an unknown function from point evaluation data $y_m = f(x^{(m)})$, $m \in [1:M]$? This question has been considered e.g. in [2, 10] with the common intent of avoiding the construction of a full approximant \hat{f} to f followed by the computation of the maximum of \hat{f} . Both articles also assumed that the datasites $x^{(m)}$ could be freely chosen and that the data y_m were exact. In contrast, in this note, the datasites are prescribed and the data are inexact, i.e., of the form

$$(1) \quad y_m = f(x^{(m)}) + e_m \quad \text{with} \quad |e_m| \leq \varepsilon_m$$

for some known $\varepsilon_1, \dots, \varepsilon_M \geq 0$ (sometimes all taken to equal the same ε , but definitely not always). One shortly writes $\mathbf{y} = \mathbf{\Lambda}_x f + \mathbf{e}$, with $\mathbf{\Lambda}_x$ being the linear map assigning $[g(x^{(1)}); \dots; g(x^{(M)})] \in \mathbb{R}^M$

*S. F. is partially supported by grants from the NSF (DMS-2505204) and the ONR (N00014-20-1-2787).

to a function g and with $\mathbf{e} \in \mathcal{E}_\varepsilon := [-\varepsilon_1, \varepsilon_1] \times \cdots \times [-\varepsilon_M, \varepsilon_M]$. Clearly, estimating $\max[f]$ from this information alone is nonsensical without an assumption preventing wild behaviors of f away from the datasites. Such an assumption is here expressed as $f \in \mathcal{K}$, where the so-called model set \mathcal{K} is the set of Lipschitz functions defined on a compact metric space \mathcal{X} , or more generally the set of Hölder functions with parameters $\alpha \in (0, 1]$ given by¹

$$(2) \quad \mathcal{H}_\alpha := \{f : \mathcal{X} \rightarrow \mathbb{R} \text{ such that } |f(x) - f(x')| \leq \text{dist}(x, x')^\alpha \text{ for all } x, x' \in \mathcal{X}\}.$$

With a focus on worst-case scenarios for assessing a procedure $\Delta : \mathbf{y} \mapsto z$ designed to estimate $\max[f]$, or another quantity of interest $\Gamma(f) \in Z$, one defines worst-case errors

- locally at \mathbf{y} via

$$(3) \quad \text{lwce}_{\mathbf{y}, \varepsilon}(z) := \sup \{ \|\Gamma(f) - z\|_Z : f \in \mathcal{H}_\alpha, \mathbf{y} - \mathbf{\Lambda}_\mathbf{x} f \in \mathcal{E}_\varepsilon \},$$

- or globally via

$$(4) \quad \text{gwce}_\varepsilon(\Delta) := \sup \{ \|\Gamma(f) - \Delta(\mathbf{\Lambda}_\mathbf{x} f + \mathbf{e})\| : f \in \mathcal{H}_\alpha, \mathbf{e} \in \mathcal{E}_\varepsilon \}.$$

Note that the ‘local’ problem of minimizing $\text{lwce}_{\mathbf{y}, \varepsilon}(z)$ over z is harder than solving the ‘global’ problem of minimizing $\text{gwce}_\varepsilon(\Delta)$ over Δ , in the sense that solutions $\hat{z}_{\mathbf{y}, \varepsilon}$ to the former directly provide a solution $\Delta : \mathbf{y} \mapsto \hat{z}_{\mathbf{y}, \varepsilon}$ to the latter—in passing, one should take notice of the identity $\text{gwce}_\varepsilon(\Delta) = \sup \{ \text{lwce}_{\mathbf{y}, \varepsilon}(\Delta(\mathbf{y})) : \mathbf{y} \in \mathbf{\Lambda}_\mathbf{x}(\mathcal{H}_\alpha) + \mathcal{E}_\varepsilon \}$. The framework just described is the one underlying the theories of Optimal Recovery and subsequently of Information-Based Complexity (where the optimal selection of $\mathbf{\Lambda}_\mathbf{x}$ is also sought). These theories often assume that the model set is convex and symmetric and that Γ is a linear map, implying in several situations that a globally optimal map Δ can be chosen linear (see [8, Chapter 4], [3, Chapters 9 and 10], and some of the author’s recent works), hinting once again at the preferability of the global problem over the local one. The case of a linear quantity of interest Γ treated here provides yet another situation where global optimality can be achieved by a linear Δ , see Theorem 3. Still, the locally optimal Δ is actually not very burdensome either, see Theorem 1. In the case $\Gamma = \text{Id}$ with exact data and a model class of Lipschitz functions, this result was already obtained in [1], where the inspiration for the McShane-extension-type functions $\ell_{\mathbf{y}, \varepsilon}$ and $u_{\mathbf{y}, \varepsilon}$ of (5)–(6) comes from. The novelty of this note is the realization that, even with inexact data, local and global Optimal Recovery problems may also be completely solved for some nonlinear quantities of interest, in particular for the maximum of a function. The situation of inexact data and of linear quantities of interest had been studied in [9], while the situation of exact data and of the maximum as quantity of interest had been treated in [7, Section 1.3.6] following the works [11, 14], but amalgamating the two situations required a fresh perspective.

Here is a preview of this note’s contribution about the case $\Gamma(f) = \max[f]$. First, Section 2 solves the local problem by showing, as a specific instance of Theorem 1, that

$$\Delta^{\text{loc}} : \mathbf{y} \in \mathbb{R}^m \mapsto \frac{\max[\ell_{\mathbf{y}, \varepsilon}] + \max[u_{\mathbf{y}, \varepsilon}]}{2} \in \mathbb{R}$$

¹The set \mathcal{H}_α can also be interpreted as a set of Lipschitz functions with respect to a modified metric defined by $\text{dist}_\alpha(x, x') = \text{dist}(x, x')^\alpha$, $x, x' \in \mathcal{X}$.

is a locally optimal procedure for the estimation of $\max[f]$, where

$$(5) \quad \ell_{\mathbf{y},\varepsilon}(x) := \max_{m \in [1:M]} (y_m - \varepsilon_m - \text{dist}(x, x^{(m)})^\alpha),$$

$$(6) \quad u_{\mathbf{y},\varepsilon}(x) := \min_{m \in [1:M]} (y_m + \varepsilon_m + \text{dist}(x, x^{(m)})^\alpha).$$

One may be content with using Δ^{loc} as a global solution—it is! Still, Section 3 re-solves the global problem, under the proviso that the ε_m 's are all equal to a common ε , by exhibiting a ‘simpler’ globally optimal procedure. Namely, Theorem 4 shows that

$$\Delta^{\text{glo}} : \mathbf{y} \in \mathbb{R}^m \mapsto \max_{m \in [1:M]} (y_m) + \frac{1}{2} \max[U],$$

is a globally optimal procedure for the estimation of $\max[f]$, where the function U is nothing else than $u_{\mathbf{0},\mathbf{0}}$, i.e., it is given by²

$$(7) \quad U(x) := \min_{m \in [1:M]} (\text{dist}(x, x^{(m)})^\alpha).$$

This result makes perfect sense in hindsight, at least for exact data ($\varepsilon = 0$). Indeed, a natural estimation of $\max[f]$ is $\max_m(y_m)$, but this always underestimates $\max[f]$, so a correction term—depending on the location of the $x^{(m)}$'s—should be added. Note that the map Δ^{glo} , unlike Δ^{loc} , is independent of ε , but its global worst-case error does depend on it.

Section 4 concludes by considering the situation where the metric dist stems from a norm on \mathbb{R}^d and the set \mathcal{X} is the unit ball for this norm. The order of the smallest possible global worst-case error over the choice of $x^{(1)}, \dots, x^{(M)}$ is uncovered there. Perhaps unsurprisingly, it features ε and $M^{-\alpha/d}$, see Theorem 6. In the particular case of the ℓ_∞ -norm on \mathbb{R}^d , for which $\mathcal{X} = [-1, 1]^d$, one can even give an exact expression for this smallest possible global worst-case error. Perhaps unsurprisingly again, the best choice for the $x^{(m)}$'s turn out to be the centers of the hypercubic cells partitioning $[-1, 1]^d$ in a regular grid, see Theorem 7 and its proof.

