NSM Converges to a k-NN Regressor Under Loose Lipschitz Estimates

Emilio T. Maddalena and Colin N. Jones, Member, IEEE

Abstract—Although it is known that having accurate Lipschitz estimates is essential for certain models to deliver good predictive performance, refining this constant in practice can be a difficult task especially when the input dimension is high. In this letter, we shed light on the consequences of employing loose Lipschitz bounds in the Nonlinear Set Membership (NSM) framework, showing that the model converges to a nearest neighbor regressor ($k$-NN with $k = 1$). This convergence process is moreover not uniform, and is monotonic in the univariate case. An intuitive geometrical interpretation of the result is then given and its practical implications are discussed.

Index Terms—Nonlinear set membership, nearest neighbors, Lipschitz continuity, regression.

I. INTRODUCTION

Non-parametric models were once considered computationally too expensive to be employed in practice. However, due to the increasing availability of computational power and the decrease of hardware costs, these tools are increasingly being embraced by the control community. In [1] and [2] for instance, the authors use scattered frequency domain data to design off-line robust controllers for an atomic force microscope and dc-dc power converters, respectively. Gaussian processes (GPs) are an example of another popular non-parametric modeling technique [3]. An early experimental investigation of GPs in control was reported in [4], where a gas-liquid separation plant was considered with a sampling period of $T_\text{samp} = 15$ s. More recent works have employed GPs to tackle the control of much faster systems such as autonomous racing cars with $T_\text{samp} = 20$ ms [5], and robotic arms with $T_\text{samp} = 1$ ms [6]. Besides their use in model predictive control (MPC) schemes as in the papers cited above, GPs have also been utilized to refine the parameters of classical state-feedback and proportional-integral-derivative (PID) compensators [7].

Certainly, the use of these statistical regression tools coming from other communities has been causing a change in the way system identification is performed in controls [8]. Moreover, there has also been an increasing concern regarding the safety of physical systems that operate in closed-loop based on the so called ‘learned models’ – see [9] for a recent review. This is especially the case when elements of on-line learning are present. In order to overcome these issues, researchers are currently carrying out rigorous analysis of these modeling procedures and assessing the associated uncertainties to design appropriate robust controllers [10], [11]. As opposed to the Bayesian techniques described previously, the Nonlinear Set Membership (NSM) approach to system identification [12] does not rely on the presence of priors, nor does it quantify uncertainties in a statistical fashion. This non-parametric methodology assumes a certain degree of regularity of the unknown target function $f(x)$, namely that it is Lipschitz continuous with a known upper bound. Based purely on collected data-points, tight deterministic bounds on the function values $f(x)$ at unobserved points can be established, making it an interesting modeling alternative. The theoretical foundations of the NSM technique were recently generalized to Hölder continuous functions [13], [14], extended to a large class of problems [15], and also to more abstract function spaces [16]. Applications of this methodology can be found in the context of vehicle yaw reference-tracking [17], electrical microgrids scheduling [18], and approximating linear MPC control laws [19]. In a broader sense, similar set-membership ideas were recently used to study the simulation of linear systems with guaranteed accuracy [20].

The quality of a prediction given by an NSM model strongly depends on how close the employed Lipschitz estimate is to the best constant, i.e., the lowest valid one. From a practical viewpoint however, refining this quantity can easily become a daunting task, especially in cases where the input dimension is large (see [21, Sec. 4.1], and the simulation results in [22]). Statistical methods exist to deal with the problem [23], but are limited to the univariate case. Ad hoc procedures followed by an augmentation to create a ‘safety margin’ seem to be a rather common way of estimating it in practical scenarios.

A question thus naturally arises: besides enlarging the error bounds, what effects do loose Lipschitz estimates have on the regressor itself?

Contributions: We answer the above question in this letter, showing that regardless of the input dimension, the...
The NSM model converges to another well-known non-parametric model, namely the nearest neighbor regressor (k-NN with \(k = 1\)). This convergence is shown under the \(L^2\) functional norm, and is moreover not uniform. The result is nevertheless valid for any finite-dimensional norm \(\| \cdot \|\) chosen to define the NSM and nearest neighbor regressors. Although not monotonic in general, the discrepancy between both models is shown to only decrease as the Lipschitz estimate is loosened in the univariate case. We emphasize that it is beyond the scope of this letter to measure the distance of any of these models and the unknown ground-truth, as our goal is to build a bridge between methodologies. Still, we do discuss the implications of this letter to measure the distance of any of these models shown to only decrease as the Lipschitz estimate is loosened in general, the discrepancy between both models is monotonic in general, the NSM Regressor: Given a dataset \(D\), a measure of distance induced by \(\| \cdot \|\) on \(X\), and a positive integer \(k \leq N\), the k-NN regressor is then defined as

