



Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

journal homepage: www.elsevier.com/locate/acha

Stable parameterization of continuous and piecewise-linear functions [☆]

Alexis Goujon ^{*}, Joaquim Campos, Michael Unser*École polytechnique fédérale de Lausanne, Switzerland*

ARTICLE INFO

Article history:

Received 26 September 2022

Received in revised form 13

February 2023

Accepted 2 August 2023

Available online 9 August 2023

Communicated by Radu Balan

Keywords:

Continuous and piecewise-linear function

Stable parameterization

Riesz basis

Condition number

Linear box spline

Hat basis functions

Triangulation

Lipschitz continuity

ReLU neural networks

ABSTRACT

Rectified-linear-unit (ReLU) neural networks, which play a prominent role in deep learning, generate continuous and piecewise-linear (CPWL) functions. While they provide a powerful parametric representation, the mapping between the parameter and function spaces lacks stability. In this paper, we investigate an alternative representation of CPWL functions that relies on local hat basis functions and that is applicable to low-dimensional regression problems. It is predicated on the fact that any CPWL function can be specified by a triangulation and its values at the grid points. We give the necessary and sufficient condition on the triangulation (in any number of dimensions and with any number of vertices) for the hat functions to form a Riesz basis, which ensures that the link between the parameters and the corresponding CPWL function is stable and unique. In addition, we provide an estimate of the $\ell_2 \rightarrow L_2$ condition number of this local representation. As a special case of our framework, we focus on a systematic parameterization of \mathbb{R}^d with control points placed on a uniform grid. In particular, we choose hat basis functions that are shifted replicas of a single linear box spline. In this setting, we prove that our general estimate of the condition number is exact. We also relate the local representation to a nonlocal one based on shifts of a causal ReLU-like function. Finally, we indicate how to efficiently estimate the Lipschitz constant of the CPWL mapping.

© 2023 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Continuous and piecewise-linear functions for supervised learning

The purpose of supervised learning is to reconstruct an unknown mapping from a set of samples [1]. Namely, given a collection of training data pairs $(\mathbf{v}_k, y_k) \in \mathbb{R}^d \times \mathbb{R}$ for $k = 1, \dots, K$, one wants to find $f: \mathbb{R}^d \rightarrow \mathbb{R}$ such that $f(\mathbf{v}_k) \approx y_k$ for $k = 1, \dots, K$, without overfitting. As such, the problem is ill-posed. To make it numerically tractable, a reconstruction space \mathcal{H} is chosen as the image of a finite-dimensional

[☆] This work was supported by the Swiss National Science Foundation, Grant 200020_184646 / 1.

^{*} Corresponding author.

E-mail address: alexis.goujon@epfl.ch (A. Goujon).

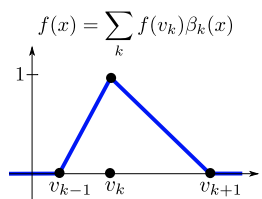
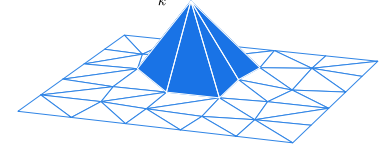
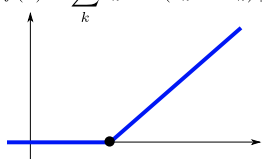
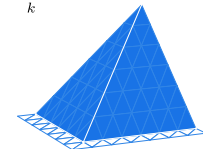
	$f: \mathbb{R} \rightarrow \mathbb{R}$	$f: \mathbb{R}^d \rightarrow \mathbb{R}$
Local	$f(x) = \sum_k f(v_k)\beta_k(x)$ 	$f(\mathbf{x}) = \sum_k f(\mathbf{v}_k)\beta_k(\mathbf{x})$ 
	B-spline (degree 1)	Hat function
Nonlocal	$f(x) = \sum_k \epsilon_k \max(a_k x - b_k)_+$ 	$f(\mathbf{x}) = \sum_k \epsilon_k \max(f_1^k(\mathbf{x}), \dots, f_d^k(\mathbf{x}))_+$ 
	ReLU	Generalized hinging hyperplane

Fig. 1. Local and nonlocal building bricks of CPWL functions, from dimension 1 to any dimension. Notice that the nonlocal basis functions have several equivalent variations; only the ReLU-like one is shown in this figure.