2 Locally Optimal Solutions

In this section, considering a fixed observation vector $\mathbf{y} \in \mathbb{R}^m$ with entries given, as in (1), by $y_m = f(x^{(m)}) + e_m$, $|e_m| \leq \varepsilon_m$, one aims at producing an estimation of $\max[f]$ —or more generally of a quantity of interest $\Gamma(f) \in Z$ —which is worst-case optimal at this particular \mathbf{y} . The optimality, relatively to the model set \mathcal{H}_α of Hölder functions introduced in (2), thus refers to the local sense

²For $\alpha = 1$ and $\text{dist}(x, x') = \|x - x'\|_{\ell_2^d}$, the quantity $\max[U]$ is sometimes called the fill distance of $\{x^{(1)}, \dots, x^{(M)}\}$ for \mathcal{X} , see [15].

as described in (3)³. For full generality (not used beyond intuitive cases), it is assumed that the normed space Z into which Γ maps is a Banach lattice. An overview and more details about Banach lattices can be found e.g. in [5, Chapters 1 and 12], see also [6], but essentially these are normed spaces equipped with an order relation, compatible with vector addition and scalar multiplication, such that one can consider the ‘maximum’ of two elements from Z , and hence the ‘absolute value’ of an element $z \in Z$ as defined by $|z| = \max\{-z, z\} \in Z$. The norm should be monotone⁴, in the sense that, for any $z, z' \in Z$,

$$(8) \quad |z| \geq |z'| \implies \|z\|_Z \geq \|z'\|_Z.$$

As for the quantity of interest $\Gamma : B(\mathcal{X}) \rightarrow Z$ with domain equal to the space of bounded functions on \mathcal{X} , it is not assumed to be linear. Instead, it is only required to be monotone, in the sense that, for any $f, f' \in B(\mathcal{X})$,

$$(9) \quad f \geq f' \implies \Gamma(f) \geq \Gamma(f').$$

Note that this covers quite a number of cases: as linear examples, one can think of Γ being the identity (i.e., $\Gamma(f) = f$), a domain restriction (i.e., $\Gamma(f) = f|_\Omega$), and a positive linear functional (e.g. $\Gamma(f) = \int_{\mathcal{X}} f$), etc.; as nonlinear examples, one can think of Γ being an increasing map applied pointwise (e.g. $\Gamma(f) = \exp(f)$) or—the main focus of this note—the maximum of a function (i.e., $\Gamma(f) = \max[f]$). The result below exactly solves the problem of minimizing the local worst-case error. The function $\ell_{\mathbf{y},\varepsilon}$ and $u_{\mathbf{y},\varepsilon}$ appearing there are the ones introduced in (5)–(6).

Theorem 1. For the recovery of a monotone quantity of interest Γ mapping into a Banach lattice Z , given the data (1) and the model set (2), a minimizer of $\text{lwce}_{\mathbf{y},\varepsilon}(z)$ over all $z \in Z$ is given by

$$\hat{z}_{\mathbf{y},\varepsilon} = \frac{\Gamma(\ell_{\mathbf{y},\varepsilon}) + \Gamma(u_{\mathbf{y},\varepsilon})}{2},$$

while the value of the minimum is

$$\hat{r}_{\mathbf{y},\varepsilon} = \frac{\|\Gamma(u_{\mathbf{y},\varepsilon}) - \Gamma(\ell_{\mathbf{y},\varepsilon})\|_Z}{2}.$$

Proof. Central to the argument is a basic observation (elucidating the notation $\ell_{\mathbf{y},\varepsilon}$ and $u_{\mathbf{y},\varepsilon}$, which stands for lower and upper bounds, respectively). It starts by writing, for any $f \in \mathcal{H}_\alpha$, any $x \in \mathcal{X}$, and any $m \in [1:M]$,

$$|f(x) - y_m| \leq |f(x^{(m)}) - y_m| + |f(x) - f(x^{(m)})| \leq \varepsilon_m + \text{dist}(x, x^{(m)})^\alpha,$$

that is to say

$$y_m - \varepsilon_m - \text{dist}(x, x^{(m)})^\alpha \leq f(x) \leq y_m + \varepsilon_m + \text{dist}(x, x^{(m)})^\alpha.$$

³‘Locally optimal’ maps are also called ‘strongly optimal’ or ‘central’ in Information-Based Complexity, see [13].

⁴There are several notions of monotone norms, listed and compared in [4].

Taking the maximum and minimum over $m \in [1:M]$, one obtains

$$(10) \quad \ell_{\mathbf{y},\varepsilon}(x) \leq f(x) \leq u_{\mathbf{y},\varepsilon}(x) \quad \text{for all } f \in \mathcal{H}_\alpha \text{ and } x \in \mathcal{X}.$$

Furthermore, one observes that both $\ell_{\mathbf{y},\varepsilon}$ and $u_{\mathbf{y},\varepsilon}$ are data-consistent and model-consistent, i.e., $|y_m - \ell_{\mathbf{y},\varepsilon}(x^{(m)})| \leq \varepsilon_m$ and $|y_m - u_{\mathbf{y},\varepsilon}(x^{(m)})| \leq \varepsilon_m$ for all $m \in [1:M]$, as well as $\ell_{\mathbf{y},\varepsilon} \in \mathcal{H}_\alpha$ and $u_{\mathbf{y},\varepsilon} \in \mathcal{H}_\alpha$. Indeed, for the first statement, using the defining expressions of $\ell_{\mathbf{y},\varepsilon}$ and $u_{\mathbf{y},\varepsilon}$ evaluated at $x^{(m)}$ for any $m \in [1:M]$, one sees that $y_m - \varepsilon_m \leq \ell_{\mathbf{y},\varepsilon}(x^{(m)})$ and $u_{\mathbf{y},\varepsilon}(x^{(m)}) \leq y_m + \varepsilon_m$, so that, in view of (10), one arrives at $y_m - \varepsilon_m \leq \ell_{\mathbf{y},\varepsilon}(x^{(m)}) \leq u_{\mathbf{y},\varepsilon}(x^{(m)}) \leq y_m + \varepsilon_m$ for any $m \in [1:M]$. This is data-consistency. For the second statement, given $x, x' \in \mathcal{X}$ and $m \in [1:M]$, one has

$$\begin{aligned} (y_m - \varepsilon_m - \text{dist}(x, x^{(m)})^\alpha) - \ell_{\mathbf{y},\varepsilon}(x') &\leq (y_m - \varepsilon_m - \text{dist}(x, x^{(m)})^\alpha) - (y_m - \varepsilon_m - \text{dist}(x', x^{(m)})^\alpha) \\ &= -\text{dist}(x, x^{(m)})^\alpha + \text{dist}(x', x^{(m)})^\alpha \leq \text{dist}(x, x')^\alpha, \end{aligned}$$

which yields $\ell_{\mathbf{y},\varepsilon}(x) - \ell_{\mathbf{y},\varepsilon}(x') \leq \text{dist}(x, x')^\alpha$ by taking the maximum over $m \in [1:M]$. The inequality $|\ell_{\mathbf{y},\varepsilon}(x) - \ell_{\mathbf{y},\varepsilon}(x')| \leq \text{dist}(x, x')^\alpha$ is obtained by exchanging the roles of x and x' , thus showing that $\ell_{\mathbf{y},\varepsilon} \in \mathcal{H}_\alpha$. This is model-consistency for $\ell_{\mathbf{y},\varepsilon}$. The case of $u_{\mathbf{y},\varepsilon}$ is treated similarly.