\[
\hat{f}_{k-NN}(x) := \frac{1}{k} \sum_{n \in \mathcal{N}(x)} y_n
\]

where \(\mathcal{N}(x)\) is the index set of the \(k\)-nearest neighbors of \(x\), i.e., points \(x_j\) that satisfy \(\|x - x_j\| \leq \|x - x_n\|, \forall n \in \mathcal{N}(x), \forall n \in [N]\)\(n\). In our rather particular case, with \(k = 1\), \(\hat{f}_{k-NN}(x)\) simply replicates at \(x\) the image of its closest data-point in \(D\).

\[\text{Remark 1: Considering noisy measurements through a scalar bound as in the original work [12] does affect ceiling and floor functions respectively as}
\]

\[
\hat{f}(x) := \min_{n \in [N]} y_n + L\|x - x_n\|
\]

\[
\underline{f}(x) := \max_{n \in [N]} y_n - L\|x - x_n\|
\]

The nonlinear set membership regressor is then defined as

\[
f_{NSM}(x) := \frac{1}{2} \left( \hat{f}(x) + \underline{f}(x) \right)
\]

The previous assumption was needed to define a proper nearest neighbors (NN) regressor without ambiguity. The selection rule does not impact our analysis since it only affects a subset of the domain with measure zero.

A Voronoi cell associated with a single point \(x_n \in D\) is defined as

\[
\mathcal{C}_n := \{x \in X | \|x - x_n\| \leq \|x - x_j\|, \forall j \in [N]\}
\]

and the Voronoi diagram (VD) of a dataset \(D\), as the partition of the domain \(X\) induced by the Voronoi cells \(\mathcal{C}_1, \ldots, \mathcal{C}_N\). The nearest neighbor regressor (\(k = 1\)) then assumes the simplified form

\[
\hat{f}_{NN}(x) = y_n, \text{ if } x \in \mathcal{C}_n, \exists n \in [N]
\]

and is defined everywhere in \(X\).

\[\text{Remark 1: The quantifier } \exists \text{ in (4), as in all other piecewise functions introduced herein, should be read as ‘for some’}.\]

The NSM Regressor: We consider noiseless measurements as already indicated in (1) to simplify our presentation. The derivations can be repeated even when the dataset \(D\) is affected by an unknown but uniformly bounded noise.

The given dataset \(D\) and a Lipschitz constant estimate \(L\) for the unknown ground-truth, define the auxiliary ‘ceiling’ and ‘floor’ functions respectively as

\[
\bar{f}(x) := \min_{n \in [N]} y_n + L\|x - x_n\|
\]

\[
\underline{f}(x) := \max_{n \in [N]} y_n - L\|x - x_n\|
\]

The nonlinear set membership regressor is then defined as

\[
f_{NSM}(x) := \frac{1}{2} \left( \bar{f}(x) + \underline{f}(x) \right)
\]

As in the previous case, it is convenient to define a domain partition over which \(\bar{f}\) and \(\underline{f}\) enjoy a simpler representation.

Additively weighted Voronoi diagrams (AVD), also known as hyperbolic Voronoi diagrams, are an extension of VD where the n-th cell is defined as

\[
\{x \in X | \|x - x_n\| \leq \|x - x_j\| + \eta_{n,j}, \forall j \in [N]\}
\]

where \(\eta_{n,j}\) can be regarded as a bias. Next, construct two AVDs with cells \(\mathcal{C}_m\) and \(\mathcal{C}_n\) as in (8), respectively with \(\bar{\eta}_{n,j} = (y_j - y_n)/L\) and \(\underline{\eta}_{n,j} = (y_n - y_j)/L\). This causes the AVD domain partition to be influenced not only by the locations \(x_n\), but also by the attained values \(y_n\). As shown in [12], the ceiling and floor functions can be rewritten as

\[
\bar{f}(x) := y_n + L\|x - x_n\|, \text{ if } x \in \mathcal{C}_m, \exists n \in [N]
\]

\[
\underline{f}(x) := y_n - L\|x - x_n\|, \text{ if } x \in \mathcal{C}_n, \exists n \in [N]
\]

describing two piecewise conic functions defined over two distinct AVD. As a result, \(f_{NSM}\) is also defined by pieces, on the intersection of the AVDs, being in general piecewise non-linear if \(\| \cdot \|\) is not the \(\ell_1\) nor the \(\ell_\infty\) norm.

\[\text{Remark 2: Considering noisy measurements through a scalar bound as in the original work [12] does affect ceiling and floor } f \text{ functions. However, both the nominal model } f_{NSM} \text{ and the two AVD domain partitions remain unchanged if the noise bound is uniform.}\]
III. THEORETICAL RESULTS

A. Main Findings

We begin by introducing the auxiliary $\mathbb{A}^{n,m}$ and $\mathbb{B}^n$ cells that define a new partition of the domain $X$.