parameter space Θ under a given synthesis operator $T: \Theta \rightarrow \mathcal{H}$. This operator maps a parameter $\theta \in \Theta$ to its continuous representation $T\{\theta\} \in \mathcal{H}$. A celebrated way to choose the synthesis operator is to pick a feedforward neural-network architecture. Given the multidimensional parameter $\theta = (\theta_1, \dots, \theta_{L+1}) \in \Theta$, we then have that

$$T\{\theta\} = (\mathbf{f}_{\theta_{L+1}} \circ \sigma_L \circ \mathbf{f}_{\theta_L} \circ \sigma_{L-1} \circ \dots \circ \sigma_2 \circ \mathbf{f}_{\theta_2} \circ \sigma_1 \circ \mathbf{f}_{\theta_1}), \quad (1)$$

where L is the number of hidden layers of the neural network, $\mathbf{f}_{\theta_k}: \mathbb{R}^{d_k} \rightarrow \mathbb{R}^{d_{k+1}}$ is an affine function parameterized by θ_k , and σ_k is an activation function that is chosen *a priori*. For the model to be able to generate a function space with good approximation properties, the activation functions need to be nonaffine—otherwise, the generated function would remain trivially affine. The pointwise rectified-linear unit (ReLU) $x \mapsto \max(x, 0)$ is one of the most popular activation functions and it usually provides state-of-the-art performance [2,3]. In this case, $T\{\theta\}$ is the composition of continuous piecewise-linear (CPWL) functions, which turns out to be a CPWL function as well [4]. The reverse also holds: any CPWL function $\mathbb{R}^d \rightarrow \mathbb{R}$ can be parameterized by a deep neural network with at most $\lceil \log_2(d+1) \rceil$ hidden layers [5]. The depth of the architecture is instrumental to improve the approximation power of the network [4,6,7] and its generalization ability [8], but it is an obstacle to the control of the model. For instance, to control the Lipschitz constant of a feedforward neural network, state-of-the-art techniques rely on theoretical upper bounds whose tightness degrades each time a new layer is added [9,10]. The depth also makes it hard to determine the role of each parameter in the constructed mapping.

1.2. Linear expansion of continuous and piecewise-linear functions

More interpretable representations of CPWL functions are provided by linear expansions. They boil down to two families: local and nonlocal representations (Fig. 1).

1.2.1. Local representation

In dimension $d = 1$, any CPWL function f can be represented by the linear expansion $f = \sum_k f(v_k)\beta_k$, where the β_k are the underlying triangular B-spline functions [11] (Fig. 1, upper left) and the $\{v_k\}$ are the knots of f , where the slope of f changes. In dimension $d > 1$, the knots are replaced by a triangulation of

the set $\{\mathbf{v}_k\}$ of vertices, which partitions the input domain into simplices. These simplices are the convex hull of a subset of $(d + 1)$ vertices and will serve to define the linear regions of f . Given an appropriate triangulation, any CPWL function can be represented by the linear expansion $f = \sum_k f(\mathbf{v}_k)\beta_k$, where the β_k now denote the hat functions—a.k.a. nodal basis functions or tent-shaped linear basis functions—that correspond to the triangulation [12] (Fig. 1, upper right). Each basis function β_k is defined as the unique CPWL function that satisfies $\beta_k(\mathbf{v}_q) = 1$ for $k = q$ and $\beta_k(\mathbf{v}_q) = 0$ otherwise, and that is affine over each simplex of the triangulation. The β_k are locally supported, hence the attribute *local*. When the vertices are regularly spaced so that they coincide with the sites of a lattice, the hat functions can be chosen as translates of a unique function, namely, a linear box spline [13,14].

There exist recent works that leverage the explicit representation offered by the local parameterization to learn a CPWL function f for regression tasks. Since the partition of f is known and typically not learned, it is possible to regularize the function during training with an explicit regularizer. For instance, one can use the total roughness of f [15–17] or the Hessian total variation of f [18,19], which is not possible with neural networks (NNs). With NNs, it is more common to regularize the parameters, which yields an implicit regularization on the learned mapping that is hard to understand. The global knowledge of the function offered by the local parameterization has, however, a cost. Since the linear regions are fixed, it suffers from the curse of dimensionality. To partially overcome this limitation, it was proposed in [16] to use the local parameterization on subspaces of the features along with ensemble techniques.