Armed with the above observations, it is time to prove the prospective result that, for any $z \in Z$,

$$\text{lwce}_{\mathbf{y},\varepsilon}(z) \geq \frac{\|\Gamma(u_{\mathbf{y},\varepsilon}) - \Gamma(\ell_{\mathbf{y},\varepsilon})\|_Z}{2} \quad \text{with equality when } z = \widehat{z}_{\mathbf{y},\varepsilon} = \frac{\Gamma(\ell_{\mathbf{y},\varepsilon}) + \Gamma(u_{\mathbf{y},\varepsilon})}{2}.$$

To justify one part of this claim, given any $z \in Z$, the fact that both $\ell_{\mathbf{y},\varepsilon}$ and $u_{\mathbf{y},\varepsilon}$ are model- and data-consistent implies that

$$\text{lwce}_{\mathbf{y},\varepsilon}(z) \geq \begin{cases} \|\Gamma(u_{\mathbf{y},\varepsilon}) - z\|_Z \\ \|\Gamma(\ell_{\mathbf{y},\varepsilon}) - z\|_Z \end{cases} \geq \frac{1}{2} (\|\Gamma(u_{\mathbf{y},\varepsilon}) - z\|_Z + \|\Gamma(\ell_{\mathbf{y},\varepsilon}) - z\|_Z) \geq \frac{1}{2} \|\Gamma(u_{\mathbf{y},\varepsilon}) - \Gamma(\ell_{\mathbf{y},\varepsilon})\|_Z.$$

For the other part of the claim, given any $f \in \mathcal{H}_\alpha$ satisfying $\mathbf{y} - \mathbf{\Lambda}_\mathbf{x} f \in \mathcal{E}_\varepsilon$, i.e., $|y_m - f(x^{(m)})| \leq \varepsilon_m$ for all $m \in [1:M]$, the inequalities (10) and the monotonicity (9) of Γ yield $\Gamma(\ell_{\mathbf{y},\varepsilon}) \leq \Gamma(f) \leq \Gamma(u_{\mathbf{y},\varepsilon})$, i.e.,

$$-\frac{\Gamma(u_{\mathbf{y},\varepsilon}) - \Gamma(\ell_{\mathbf{y},\varepsilon})}{2} \leq \Gamma(f) - \frac{\Gamma(\ell_{\mathbf{y},\varepsilon}) + \Gamma(u_{\mathbf{y},\varepsilon})}{2} \leq \frac{\Gamma(u_{\mathbf{y},\varepsilon}) - \Gamma(\ell_{\mathbf{y},\varepsilon})}{2}.$$

From here, it follows (with the ‘absolute value’ being the one relative to the Banach lattice Z) that

$$\left| \Gamma(f) - \frac{\Gamma(\ell_{\mathbf{y},\varepsilon}) + \Gamma(u_{\mathbf{y},\varepsilon})}{2} \right| \leq \frac{\Gamma(u_{\mathbf{y},\varepsilon}) - \Gamma(\ell_{\mathbf{y},\varepsilon})}{2},$$

allowing to conclude, using the monotonicity (8) of the norm on Z , that

$$\left\| \Gamma(f) - \frac{\Gamma(\ell_{\mathbf{y},\varepsilon}) + \Gamma(u_{\mathbf{y},\varepsilon})}{2} \right\|_Z \leq \frac{\|\Gamma(u_{\mathbf{y},\varepsilon}) - \Gamma(\ell_{\mathbf{y},\varepsilon})\|_Z}{2}.$$

Taking the supremum over f shows that $\text{lwce}_{\mathbf{y},\varepsilon}(\widehat{z}_{\mathbf{y},\varepsilon}) \leq \|\Gamma(u_{\mathbf{y},\varepsilon}) - \Gamma(\ell_{\mathbf{y},\varepsilon})\|_Z/2$, thus justifying the other part of the claim. The proof is now complete. \square

The elements $\hat{z}_{\mathbf{y},\varepsilon} \in Z$ and $\hat{r}_{\mathbf{y},\varepsilon} := \text{lwce}_{\mathbf{y},\varepsilon}(\hat{z}_{\mathbf{y},\varepsilon}) \in \mathbb{R}_+$ are called Chebyshev center and radius, respectively, for the set $\{\Gamma(f) : f \in \mathcal{H}_\alpha, \mathbf{y} - \mathbf{\Lambda}_\mathbf{x}f \in \mathcal{E}_\varepsilon\}$, as they are solutions to

$$\underset{z,r}{\text{minimize}} \quad r \quad \text{subject to} \quad \|\Gamma(f) - z\|_Z \leq r \text{ whenever } f \in \mathcal{H}_\alpha \text{ and } \mathbf{y} - \mathbf{\Lambda}_\mathbf{x}f \in \mathcal{E}_\varepsilon.$$

In other words, they are center and radius of a smallest Z -ball containing this set. Using the notation $\hat{f}_{\mathbf{y},\varepsilon}$ for the Chebyshev center in the special case $\Gamma = I$, Theorem 1 reveals in particular that, if the quantity of interest Γ is linear, then $\hat{z}_{\mathbf{y},\varepsilon} = \Gamma(\hat{f}_{\mathbf{y},\varepsilon})$, i.e., a locally optimal estimation of $\Gamma(f)$ can be obtained by construction a full approximant and plugging it into Γ . This fortuity does not hold anymore for nonlinear quantities of interest, say for $\Gamma = \max$. Indeed, the plug-in estimate $\max[\hat{f}_{\mathbf{y},\varepsilon}] = \max[(\ell_{\mathbf{y},\varepsilon} + u_{\mathbf{y},\varepsilon})/2]$ is generally smaller than $\hat{z}_{\mathbf{y},\varepsilon} = (\max[\ell_{\mathbf{y},\varepsilon}] + \max[u_{\mathbf{y},\varepsilon}])/2$, as confirmed from Figure 1 below.

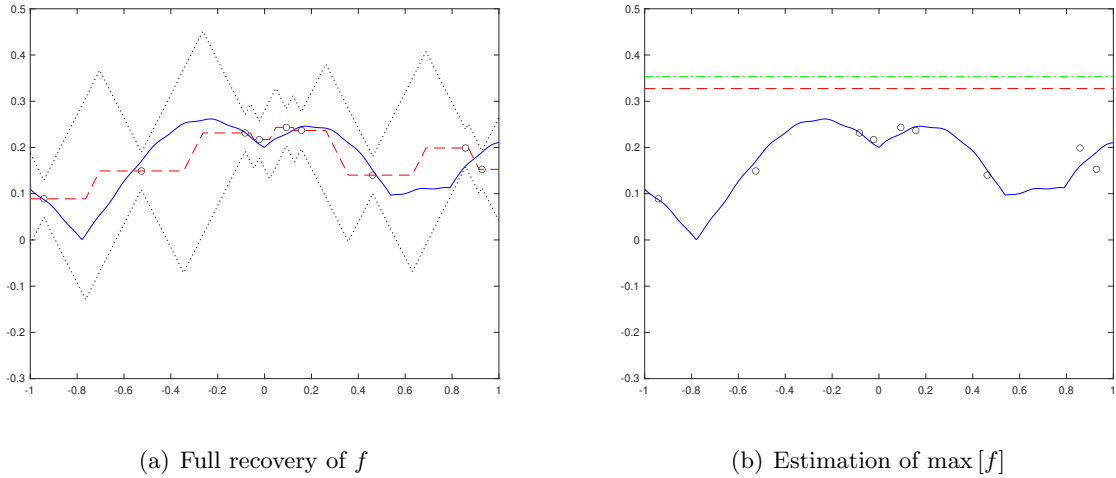


Figure 1: An example of a univariate Lipschitz function f (solid blue curve) acquired via inexact point values (black circles). The left panel displays the lower function $\ell_{\mathbf{y},\varepsilon}$ and the upper function $u_{\mathbf{y},\varepsilon}$ from (5)-(6) (dotted curves), as well as the locally optimal estimation of f from Theorem 1 (dashed red curve). The right panel shows, as horizontal lines, the locally optimal estimation of $\max[f]$ from Theorem 1 (dashed red line) and the globally optimal estimation of $\max[f]$ from Theorem 4 (dash-dotted green line).

The Chebyshev center $\hat{z}_{\mathbf{y},\varepsilon}$ is not supplied as explicitly as ideally desired, even for $\Gamma(f) = \max[f]$. For the specific case of univariate ($\mathcal{X} = [-1, 1]$) Lipschitz ($\alpha = 1$) functions acquired via exact point evaluations ($\varepsilon_m = 0$), it is actually possible to be completely explicit.