Definition 1: Let $\mathbb{A}^{n,m} := (\mathbb{C}^d \cap \mathbb{C}^e), n, m \in [N], m \neq n$, and $\mathbb{B}^n := (\mathbb{C}^e \cap \mathbb{C}^e), n \in [N]$.

Proposition 1: For any $n \in [N], \mathbb{B}^n \neq \emptyset$. Moreover, let $x \in \mathbb{B}^n$, then $x \in \mathbb{C}^d$.

Proof: From the definition of $\mathbb{B}^n$ we have

$$\mathbb{B}^n = \left\{ x \mid \|x - x_n\| \leq \|x - x_j\| + \frac{y_j - y_n}{L}, \forall j \in [N] \right\} \quad (11)$$

Since for all $j \in [N]$ either $y_j - y_n \leq 0$ or $y_n - y_j \leq 0$, then $\|x - x_n\| - \|x - x_j\| \leq 0 \implies x \in \mathbb{C}^d$.

The domain was partitioned into three different diagrams: a VD, a ceiling AVD, and a floor AVD. Each point in the domain thus belongs to a cell in each one of them. Given the cells to which a specific $x$ belongs, we establish in the following proposition an order for their associated images.

Proposition 2: If $x \in \mathbb{A}^{n,m}$ and also $x \in \mathbb{C}^d, n, m, p \in [N], p \neq n, p \neq m$, then $y_n \leq y_p \leq y_m$ must hold.

Proof: Follows from the feasibility of the set of inequities in $\mathbb{C}^d \cap \mathbb{C}^e \cap \mathbb{C}^e$.

Proposition 3: Let $F_{\text{NSM}} := \max_{x \in [N]} |y_n|$ be the maximum over the dataset labels absolute values, then $F_{\text{NSM}} \geq |f_{\text{NSM}}(x)|, \forall x \in X$, regardless of $L$.

Proof: The analysis is done region by region since there is only a finite number of them. From (9) and (10), inside the $\mathbb{B}^n$ cells, $f_{\text{NSM}}(x)$ always attains $y_n$. Assume now $\exists x \in \mathbb{A}^{n,m} : f_{\text{NSM}}(x) > y_m$. This is equivalent to $\frac{1}{2}(y_n + y_m + L(\|x - x_n\| - \|x - x_m\|)) > y_m$, which implies $\|x - x_n\| > \|x - x_m\| + \frac{y_m - y_n}{L} \implies x \notin \mathbb{C}^d \implies x \notin \mathbb{A}^{n,m}$, leading to a contradiction. Therefore, $\forall x \in \mathbb{A}^{n,m}, |f_{\text{NSM}}(x)| \leq y_m$.

An analogous contradiction can be constructed to show that $\forall x \in \mathbb{A}^{n,m}, |f_{\text{NSM}}(x)| \geq y_n$. Hence, on each $\mathbb{A}^{n,m}$, $|f_{\text{NSM}}(x)|$ is bounded by $\max(\{y_m,|y_n\})$. On the whole domain, the regressor is thus tightly bounded by $F_{\text{NSM}}$.

In what follows, $L_i, L_{i+1}, \ldots$ denotes an unbounded sequence of strictly increasing Lipschitz constants indexed by $i$, which give rise to a sequence of NSM regressors $f_{\text{NSM},i}$, $f_{\text{NSM},i+1}, \ldots$ The error is then defined as

$$e_i := f_{\text{NN}} - f_{\text{NSM},i} \quad (12)$$

with a point-wise subtraction.

To simplify our notation, the dependence of the functions and sets on the index $i$ will be omitted, and made explicit only when convenient. The main convergence theorem is stated next, and its proof relies on bounding the worst-case error between the NSM and NN functions, as well as on the strict expansion of certain $\mathbb{B}^n$ sets.

Theorem 1: Under the $\| \cdot \|_2$ functional, $f_{\text{NSM}} \rightarrow f_{\text{NN}}$ as $l \rightarrow \infty$.