Proposition 2. For functions $f : [-1, 1] \rightarrow \mathbb{R}$ satisfying the condition $|f(x) - f(x')| \leq |x - x'|$ for all $x, x' \in [-1, 1]$ and observed via $y_1 = f(x^{(1)})$, ..., $y_M = f(x^{(M)})$ at $x^{(1)} < \dots < x^{(M)} \in (-1, 1)$, the locally optimal estimation of $\max[f]$ is expressed as

$$\hat{z}_{\mathbf{y},0} = \frac{1}{2} \left[\max_{m \in [1:M]} (y_m) + \max_{m \in [0:M]} \left(\frac{x^{(m+1)} - x^{(m)} + y_{m+1} + y_m}{2} \right) \right],$$

having set $x^{(0)} = -1$, $y_0 = 0$, $x^{(M+1)} = 1$, and $y_{M+1} = 0$.

Proof. Based on Theorem 1, one needs to untangle $\max[\ell_{\mathbf{y},\mathbf{0}}]$ and $\max[u_{\mathbf{y},\mathbf{0}}]$. The former can be dealt with in full generality. Indeed, in view of $\ell_{\mathbf{y},\mathbf{0}}(x) = \max_m (y_m - \text{dist}(x, x^{(m)})^\alpha)$, one observes on the one hand that $\ell_{\mathbf{y},\mathbf{0}}(x) \leq \max_m (y_m)$ for all $x \in \mathcal{X}$ and on the other hand that $\ell_{\mathbf{y},\mathbf{0}}(x^{(m)}) \geq y_m$ for all $m \in [1:M]$. These two facts combine to show that

$$\max[\ell_{\mathbf{y},\mathbf{0}}] = \max_m (y_m).$$

It remains to determine $\max[u_{\mathbf{y},\mathbf{0}}]$ explicitly in the specific univariate setting of the proposition. Precisely, one shall show that

$$(11) \quad \max[u_{\mathbf{y},\mathbf{0}}] = \max_{m \in [0:M]} \left(\frac{x^{(m+1)} - x^{(m)} + y_{m+1} + y_m}{2} \right).$$

To this end, one considers the Voronoi-type cells $V_{\mathbf{y},1}, \dots, V_{\mathbf{y},M}$ defined for $m \in [1:M]$ by

$$V_{\mathbf{y},m} := \{x \in [-1, 1] : y_m + |x - x^{(m)}| \leq y_n + |x - x^{(n)}| \text{ for all } n \in [1:M]\},$$

so that

$$(12) \quad \max[u_{\mathbf{y},\mathbf{0}}] = \max_{x \in [-1,1]} \min_{m \in [1:M]} (y_m + |x - x^{(m)}|) = \max_{m \in [1:M]} \max_{x \in V_{\mathbf{y},m}} (y_m + |x - x^{(m)}|).$$

At a fixed $m \in [1:M]$, introducing $V_{\mathbf{y},m,n} := \{x \in [-1, 1] : |x - x^{(m)}| - |x - x^{(n)}| \leq y_n - y_m\}$ for each $n \in [1:M]$, one has $V_{\mathbf{y},m} = \bigcap_{n \neq m} V_{\mathbf{y},m,n}$. For $n < m$, since $x \mapsto |x - x^{(m)}| - |x - x^{(n)}|$ is a ramp function joining $|x^{(n)} - x^{(m)}|$ on the left of $x^{(n)}$ to $-|x^{(n)} - x^{(m)}|$ on the right of $x^{(m)}$ and since $|y_n - y_m| \leq |x^{(n)} - x^{(m)}|$ by the Lipschitz condition, a point $x \in [-1, 1]$ belongs to $V_{\mathbf{y},m,n}$ if and only if $x \geq \alpha^{(m,n)}$, where $\alpha^{(m,n)} \in [x^{(n)}, x^{(m)}]$ is implicitly defined by the equation $|\alpha^{(m,n)} - x^{(m)}| - |\alpha^{(m,n)} - x^{(n)}| = y_n - y_m$, or explicitly given by $\alpha^{(m,n)} = (x^{(m)} + x^{(n)} + y_m - y_n)/2$. For $n > m$, one also obtains that a point $x \in [-1, 1]$ belongs to $V_{\mathbf{y},m,n}$ if and only if $x \leq \beta^{(m,n)}$, where $\beta^{(m,n)} \in [x^{(m)}, x^{(n)}]$ is implicitly defined by the equation $|\beta^{(m,n)} - x^{(m)}| - |\beta^{(m,n)} - x^{(n)}| = y_n - y_m$, or explicitly given by $\beta^{(m,n)} = (x^{(m)} + x^{(n)} - y_m + y_n)/2$. As a result, for $m \in [2 : M - 1]$, one derives that

$$\begin{aligned} V_{\mathbf{y},m} &= \left(\bigcap_{n < m} [\alpha^{(m,n)}, 1] \right) \bigcap \left(\bigcap_{n > m} [-1, \beta^{(m,n)}] \right) = [\max_{n < m} \alpha^{(m,n)}, 1] \bigcap [-1, \min_{n > m} \beta^{(m,n)}] \\ &= [\alpha^{(m,m-1)}, \beta^{(m,m+1)}], \end{aligned}$$

where one used the easily verifiable facts that $\alpha^{(m,n)} \leq \alpha^{(m,m-1)}$ for $n < m$ and $\beta^{(m,n)} \geq \beta^{(m,m+1)}$ for $n > m$. Now, noticing that $\beta^{(m,m+1)}$ coincides with $\alpha^{(m+1,m)}$ with equal value

$$\xi^{(m)} := \frac{x^{(m)} + x^{(m+1)} - y_m + y_{m+1}}{2}, \quad m \in [1:M-1],$$

one deduces that the Voronoi-type cells reduce to

$$V_{\mathbf{y},m} = [\xi^{(m-1)}, \xi^{(m)}] \quad \text{for } m \in [2 : M-1].$$

This remains valid for $m = 1$ if one sets $\xi^{(0)} = -1$ and for $m = M$ if one sets $\xi^{(M)} = 1$ —for instance, $V_{\mathbf{y},M} = \cap_{n < M} [\alpha^{(M,n)}, 1] = [\alpha^{(M,M-1)}, 1] = [\xi^{(M-1)}, 1]$. It follows that, for any $m \in [1:M]$,

$$\begin{aligned} \max_{x \in V_{\mathbf{y},m}} (y_m + |x - x^{(m)}|) &= \max\{y_m + x^{(m)} - \xi^{(m-1)}, y_m + \xi^{(m)} - x^{(m)}\} \\ &= \max\left\{\frac{x^{(m)} - x^{(m-1)} + y_m + y_{m-1}}{2}, \frac{x^{(m+1)} - x^{(m)} + y_{m+1} + y_m}{2}\right\}, \end{aligned}$$

where the boundary cases $m = 1$ and $m = M$ require some special attention. Substituting the latter into (12) and eliminating the repetitions leads to the expression announced in (11). \square

3 Globally Optimal Solutions

In this section, one still aims at producing a recovery map $\Delta : \mathbb{R}^m \rightarrow Z$ which is worst-case optimal for the estimation of a quantity of interest $\Gamma(f) \in Z$, but the optimality here refers to the global sense as described in (4). As mentioned earlier, the locally optimal recovery map $\Delta^{\text{loc}} : \mathbf{y} \in \mathbb{R}^m \mapsto \hat{z}_{\mathbf{y},\epsilon} \in Z$ from Theorem 1 is already automatically globally optimal. Nevertheless, one may be interested in other maps $\Delta : \mathbb{R}^m \rightarrow Z$ minimizing $\text{gwce}_\epsilon(\Delta)$, one reason being that they could be easier to construct and manipulate, e.g. they could be linear in selected instances where the quantity of interest Γ is itself linear. Contrary to Section 2, the cases of linear and nonlinear Γ can seemingly not be treated in one swift stroke, so they are separated from now on.

3.1 The case of linear quantities of interest

In the result below, the generic function $\mathbb{1}_V$ denotes the indicator function of a set V , meaning that $\mathbb{1}_V(x) = 1$ if $x \in V$ and $\mathbb{1}_V(x) = 0$ if $x \notin V$, while the sets $V_{\epsilon,1}, \dots, V_{\epsilon,M} \subseteq \mathcal{X}$ are the Voronoi-type cells defined by

$$V_{\epsilon,m} := \{x \in \mathcal{X} : \epsilon_m + \text{dist}(x, x^{(m)})^\alpha \leq \epsilon_n + \text{dist}(x, x^{(n)})^\alpha \text{ for all } n \in [1:M]\}.$$

Theorem 3. For the recovery of a monotone linear quantity of interest Γ from $B(\mathcal{X})$ into a Banach lattice Z , given the datasites $x^{(1)}, \dots, x^{(M)} \in \mathcal{X}$ and the model set (2), the minimal global worst-case error is

$$\inf_{\Delta: \mathbb{R}^M \rightarrow Z} \text{gwce}_\epsilon(\Delta) = \|\Gamma(u_{\mathbf{0},\epsilon})\|_Z, \quad \text{where } u_{\mathbf{0},\epsilon}(x) = \min_{m \in [1:M]} (\epsilon_m + \text{dist}(x, x^{(m)})^\alpha).$$

Furthermore, the infimum is achieved for the recovery map $\Gamma \circ \Delta^{\text{glo}} : \mathbb{R}^M \rightarrow Z$, where

$$\Delta^{\text{glo}} : \mathbf{y} \in \mathbb{R}^M \mapsto \sum_{m=1}^M y_m \mathbb{1}_{V_{\epsilon,m}} \in B(\mathcal{X}).$$

Proof. For the first part of the proof, i.e., the fact that $\text{gwce}_\varepsilon(\Delta) \geq \|\Gamma(u_{\mathbf{0},\varepsilon})\|_Z$ for any $\Delta : \mathbb{R}^m \rightarrow Z$, one relies of the aforementioned identity $\text{gwce}_\varepsilon(\Delta) = \sup\{\text{lwce}_{\mathbf{y},\varepsilon}(\Delta(\mathbf{y})) : \mathbf{y} \in \mathbf{\Lambda}_{\mathbf{x}}(\mathcal{H}_\alpha) + \mathcal{E}_\varepsilon\}$. One immediately derives that

$$\text{gwce}_\varepsilon(\Delta) \geq \text{lwce}_{\mathbf{0},\varepsilon}(\Delta(\mathbf{y})) \geq \widehat{r}_{\mathbf{0},\varepsilon} = \frac{\|\Gamma(u_{\mathbf{0},\varepsilon}) - \Gamma(\ell_{\mathbf{0},\varepsilon})\|_Z}{2} = \|\Gamma(u_{\mathbf{0},\varepsilon})\|_Z,$$

where the last equality used the fact (seen from (5)–(6)) that $\ell_{\mathbf{0},\varepsilon} = -u_{\mathbf{0},\varepsilon}$ and the linearity of Γ .

For the second part of the proof, it is enough to show that $\text{gwce}_\varepsilon(\Gamma \circ \Delta^{\text{glo}}) \leq \|\Gamma(u_{\mathbf{0},\varepsilon})\|_Z$. To this end, let $f \in \mathcal{H}_\alpha$, $\mathbf{e} \in \mathcal{E}_\varepsilon$, and let the shorthand $\mathbf{y} := \mathbf{\Lambda}_{\mathbf{x}}f + \mathbf{e}$ be employed below. Then, for any $x \in \mathcal{X}$, select $n \in [1:M]$ such that $x \in V_{\varepsilon,n}$ and observe that

$$\begin{aligned} |f(x) - \Delta^{\text{glo}}(\mathbf{y})(x)| &= |f(x) - \sum_{m=1}^M y_m \mathbb{1}_{V_{\varepsilon,m}}(x)| = |f(x) - y_n| = |f(x) - (f(x^{(n)}) + e_n)| \\ &\leq |e_n| + |f(x) - f(x^{(n)})| \leq \varepsilon_n + \text{dist}(x, x^{(n)})^\alpha = u_{\mathbf{0},\varepsilon}(x). \end{aligned}$$

This reads $-u_{\mathbf{0},\varepsilon} \leq f - \Delta^{\text{glo}}(\mathbf{y}) \leq u_{\mathbf{0},\varepsilon}$ and, thanks the monotonicity (9) and linearity of Γ , it follows that $-\Gamma(u_{\mathbf{0},\varepsilon}) \leq \Gamma(f) - (\Gamma \circ \Delta^{\text{glo}})(\mathbf{y}) \leq \Gamma(u_{\mathbf{0},\varepsilon})$, i.e., $|\Gamma(f) - (\Gamma \circ \Delta^{\text{glo}})(\mathbf{y})| \leq \Gamma(u_{\mathbf{0},\varepsilon})$. Finally, thanks to the monotonicity (8) of the norm, one concludes that $\|\Gamma(f) - (\Gamma \circ \Delta^{\text{glo}})(\mathbf{y})\|_Z \leq \|\Gamma(u_{\mathbf{0},\varepsilon})\|_Z$. Taking the maximum over f and \mathbf{e} finishes the proof. \square

A few comments are appropriate to conclude on the subject of monotone linear quantities of interest:

- the two optimal recovery maps that were exposed, the locally—hence globally—optimal one from Theorem 1 and the globally optimal one from Theorem 3, both take the form $\Gamma \circ \Delta$, with Δ given by either

$$\Delta^{\text{loc}}(\mathbf{y}) = \frac{\ell_{\mathbf{y},\varepsilon} + u_{\mathbf{y},\varepsilon}}{2} \quad \text{or} \quad \Delta^{\text{glo}}(\mathbf{y}) = \sum_{m=1}^M y_m \mathbb{1}_{V_{\varepsilon,m}},$$

i.e., they are obtained by applying the quantity of interest after producing a full estimation;

- when all the ε_m 's are equal to a common ε , both Δ^{loc} and Δ^{glo} are independent of this ε (but of course the minimal worst-case errors do depend on ε);
- the local estimate $\Delta^{\text{loc}}(\mathbf{y})$ is data- and model-consistent, while the global estimate $\Delta^{\text{glo}}(\mathbf{y})$ is not model-consistent, as it is discontinuous, hence not Hölder;
- the global estimate $\Delta^{\text{glo}}(\mathbf{y})$ depends linearly on $\mathbf{y} \in \mathbb{R}^m$, while local estimate $\Delta^{\text{loc}}(\mathbf{y})$ does not—this does not necessarily mean, however, that the former is simpler than the latter, as producing $\Delta^{\text{loc}}(\mathbf{y})$ via the defining expressions (5)–(6) of $\ell_{\mathbf{y},\varepsilon}$ and $u_{\mathbf{y},\varepsilon}$ can be viewed as easier than producing $\Delta^{\text{glo}}(\mathbf{y})$, if done through the construction all the Voronoi-type cells;
- the local estimate $\Delta^{\text{loc}}(\mathbf{y})$ seems more convenient for streaming data than the global estimate $\Delta^{\text{glo}}(\mathbf{y})$, as the former is easy to update (via the defining expressions (5)–(6) of $\ell_{\mathbf{y},\varepsilon}$ and $u_{\mathbf{y},\varepsilon}$)

when a new $(x^{(M+1)}, y_{M+1})$ is added, while this addition potentially affects a large number of Voronoi-type cells for the latter. But if one is primarily interested in prediction at a fixed $x \in \mathcal{X}$, then the global estimate is also straightforward to update, as it essentially returns the value y_m at the datasite $x^{(m)}$ closest to x , so one only needs to check if x is closer to $x^{(M+1)}$ than to the previous closest datasite.

3.2 The case of the maximum

Turning to the estimation of nonlinear quantities of interest, as an encompassing result seems out of reach, one concentrates specifically on the maximum, i.e., $\Gamma(f) = \max[f]$. To start, consider the situation of exact observations, i.e., $y_m = f(x^{(m)})$ for all $m \in [1:M]$. To estimate $\max[f]$, it might seem natural to simply output $\max_m(y_m)$, but this is always an underestimation. One probably needs to add a slight correction. To guess what this correction could be, look at the local optimality result instantiated at $\mathbf{y} = \mathbf{0}$, which suggests $(\max[\ell_{\mathbf{0},\mathbf{0}}] + \max[u_{\mathbf{0},\mathbf{0}}])/2$. Since $\max[\ell_{\mathbf{0},\mathbf{0}}] = 0$ and $u_{\mathbf{0},\mathbf{0}}$ coincides with the function U defined in (7), this correction simplifies to $\max[U]/2$. It turns to be the right guess, remaining valid when the observation error bounds ε_m are all equal to a common ε . In this case, a globally optimal recovery map can be chosen independently of this ε , as formally stated below.