Proof: The aim is to show that $\forall \epsilon > 0, \exists l$ with its associated $L_i$ s.t. $\|f_{\text{NN}} - f_{\text{NSM}}\|_2 < \epsilon$ for any equal or larger constant $L_i, L_{i+1}, \ldots$

From (9) and (10), begin by writing $f_{\text{NSM}}$ as

$$f_{\text{NSM}}(x) = \begin{cases} y_n, & \text{if } x \in \mathbb{B}^n, \exists n \in [N] \\ y_n - \frac{y_n + y_m}{2L_i} \|x - x_n\|, & \text{if } x \in \mathbb{A}^{n,m}, \exists n, m \in [N] \end{cases} \quad (13)$$

From (12) and Proposition 1, the error becomes

$$e_i(x) = \begin{cases} 0, & \text{if } x \in \mathbb{B}^n, \exists n \in [N] \\ y_n - \frac{y_n + y_m}{2L_i} \|x - x_n\| - \|x - x_m\|, & \text{if } x \in \mathbb{A}^{n,m}, \exists n, m \in [N] \end{cases} \quad (14)$$

where $N(x)$ is the index of the nearest neighbor cell $\mathbb{C}^e$.

From (14), it is clear that the union of all $\mathbb{A}^{n,m}$ also contains the support of $e_i(x)$. From (11), the dependence of $\mathbb{B}^n$ on the Lipschitz estimate $L_i$ is

$$\mathbb{B}^n_i = \left\{ x \mid \|x - x_n\| \leq \|x - x_i\| + \frac{|y_i - y_n|}{L_i}, \forall i \in [N] \right\} \quad (15)$$

which shows that for every $i$ the inequality is restricted by the bias if $y_i \neq y_n$, and that this restriction depends inversely on $L_i$. Due to Assumption 1, at least one $\mathbb{B}^n_i$ cell will be restricted by the bias; therefore, for that cell $\mathbb{B}^n_i \subset \mathbb{B}^n_{i+1} \subset \cdots$ since $L_i < L_{i+1} < \cdots$. This implies that

$$\bigcup_{n,m} \mathbb{B}^n_i \subset \mathbb{B}^n_{i+1} \subset \cdots \quad (16)$$

which in turn implies

$$\bigcup_{n,m} \mathbb{A}^{n,m}_i \supset \bigcup_{n,m} \mathbb{A}^{n,m}_{i+1} \supset \cdots \quad (17)$$

as the union over $n$ and $m$ of all $\mathbb{B}^n_i$ and $\mathbb{A}^{n,m}_i$ cells forms a partition of the domain $X$.

Given that $f_{\text{NSM}}$ in (4) is bounded and has only a finite number of discontinuities, which have measure zero, $f_{\text{NSM}}$ is integrable. In view of this fact and the continuity of $f_{\text{NSM}}, e_i$ is integrable $\forall i$ and its norm $\|e_i\|_2$ is defined. Let $F_{\text{NSM}} > 0$ and $F_{\text{NN}} > 0$ be constants such that $F_{\text{NSM}} > |f_{\text{NSM}}(x)|$ and $F_{\text{NN}} > |f_{\text{NN}}(x)|, \forall x \in X$ (the former was defined in Proposition 3). The inequality $F_{\text{NSM}} + F_{\text{NN}} = \geq F_0 > |e_i(x)|$ then holds $\forall x, i$.

Now define the auxiliary constant $\gamma_i^{n,m} := \int_{\mathbb{A}^{n,m}_i} 1 \, dx$, which is guaranteed to exist and to be finite due to the sets $\mathbb{A}^{n,m}_i$ being compact, and $e_i := \sum_{n,m} \gamma_i^{n,m}$. We can then upper-bound the square of a given error norm by

$$\|e_i\|_2^2 = \int_X e_i^2(x) \, dx \quad (18a)$$

$$= \sum_{n,m} \int_{\mathbb{A}^{n,m}_i} e_i^2(x) \, dx + \sum_i \int_{\mathbb{B}^n_i} e_i^2(x) \, dx \quad (18b)$$

$$= \sum_{n,m} \int_{\mathbb{A}^{n,m}_i} e_i^2(x) \, dx + 0 \quad (18c)$$
< F^2 \sum_{n,m} y_{1,n,m}^{n,m} \\
= F^2 \sigma_l \quad (18d) 
\]

Moreover, \(\|e_l\|_2^2 < F^2 \sigma_l \implies \|e_l\|_2 < F \sqrt{\sigma_l}\). As the sequence \(L_l\) is unbounded, we see from (15) that the union of sets in (16) tends to the whole domain \(X\), thus concentrating all the measure. This in turn implies that the union in (17) has a decreasing measure and leads to \(\lim_{n \to \infty} \|e_l\|_2 = 0\). Due to the latter fact and \(\|e_l\|_2\) being bounded from below by zero, it follows from the squeeze theorem that \(\lim_{n \to \infty} \|e_l\|_2 = 0\) and therefore \(f_{\text{NSM}}(x) \to \text{fnN}\).