Theorem 4. For the recovery of the quantity of interest $f \in B(\mathcal{X}) \mapsto \max[f] \in \mathbb{R}$, given the data (1) with $\varepsilon_m = \varepsilon$ for all $m \in [1:M]$ and the model set (2), the minimal global worst-case error is

$$\inf_{\Delta: \mathbb{R}^M \rightarrow \mathbb{R}} \text{gwce}_\varepsilon(\Delta) = \frac{1}{2} \max[U] + \varepsilon, \quad \text{where } U(x) = \min_{m \in [1:M]} (\text{dist}(x, x^{(m)})^\alpha).$$

Furthermore, the infimum is achieved for the recovery map

$$\Delta^{\text{glo}} : \mathbf{y} \in \mathbb{R}^M \mapsto \max_{m \in [1:M]} (y_m) + \frac{1}{2} \max[U] \in \mathbb{R}.$$

This result is in fact obtained by setting $\varepsilon_m = \varepsilon$ in the more general theorem appearing next. Before that, one takes notice that the above global recovery map Δ^{glo} is different, though close, to the local recovery map Δ^{loc} from Theorem 1. Indeed, imposing $\varepsilon_m = \varepsilon$ for all $m \in [1:M]$, one easily arrives at the expressions

$$\begin{aligned} \Delta^{\text{glo}}(\mathbf{y}) &= \max_{m \in [1:M]} (y_m) + \frac{1}{2} \max \left[\min_{m \in [1:M]} (\text{dist}(\cdot, x^{(m)})^\alpha) \right], \\ \Delta^{\text{loc}}(\mathbf{y}) &= \frac{1}{2} \max_{m \in [1:M]} \left(y_m + \frac{1}{2} \max \left[\min_{m \in [1:M]} (y_m + \text{dist}(\cdot, x^{(m)})^\alpha) \right] \right), \end{aligned}$$

which could even be made more explicit in the univariate setting by invoking Proposition 2. These expressions incidentally reveal that the global estimate is never smaller than the local estimate, i.e., that $\Delta^{\text{glo}}(\mathbf{y}) \geq \Delta^{\text{loc}}(\mathbf{y})$, as illustrated in Figure 1(b).

Theorem 5. For the recovery of the quantity of interest $f \in B(\mathcal{X}) \mapsto \max[f] \in \mathbb{R}$, given the data (1) and the model set (2), the minimal global worst-case error is bounded below as

$$\inf_{\Delta: \mathbb{R}^M \rightarrow \mathbb{R}} \text{gwce}_\varepsilon(\Delta) \geq \frac{1}{2} \max[u_{\mathbf{0}, \varepsilon}] + \frac{1}{2} \min_{m \in [1:M]} (\varepsilon_m),$$

while the recovery map $\tilde{\Delta}: \mathbf{y} \in \mathbb{R}^M \mapsto \max_{m \in [1:M]} (y_m) + \frac{1}{2} \max[u_{\mathbf{0}, \varepsilon}] - \frac{1}{2} \max_{m \in [1:M]} (\varepsilon_m) \in \mathbb{R}$ has global worst-case error bounded above as

$$\text{gwce}_\varepsilon(\tilde{\Delta}) \leq \frac{1}{2} \max[u_{\mathbf{0}, \varepsilon}] + \frac{1}{2} \max_{m \in [1:M]} (\varepsilon_m).$$

Proof. For the lower bound, one points out that $\ell_{\mathbf{0}, \varepsilon} \in \mathcal{H}_\alpha$ and $u_{\mathbf{0}, \varepsilon} \in \mathcal{H}_\alpha$ based on an argument similar to one in the proof of Theorem 1. One also sees from their defining expressions that $-\varepsilon_m \leq \ell_{\mathbf{0}, \varepsilon}(x^{(m)}) \leq 0 \leq u_{\mathbf{0}, \varepsilon}(x^{(m)}) \leq \varepsilon_m$ for all $m \in [1:M]$, and hence $\ell_{\mathbf{0}, \varepsilon}(x^{(m)}) + e_m = 0$ for some $\mathbf{e} \in \mathcal{E}_\varepsilon$ and $u_{\mathbf{0}, \varepsilon}(x^{(m)}) + e_m = 0$ for some $\mathbf{e} \in \mathcal{E}_\varepsilon$. Therefore, one derives that, for any $\Delta: \mathbb{R}^M \rightarrow \mathbb{R}$,

$$\begin{aligned} \text{gwce}_\varepsilon(\Delta) &\geq \begin{cases} |\max[\ell_{\mathbf{0}, \varepsilon}] - \Delta(0)| \\ |\max[u_{\mathbf{0}, \varepsilon}] - \Delta(0)| \end{cases} \geq \frac{1}{2} (|\max[u_{\mathbf{0}, \varepsilon}] - \Delta(0)| + |\max[\ell_{\mathbf{0}, \varepsilon}] - \Delta(0)|) \\ &\geq \frac{1}{2} (\max[u_{\mathbf{0}, \varepsilon}] - \max[\ell_{\mathbf{0}, \varepsilon}]). \end{aligned}$$

The result follows from the observation that $\max[\ell_{\mathbf{0}, \varepsilon}] = \max_m (-\varepsilon_m) = -\min_m (\varepsilon_m)$.

For the upper bound on $\text{gwce}_\varepsilon(\tilde{\Delta})$, one needs to establish that, given $f \in \mathcal{H}_\alpha$, $\mathbf{e} \in \mathcal{E}_\varepsilon$, and writing $\mathbf{y} = \mathbf{\Lambda}_x f + \mathbf{e}$ for short,

$$-\frac{1}{2} \max[u_{\mathbf{0}, \varepsilon}] - \frac{1}{2} \max_{m \in [1:M]} (\varepsilon_m) \leq \max[f] - \tilde{\Delta}(\mathbf{y}) \leq \frac{1}{2} \max[u_{\mathbf{0}, \varepsilon}] + \frac{1}{2} \max_{m \in [1:M]} (\varepsilon_m).$$

Taking the expression of $\tilde{\Delta}(\mathbf{y})$ into account in these prospective inequalities, one notices that the leftmost one reduces to $\max[f] \geq \max_m (y_m) - \max_m (\varepsilon_m)$, while the rightmost one reduces to $\max[f] \leq \max_m (y_m) + \max[u_{\mathbf{0}, \varepsilon}]$. The facts that $\max[f] \geq \max_m (f(x^{(m)})) = \max_m (y_m - e_m)$ and $f(x) \leq u_{\mathbf{y}, \varepsilon}(x) = \min_m (y_m + \varepsilon_m + \text{dist}(x, x^{(m)})^\alpha) \leq \max_m (y_m) + u_{\mathbf{0}, \varepsilon}(x)$ for all $x \in \mathcal{X}$ yield the former and the latter, respectively. \square

Remark. Theorem 4 could be extended in a further direction, incorporating jittered observations, so as to receive data of the form $y_m = f(x^{(m)} + \xi^{(m)}) + e_m$ with $|e_m| \leq \varepsilon$ and $\|\xi^{(m)}\| \leq \delta$. It is hereby assumed that this norm is the one generating the metric dist and the domain $\mathcal{X} \subseteq \mathbb{R}^d$ as its unit ball. Defining a new notion of global error including the worst case over $\xi^{(1)}, \dots, \xi^{(M)} \in \mathcal{X}$, one would obtain $\text{gwce}_{\varepsilon, \delta}(\Delta) \geq \max[U]/2 + \varepsilon$ for any $\Delta: \mathbb{R}^M \rightarrow \mathbb{R}$, simply because this new notion $\text{gwce}_{\varepsilon, \delta}$ of global worst-case error is larger than the previous notion gwce_ε . Moreover, for the recovery map Δ^{glo} from Theorem 4, one can obtain $\text{gwce}_{\varepsilon, \delta}(\Delta^{\text{glo}}) \leq \max[U]/2 + \varepsilon + \delta$ by making the necessary adjustments in the second part of the proof of Theorem 5. Note that it makes sense

to assume that the balls $B(x^{(m)}, \delta)$ do not intersect, otherwise a jittered observation y_m could also come from another datasite $x^{(n)}$ with $n \neq m$. Consequently, if μ is the volumetric measure, then

$$M\delta^d \mu(B(0, 1)) = \sum_{m=1}^M \mu(B(x^{(m)}, \delta)) = \mu\left(\bigcup_{m=1}^M B(x^{(m)}, \delta)\right) \leq \mu(B(0, 1 + \delta)) \leq (1 + \delta)^d \mu(B(0, 1)),$$

yielding $M^{1/d} \leq 1 + 1/\delta \leq 2/\delta$, i.e., $\delta \leq 2/M^{1/d}$. Since the next section will establish that $\max[U] \geq 1/(2M^{\alpha/d}) \geq 1/(2M^{1/d})$, one derives that $\delta \leq 4\max[U]$, which in turn implies that $\text{gwce}_{\varepsilon, \delta}(\Delta^{\text{glo}}) \leq \max[U]/2 + \varepsilon + 4\max[U] \leq 9(\max[U]/2 + \varepsilon) \leq 9\text{gwce}_{\varepsilon, \delta}(\Delta)$ for any $\Delta : \mathbb{R}^M \rightarrow \mathbb{R}$. In other words, one can assert that Δ^{glo} is near optimal with a factor 9.

4 Estimations of the Minimal Global Worst-Case Errors

Throughout this section, given a norm $\|\cdot\|$ on \mathbb{R}^d , the domain $\mathcal{X} \subseteq \mathbb{R}^d$ is chosen to be its unit ball $B(0, 1)$ equipped with the Lebesgue measure μ normalized so that $\mu(\mathcal{X}) = 1$. The metric involved in the Hölder model set \mathcal{H}_α defined in (2) is also chosen to be the one subordinated to this norm, i.e., $\text{dist}(x, x') = \|x - x'\|$. As shown in the previous section, the minimal global worst-case errors are dictated by the function $u_{\mathbf{0}, \varepsilon}$, e.g. via its L_p -norm for the recovery of $\Gamma(f) = f$, Γ being the identity from $B(\mathcal{X})$ to $L_p(\mathcal{X})$ (see Theorem 3), or via its L_∞ -norm for the recovery of $\max[f]$ (see Theorem 5). The goal here is to evaluate these quantities optimized over the selection of datasites $x^{(1)}, \dots, x^{(M)}$. The first result applies to a general $\mathcal{X} = B(0, 1)$ and gives a precise two-sided estimate valid for any $p \in [0, \infty]$. It is followed by an exact expression applying to the case of the ℓ_∞ -norm on \mathbb{R}^d , so that $\mathcal{X} = [-1, 1]^d$.

Theorem 6. For \mathcal{X} being the unit ball of \mathbb{R}^d relatively to a norm $\|\cdot\|$ and for $p \in [1, \infty]$, recalling that $u_{\mathbf{0}, \varepsilon}(x) = \min_m (\varepsilon_m + \|x - x^{(m)}\|^\alpha)$, one has

$$\min_{m \in [1:M]} (\varepsilon_m) + \frac{c}{M^{\alpha/d}} \leq \inf_{x^{(1)}, \dots, x^{(M)} \in \mathcal{X}} \|u_{\mathbf{0}, \varepsilon}\|_{L_p(\mathcal{X})} \leq \max_{m \in [1:M]} (\varepsilon_m) + \frac{C}{M^{\alpha/d}},$$

where the absolute constants $c, C > 0$ can be taken as $c = 1/2$ and $C = 3$.

Proof. For the lower bound, given any $x^{(1)}, \dots, x^{(M)} \in \mathcal{X}$, one uses $\|u_{\mathbf{0}, \varepsilon}\|_{L_p(\mathcal{X})} \geq \|u_{\mathbf{0}, \varepsilon}\|_{L_1(\mathcal{X})}$ and $u_{\mathbf{0}, \varepsilon}(x) \geq \min_m (\varepsilon_m) + U(x)$ to deduce that $\|u_{\mathbf{0}, \varepsilon}\|_{L_p(\mathcal{X})} \geq \min_m (\varepsilon_m) + \|U\|_{L_1(\mathcal{X})}$. Then, in order to bound $\|U\|_{L_1(\mathcal{X})}$ from below, one adapts the proof of [12], where Sukharev was interested in the optimal recovery of the integral of a Lipschitz function on $[0, 1]^d$, but otherwise the argument is

essentially the same. Namely, for $\tau > 0$ to be chosen later, one writes

$$\begin{aligned}
\|U\|_{L_1(\mathcal{X})} &= \int_{\mathcal{X}} U(x) d\mu(x) = \int_0^\infty \mu\{x \in \mathcal{X} : U(x) > t\} dt \geq \int_0^\tau [1 - \mu\{x \in \mathcal{X} : U(x) \leq t\}] dt \\
&= \int_0^\tau [1 - \mu\{x \in \mathcal{X} : \|x - x^{(m)}\|^\alpha \leq t \text{ for some } m \in [1:M]\}] dt \\
&\geq \int_0^\tau \left[1 - \sum_{m=1}^M \mu\{x \in \mathcal{X} : \|x - x^{(m)}\| \leq t^{1/\alpha}\}\right] dt \geq \tau - \sum_{m=1}^M \int_0^\tau (t^{1/\alpha})^d dt \\
&= \tau - \frac{M}{d/\alpha + 1} \tau^{d/\alpha + 1}.
\end{aligned}$$

This lower bound is optimized when $1 - M\tau^{d/\alpha} = 0$, i.e., $\tau = M^{-\alpha/d}$, yielding

$$\|U\|_{L_1(\mathcal{X})} \geq M^{-\alpha/d} - \frac{M}{d/\alpha + 1} (M^{-\alpha/d})^{d/\alpha + 1} = \left(1 - \frac{1}{d/\alpha + 1}\right) M^{-\alpha/d} = \frac{d}{d + \alpha} M^{-\alpha/d}.$$

The announced inequality now simply follows from $d/(d + \alpha) \geq 1/2$.

For the upper bound, one uses $\|u_{\mathbf{0}, \varepsilon}\|_{L_p(\mathcal{X})} \leq \|u_{\mathbf{0}, \varepsilon}\|_{L_\infty(\mathcal{X})}$ and $u_{\mathbf{0}, \varepsilon}(x) \leq \max_m(\varepsilon_m) + U(x)$ to deduce that $\|u_{\mathbf{0}, \varepsilon}\|_{L_p(\mathcal{X})} \leq \max_m(\varepsilon_m) + \|U\|_{L_\infty(\mathcal{X})}$. Then, in order to bound $\|U\|_{L_\infty(\mathcal{X})}$ from above, one relies on a folklore result about covering numbers. Namely, one can find a set $\{x^{(1)}, \dots, x^{(N)}\} \subseteq \mathcal{X}$ of $N \leq (1 + 2/\theta)^d$ points such that, for any $x \in \mathcal{X}$, there exists $n \in [1:N]$ such that $\|x - x^{(n)}\| \leq \theta$. Selecting $\theta = 3/M^{1/d}$, one has $N \leq (3/\theta)^d = M$, so one can complete the above set to form a θ -net $\{x^{(1)}, \dots, x^{(M)}\} \subseteq \mathcal{X}$ consisting of M points. Using these datasites, it is now clear that $\|U\|_{L_\infty(\mathcal{X})} = \max_x \min_m \|x - x^{(m)}\|^\alpha \leq (3/M^{1/d})^\alpha \leq 3/M^{\alpha/d}$. \square

The result of Theorem 6 can be refined for the standard domain $\mathcal{X} = [-1, 1]^d$, as uncovered below. The case $p = \infty$, not included but obtainable as a limiting case, would in fact be easier to establish.