**Theorem 2:** The convergence in Theorem 1 is not uniform, i.e., \(f_{\text{NSM}} \not\to f_{\text{fnN}}\) as \(l \to \infty\) w.r.t. \(\|\cdot\|_2\).

**Proof:** It suffices to show that \(\|e_l\|_2 = \|e_{l+1}\|_2\), \(\forall l\). Consider the error expression in (14) with \(x \in A^{n,m}\) and \(x \in C^p\). We analyze the case when \(y_p > y_n + \frac{y_m}{2}\), whereas \(y_p < \frac{y_n + y_m}{2}\) follows by analogy. Let \(y_p > \frac{y_n + y_m}{2}\) and \(\|x - x_0\| \geq \|x - x_m\|\), then \(|e_l(x)| = \frac{y_p - \frac{y_n + y_m}{2}}{L_l}\), leading to \(|e_l(x)| = |y_p - \frac{y_n + y_m}{2}|\). On the other hand, if \(\|x - x_n\| \geq \|x - x_m\|\), then \(|e_l(x)| = \frac{y_m - y_n}{L_l}\), leading to \(|e_l(x)| = |y_m - y_n|\). The previous analysis is valid for any region of \(e_l(x)\). Since all local maxima are independent of the Lipschitz estimate \(L_l\) and \(\|e_l\|_\infty\) is the maximum over all of them, we conclude that \(\|e_l\|_\infty = \|e_{l+1}\|_\infty\). \(\Box\)

**B. Additional Analysis**

In this section we inspect parts of the domain where the absolute error is guaranteed to decrease monotonically. This will be expressed in terms of two sufficient conditions later used to establish an additional result.

**Proposition 4:** Let \(x \in A^{n,m}, x \in \text{int}(C^p)\), and also \(x \in A^{n,m}_{l+1}\) at the next index. In this case, \(|e_l(x)| > |e_{l+1}(x)|\). The same conclusion holds if \(x \in \text{int}(C^p)\) instead.

**Proof:** Consider \(x \in A^{n,m}_{l+1}, x \in \text{int}(C^p)\) and \(x \in A^{n,m}\). In view of this and of Proposition 2, the error in (14) becomes

\[
e_l(x) = \frac{y_n - y_m}{2} \frac{L_l}{2} (\|x - x_n\| - \|x - x_m\|) \tag{19}
\]

Note that \(|a| \geq |b|\) since \(|b| > |a| \implies x \notin A^{n,m}\) (see the proof of Proposition 3). At the next index sequence, \(l+1\), the \(a\) stays constant, but \(b\) decreases as \(L_{l+1} > L_l\). The absolute error has therefore decreased, i.e., \(|e_l(x)| > |e_{l+1}(x)|\). Analogous arguments can be made if \(x \in \text{int}(C^p)\).

As the Lipschitz estimate changes, the AVD cells are affected. According to the following proposition, given a point in the domain, if the pair of AVD cells that contains it changes from \(l\) to \(l+1\), then the difference between images must have decreased.

**Proposition 5:** Let \(x \in A^{n,m}_{l+1}\) at step \(l\) and \(x \in A^{n,m}_{l+1}\) at step \(l+1\), \(m \neq n\), then necessarily \(y_0 \geq y_n\). Similarly, if \(x \in A^{n,m}_{l+1}\) and \(x \in A^{n,m}_{l+1}\), \(m \neq n\), then necessarily \(y_0 \leq y_m\). If, more specifically, \(x \in A^{n,m}_{l+1}\) and \(x \in B^n_{l+1}\) or \(x \in B^n_{l+1}\), then \(|e_l(x)| > |e_{l+1}(x)|\).

**Proof:** The first case is shown, whereas the second follows by analogy. Given \(x \in A^{n,m}\) and \(x \in A^{n,m}_{l+1}\) and the definitions of the AVD cells in (8), we have that

\[
\|x - x_n\| \leq \|x - x_0\| + \frac{y_0 - y_n}{L_l} \tag{19a}
\]

\[
\|x - x_0\| \leq \|x - x_n\| + \frac{y_n - y_0}{L_{l+1}} \tag{19b}
\]

leading to \(\|x - x_n\| - \frac{y_0 - y_n}{L_l} \leq \|x - x_n\| \leq \|x - x_0\| + \frac{y_n - y_0}{L_{l+1}}\), which is only feasible if \(y_0 \geq y_n\).

The strict error decrease is implied by \(|e_l(x)| > 0, \forall x \in \text{int}(A^{n,m}_{l+1}), \|e_{l+1}(x)\| = 0, \forall x \in B^n_{l+1}\) and \(\forall x \in B^n_{l+1}\).

Even though Proposition 4 and Proposition 5 present local monotonic error decrease results, there might be areas of the domain where \(|e_l(x)|\) can increase. Whether or not the \(L_2\) norm of the function will strictly decrease at every new index depends on its integral over the whole domain. Nevertheless, in the particular case of a univariate model, an additional property can be established.

**Theorem 3:** If the model is univariate, i.e., \(X \subset \mathbb{R}\), then \(f_{\text{NSM}} \to f_{\text{fnN}}\) as \(l \to \infty\) w.r.t. \(\|\cdot\|_2\) monotonically.

**Proof:** Begin by sorting the dataset \(D\) in an increasing order of features \(x_1 \leq x_2 \leq \ldots \leq x_N\). Consider two cases:

**Case 1:** If \(3n, m \in [N]\) such that \(x \) lies in between them, \(x_n \leq x \leq x_m\), then clearly \(x \in C^n\) or \(x \in C^m\). Let \(y_n \leq y_m\). Suppose for the sake of building a contradiction that \(x \in C^p\). If \(x\) is greater than \(x_n\) and \(x_m\), then \(|x - x_p| = |x - x_n| = |x - x_m|\) and \(x \in C^p \implies \|x - x_p\| \leq \|x - x_n\| + \frac{y_m - y_n}{L_{n+1}} \implies \|x - y_n\| \geq L_{n+1}\) is a valid Lipschitz constant, establishing a contradiction. Similar arguments can be made for \(x_p \geq x_m\), and in the case \(x \in C^p\). Thus, \(x\) can only be contained in the AVD cells \(C^n\) and \(C^m\).

**Case 2:** If \(x \leq x_1\), then \(x \in C^1\). Moreover, \(x \in (C^1 \cap C^1)\) since belonging to any other AVD cell would invalidate \(L\), leading to a contradiction. Analogous arguments hold if \(x \geq x_N\). Therefore, \(x\) can only be contained in AVD cells with the same index of its VD cell.

Due to the previous facts, the ND and AVD cells to which any point \(x \in X\) belongs must be the ones associated with either the two feature points \(x_n\) and \(x_m\) that surround it (Case 1), or the smallest/largest point \(x_1/\chi_N\) (Case 2). Hence, if \(x \in A^{n,m}_{l+1}\), then \(x \in A^{n,m}_{l+1}\) or \(x \in B^{n}_{l+1}\) or \(x \in B^{m}_{l+1}\). In any of the three cases, by Proposition 4 and Proposition 5, \(e_l(x) \geq e_{l+1}(x)\). If, on the other hand, \(x \in B^{n}_{l+1}\), \(x \in B^{m}_{l+1}\) and \(e_l(x) = e_{l+1}(x)\). Therefore, for any \(x \in X, e_l(x) \geq e_{l+1}(x) \implies e_l^2(x) \geq e_{l+1}^2(x) \implies\) the convergence is monotonic. \(\Box\)

**IV. DISCUSSION**

Refining Lipschitz estimates of an unknown function is not straightforward in the NSM framework, where the ground-truth \(f\) and clearly also its gradient \(\nabla f\) are assumed to be unknown. This is referred to as the ‘black-box’ case, as opposed to scenarios where either \(f\) or \(\nabla f\) are known [23]. Methods available in the literature usually prove convergence \(L \to L^*\) from below, thus yielding lower bounds instead of upper bounds (see for instance [13], [24]). Other approaches propose the optimization of \(L\) as a decision variable [25]. A statistical procedure for univariate models was proposed in [23];
A Lipschitz constant is a way to quantify how fast a function values change with respect to the input.


\[ f(x) = \cos(x_1) + \sin(x_2), \quad x = [x_1 \quad x_2] \]

was chosen for this example\(^2\) and \( N = 30 \) samples were collected employing a uniform distribution over the domain \( X = [x \in \mathbb{R}^2|0 \leq x \leq [10 \quad 10]| \). Since the domain is compact and convex, the best Lipschitz constant of \( f(x) \) is the maximum value of \( \|\nabla f(x)\| \), which is clearly \( L^* = \sqrt{2} \) using the \( \ell_2 \) norm.