Theorem 7. If \mathbb{R}^d is equipped with the ℓ_∞ -norm, so that $\mathcal{X} = [-1, 1]^d$ and $\text{dist}(x, x') = \|x - x'\|_{\ell_\infty^d}$, and if all the ε_m 's are equal to a common ε , then, for any $p \in [1, \infty)$,

$$\inf_{x^{(1)}, \dots, x^{(M)} \in \mathcal{X}} \|u_{\mathbf{0}, \varepsilon}\|_{L_p(\mathcal{X})}^p \geq \left(\varepsilon + \frac{1}{M^{\alpha/d}}\right)^p - M \int_{\varepsilon^p}^{(\varepsilon + 1/M^{\alpha/d})^p} (t^{1/p} - \varepsilon)^{d/\alpha} dt,$$

with equality being achieved when $M^{1/d}$ is an integer. Furthermore, if $\varepsilon = 0$, then

$$\inf_{x^{(1)}, \dots, x^{(M)} \in \mathcal{X}} \|u_{\mathbf{0}, \varepsilon}\|_{L_p(\mathcal{X})} \geq \left(\frac{d}{d + \alpha p}\right)^{1/p} \frac{1}{M^{\alpha/d}} \quad \text{with equality when } M^{1/d} \text{ is an integer.}$$

Proof. For arbitrary points $x^{(1)}, \dots, x^{(M)} \in \mathcal{X}$, let V_1, \dots, V_M denote the (genuine) Voronoi cells,

i.e., $V_m = \{x \in \mathcal{X} : \|x - x^{(m)}\|_{\ell_\infty^d} \leq \|x - x^{(n)}\|_{\ell_\infty^d} \text{ for all } n \in [1:M]\}$. One starts by writing

$$\begin{aligned}
 \|u_{\mathbf{0},\varepsilon}\|_{L_p(\mathcal{X})}^p &= \int_{\mathcal{X}} \min_{m \in [1:M]} (\varepsilon + \|x - x^{(m)}\|_{\ell_\infty^d}^\alpha)^p d\mu(x) = \sum_{m=1}^M \int_{V_m} (\varepsilon + \|x - x^{(m)}\|_{\ell_\infty^d}^\alpha)^p d\mu(x) \\
 &= \sum_{m=1}^M \int_0^\infty \mu\{x \in V_m : (\varepsilon + \|x - x^{(m)}\|_{\ell_\infty^d}^\alpha)^p > t\} dt \\
 &= \sum_{m=1}^M \int_0^\infty \mu\{x \in V_m : \|x - x^{(m)}\|_{\ell_\infty^d} > (t^{1/p} - \varepsilon)^{1/\alpha}\} dt \\
 (13) \quad &= \sum_{m=1}^M \int_0^\infty \mu(V_m \setminus B(x^{(m)}, (t^{1/p} - \varepsilon)^{1/\alpha})) dt.
 \end{aligned}$$

For $\tau > 0$ to be chosen later, one continues with

$$\begin{aligned}
 \|u_{\mathbf{0},\varepsilon}\|_{L_p(\mathcal{X})}^p &\geq \sum_{m=1}^M \int_0^\tau \mu(V_m \setminus B(x^{(m)}, (t^{1/p} - \varepsilon)^{1/\alpha})) dt \\
 &\geq \sum_{m=1}^M \int_0^\tau (\mu(V_m) - \mu(B(x^{(m)}, (t^{1/p} - \varepsilon)^{1/\alpha}))) dt \\
 &= \tau - M \int_{\varepsilon^p}^\tau (t^{1/p} - \varepsilon)^{d/\alpha} dt.
 \end{aligned}$$

Selecting the optimal $\tau = (\varepsilon + M^{-\alpha/d})^p$ gives the announced lower bound.

Assuming now that $M = N^d$ for some integer $N \in \mathbb{N}$, one aims at showing that the lower bound can be achieved. To this end, one partitions $\mathcal{X} = [-1, 1]^d$ into a regular grid made of N^d hypercubes of side $2/N$ and one picks $x^{(1)}, \dots, x^{(M)}$ as their centers. Thus, the hypercubes are the Voronoi cells associated with $x^{(1)}, \dots, x^{(M)}$, i.e., $B(x^{(m)}, 1/N) = V_m$. Coming back to (13) while remarking that

$$\mu(V_m \setminus B(x^{(m)}, (t^{1/p} - \varepsilon)^{1/\alpha})) = \begin{cases} N^{-d}, & \text{if } t \leq \varepsilon^p, \\ N^{-d} - (t^{1/p} - \varepsilon)^{d/\alpha}, & \text{if } \varepsilon^p < t \leq (\varepsilon + N^{-\alpha})^p, \\ 0, & \text{if } t > (\varepsilon + N^{-\alpha})^p, \end{cases}$$

one obtains, for this choice of $x^{(1)}, \dots, x^{(M)}$,

$$\begin{aligned}
 \|u_{\mathbf{0},\varepsilon}\|_{L_p(\mathcal{X})}^p &= \sum_{m=1}^M \left(\int_0^{\varepsilon^p} \frac{1}{N^d} dt + \int_{\varepsilon^p}^{(\varepsilon + N^{-\alpha})^p} \left(\frac{1}{N^d} - (t^{1/p} - \varepsilon)^{d/\alpha} \right) dt \right) \\
 &= (\varepsilon + N^{-\alpha})^p - M \int_{\varepsilon^p}^{(\varepsilon + N^{-\alpha})^p} (t^{1/p} - \varepsilon)^{d/\alpha} dt.
 \end{aligned}$$

Substituting $N = M^{1/d}$ shows that the latter indeed equals the previous lower bound. \square

With $p = 1$ and $\alpha = 1$, the infimum from Theorem 7 becomes $\varepsilon + (d/(d+1))M^{-1/d}$. Similarly, Theorem 6 also demonstrated that in general the minimal global-worst case error over all possible recovery procedures and datasites is composed of two terms: ε , the observation accuracy, and $M^{-1/d}$, where M is the number of observations. The term $M^{-1/d}$ is bad news: even for perfectly accurate observations, an error of order $\eta < 1$ can only be achieved with a number M of observations of order η^{-d} —this is the curse of dimensionality, as formalized in Information-Based Complexity. But the high dimensionality seems to come with a silver lining, too: in a scenario when the number M of observations is fixed, independent of d , and when the observation accuracy ε can be selected, the smallest error achieved at $\varepsilon = 0$ is of the same order $M^{-1/d}$ as if one chose $\varepsilon = M^{-1/d}$, which is less stringent as the dimension d grows.

References

- [1] Beliaikov, G., 2006. *Interpolation of Lipschitz functions*. Journal of Computational and Applied Mathematics, 196(1), 20–44.
- [2] Cohen, A., DeVore, R., Petrova, G., Wojtaszczyk, P., 2013. *Finding the minimum of a function*. Methods and Applications of Analysis, 20(4), 365–382.
- [3] Foucart, S., 2022. Mathematical Pictures at a Data Science Exhibition. Cambridge University Press.
- [4] Johnson, C. R., Nymen, P., 1991. *Monotonicity properties of norms*. Linear Algebra and its Applications, 148, 43–58.
- [5] Johnson, W. B., Lindenstrauss, J., 2001. Handbook of the Geometry of Banach Spaces. Elsevier.
- [6] Meyer-Nieberg, P., 2012. Banach Lattices. Springer Science & Business Media.
- [7] Novak, E., 2006. Deterministic and Stochastic Error Bounds in Numerical Analysis. Springer.
- [8] Novak, E., Wozniakowski, H., 2010. Tractability of Multivariate Problems (Vol. 1). European Mathematical Society.
- [9] Plaskota, L., 1996. Noisy Information and Computational Complexity. Cambridge University Press.
- [10] Rudi, A., Marteau-Ferey, U., Bach, F., 2025. *Finding global minima via kernel approximations*. Mathematical Programming, 209, 703–784.
- [11] Sukharev, A. G., 1971. *Optimal strategies of the search for an extremum*. USSR Computational Mathematics and Mathematical Physics, 11(4), 119–137.
- [12] Sukharev, A. G., 1979. *Optimal numerical integration formulas for some classes of functions of several variables*. Doklady Akademii Nauk, vol. 246. Russian Academy of Sciences.

- [13] Traub, J. F., Wasilkowski, G. W., and Woźniakowski, H., 1988. Information-Based Complexity. Academic Press
- [14] Wasilkowski, G. W., 1984. *Some nonlinear problems are as easy as the approximation problem.* Computers & Mathematics with Applications, 10(4-5), 351–363.
- [15] Wendland, H., 2004. Scattered Data Approximation. Cambridge University Press.