NSM regressors were then computed based on three Lipschitz estimates, \( L = 2, 4, \) and 16, and are shown in the top row of Figure 1. Discrepancy plots between the NSM models and a fixed nearest neighbor regressor, \((f_{\text{NN}} - f_{\text{NSM}})^2\), are instead presented in the bottom row of Figure 1.

It is possible to see in the upper plots of Figure 1 constant ‘plateau’ regions around each sample, that is, the \( B^n \) cells. Those are interfaced by transition areas in which there are gradients of colors, the \( A^n,m \) cells. We moreover note that the blurred regions that separate the data-points become sharper as \( L \) increases, closely resembling a standard Voronoi diagram partition when \( L = 16 \). In the lower plots, the zero-error \( B^n \) areas can be again identified around each sample with \( A^n,m \) cells interfacing them. As \( L \) increases, some regions where \((f_{\text{NN}} - f_{\text{NSM}})^2 \neq 0 \) shrink, preserving however their maximum values. The error \( L^2 \) norm \( \|f_{\text{NN}} - f_{\text{NSM}}\| \) was then estimated by gridding the domain with a fine mesh, and the resulting values for Lipschitz constants ranging from 2 to 30 are shown in Figure 1.

\(^2\)This and an additional univariate example can be found on https://github.com/emilioMaddalena/NSMtoKNN.
in Figure 2. As can be seen in the normalized plot, the discrepancy between models was reduced as the Lipschitz estimate became looser, but this process was not monotonic in this particular example.

VI. CONCLUSION

Loose Lipschitz estimates have the effect of causing the NSM regressor to converge to a nearest neighbor model. The latter is known to exhibit a strong local behavior and to possess high variance error. The looser the \( L \) parameter, the closer the two additively weighted Voronoi diagrams associated with the NSM will be to the NN Voronoi partition. More precisely, the union of the \( B^n \) cells – which are a subset of the VD \( C^n \) cells and where both regressors attain the same values – is expanded. This convergence is not uniform in the domain, as the maximum error value is invariant to Lipschitz constant modifications. Monotonicity is present in the univariate case, but not necessarily in higher dimensions.

REFERENCES

[1] C. Kammer, A. P. Nievergelt, G. E. Fantner, and A. Karimi, “Data-driven controller design for atomic-force microscopy,” IFAC PapersOnLine, vol. 50, no. 1, pp. 10437–10442, 2017.

[2] A. Nicoletti, M. Martino, and A. Karimi, “A data-driven approach to model-reference control with applications to particle accelerator power converters,” Control Eng. Pract., vol. 83, pp. 372–384, 2019.

[3] M. Liu, G. Chowdhary, B. C. Da Silva, S.-Y. Liu, and J. P. How, “Gaussian processes for learning and control: A tutorial with examples,” IEEE Control Syst. Mag., vol. 38, no. 5, pp. 53–86, Oct. 2018.

[4] B. Likar and J. Kocijan, “Predictive control of a gas-liquid separation plant based on a Gaussian process model,” Comput. Chem. Eng., vol. 31, no. 3, pp. 142–152, 2007.

[5] L. Hewing, J. Kabzán, and M. N. Zeilinger, “Cautious model predictive control using Gaussian process regression,” IEEE Trans. Control Syst. Technol., early access, Nov. 21, 2019, doi: 10.1109/TCST.2019.2949757.

[6] A. Carron, E. Arcari, M. Wermelinger, L. Hewing, M. Hutter, and M. N. Zeilinger, “Data-driven model predictive control for trajectory tracking with a robotic arm,” IEEE Robot. Autom. Lett., vol. 4, no. 4, pp. 3758–3765, Oct. 2019.

[7] F. Berkenkamp, A. P. Schoellig, and A. Krause, “Safe controller optimization for quadrupeds with Gaussian processes,” in Proc. IEEE Int. Conf. Robot. Autom. (ICRA), Stockholm, Sweden, 2016, pp. 491–496.

[8] L. Ljung, T. Chen, and B. Mu, “A shift in paradigm for system identification,” Int. J. Control, vol. 95, no. 2, pp. 173–180, 2020.

[9] L. Hewing, K. P. Wabersich, M. Menner, and M. N. Zeilinger, “Learning-based model predictive control: Toward safe learning in control,” Annu. Rev. Control Robot. Auton. Syst., vol. 3, pp. 269–296, Dec. 2019.

[10] J. Ullmann and S. Hirsch, “Feedback linearization based on Gaussian processes with event-triggered online learning,” IEEE Trans. Autom. Control, early access, Dec. 10, 2019, doi: 10.1109/TAC.2019.2949840.

[11] T. Beckers, D. Kulić, and S. Hirsch, “Stable Gaussian process based tracking control of Euler–Lagrange systems,” Automatica, vol. 103, pp. 390–397, May 2019.

[12] M. Milanese and C. Novara, “Set membership identification of nonlinear systems,” Automatica, vol. 40, no. 6, pp. 957–975, 2004.

[13] J.-P. Calliess, “Lazyly adapted constant kinky inference for nonparametric regression and model-reference adaptive control,” 2016. [Online]. Available: arXiv:1701.0017.

[14] A. Blaas, J. Manzano, D. Limon, and J. Calliess, “Localised kinky inference,” in Proc. 18th Eur. Congr. Conf. (ECC), Naples, Italy, 2019, pp. 985–992.

[15] M. Milanese and C. Novara, “Unified set membership theory for identification, prediction and filtering of nonlinear systems,” Automatica, vol. 47, no. 10, pp. 2141–2151, 2011.

[16] C. Novara, A. Nicoli, and G. C. Calafitore, “Nonlinear system identification in Sobolev spaces,” 2019. [Online]. Available: arXiv:1911.02930.

[17] M. Canale and L. Fagiano, “Vehicle yaw control using a fast NMPC approach,” in Proc. 47th IEEE Conf. Decis. Control (CDC), Cancun, Mexico, 2008, pp. 5360–5365.

[18] A. La Bella, L. Fagiano, and R. Scattolini, “Set membership estimation of day-ahead microgrid scheduling,” in Proc. 18th Eur. Congr. Conf. (ECC), Naples, Italy, 2019, pp. 1412–1417.

[19] M. Canale, L. Fagiano, and M. Milanese, “Set membership approximation theory for fast implementation of model predictive control laws,” Automatica, vol. 45, no. 1, pp. 45–54, 2009.

[20] M. Lauricella and L. Fagiano, “Set membership identification of linear systems with guaranteed simulation accuracy,” IEEE Trans. Autom. Control, early access, Jan. 28, 2020, doi: 10.1109/TAC.2020.2970146.

[21] G. A. Bunin, G. François, and D. Bonvin, “Sufficient conditions for feasibility and optimality of real-time optimization schemes—II. Implementation issues,” 2013. [Online]. Available: arXiv:1308.2625.

[22] M. Canale, L. Fagiano, and M. Signorile, “Nonlinear model predictive control from data: A set membership approach,” Int. J. Robust Nonlinear Control, vol. 24, no. 1, pp. 123–139, 2014.

[23] G. Wood and B. Zhang, “Estimation of the Lipschitz constant of a function,” J. Global Optim., vol. 8, no. 1, pp. 91–103, 1996.

[24] A. Blaas, A. D. Cobb, J.-P. Calliess, and S. J. Roberts, “Scalable bounding of predictive uncertainty in regression problems with SLAC,” in Proc. Int. Conf. Scalable Uncertainty Manag., 2018, pp. 373–379.

[25] J.-P. Calliess, “Lipschitz optimisation for Lipschitz interpolation,” in Proc. Amer. Control Conf. (ACC), Seattle, WA, USA, 2017, pp. 3141–3146.

[26] N. S. Altman, “An introduction to kernel and nearest-neighbor nonparametric regression,” Amer. Stat., vol. 46, no. 3, pp. 175–185, 1992.

[27] J. M. Manzano, D. Limon, D. M. de la Peña, and J. Calliess, “Output feedback MPC based on smoothed projected kinky inference,” IET Control Theory Appl., vol. 13, no. 6, pp. 795–805, Apr. 2019.

[28] M. Maddalena and C. Jones, “Learning non-parametric models with guarantees: A smooth Lipschitz interpolation approach,” in Proc. IFAC World Congr., 2020.

[29] S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Y. Wu, “An optimal algorithm for approximate nearest neighbor searching fixed dimensions,” J. ACM, vol. 45, no. 6, pp. 981–923, 1998.

[30] S. Har-Peled and N. Kumar, “Approximating minimization diagrams and generalized proximity search,” SIAM J. Comput., vol. 44, no. 4, pp. 944–974, 2015.