1.2.2. Nonlocal representations

In dimension $d = 1$, a CPWL function f with control points v_k can also be represented by the nonlocal representation [20]

$$f(x) = a_0 + a_1(x - v_1) + \sum_{k=2}^{K-1} a_k \text{ReLU}(x - v_k). \tag{2}$$

Note that (2) is one instance out of many nonlocal representations, another one being $f(x) = c_0 + c_1(x - v_1) + \sum_{k=2}^{K-1} c_k |x - v_k|$. The generalization to any dimension is due to Wang and Sun [21]. Their generalized hinging-hyperplanes (GHH) model can represent any CPWL function as $f(\mathbf{x}) = \sum_k \epsilon_k \max(f_1^k(\mathbf{x}), \dots, f_{m_k}^k(\mathbf{x}))$, where $f_1^k, \dots, f_{m_k}^k$ are affine functions, $\epsilon_k = \pm 1$, and $m_k \leq d + 1$. This expansion has also different variations and can be recast as $f(\mathbf{x}) = \sum_k \epsilon_k \max(g_1^k(\mathbf{x}), \dots, g_{m_k-1}^k(\mathbf{x}))_+$, where $g_1^k, \dots, g_{m_k-1}^k$ are affine functions, $\epsilon_k = \pm 1$, $m_k \leq d + 1$, and for any $x \in \mathbb{R}$, $(x)_+ := \max(x, 0) = \text{ReLU}(x)$.

The basis functions of nonlocal representations are the building blocks of many feedforward neural networks. They play the role of activation functions, including ReLU, Leaky ReLU, PReLU, CReLU, and maxout [2,22–25].

1.3. Stability of the parameterizations

The stability of a parameterization can be analyzed through various metrics, including the condition number and the Lipschitz constant of appropriate mappings (see Section 2 for the mathematical definitions).

Condition number of the interpolation problem: Consider the problem of finding the parameter θ such that $T\{\theta\}$ interpolates the data (\mathbf{v}_k, y_k) . For linear models, this amounts to solving a system of linear equations with unknown θ and input (y_k) . The stability of this problem is usually quantified via its condition number, which is the ratio of the largest to the smallest singular value of the system matrix. For the local representation, the explicit solution is $f = \sum_k y_k \beta_k$ in any dimension. The associated condition number is unity and optimal. The situation is less favorable for nonlocal representations. In Fig. 2, we illustrate a case where a small change in a single target value implies a significant change of the nonlocal parameter θ for

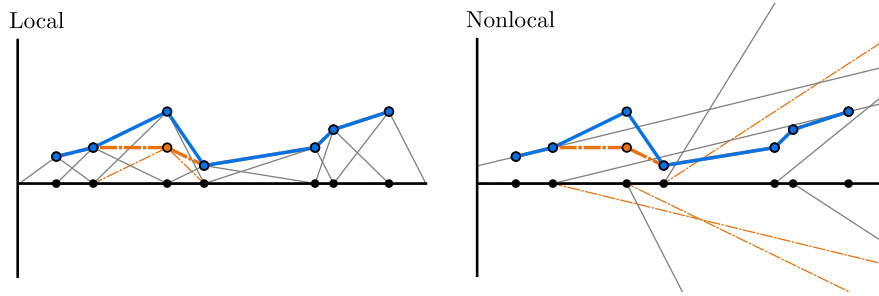


Fig. 2. Linear expansions of two nearly identical CPWL one-dimensional functions (solid thick blue line and mixed thick orange line). The weighted basis functions are represented with thin lines (solid gray and solid orange respectively if different). Left: the local expansions of the two curves are very similar. Right: the small difference between the two curves induces a great difference in their nonlocal parameterization. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

$T\{\boldsymbol{\theta}\}$ to remain interpolatory. In Appendix A, we prove that the interpolation condition number given K data points is at least $\mathcal{O}(K^{3/2})$ for the nonlocal representation (2) in the one-dimensional case. Thus the problem is generally ill-conditioned.

Riesz basis and Riesz ratio: In the context of the local parameterization, we consider the parameter space $\ell_2(\mathbb{Z})$ of finite-energy sequences and equip the function space with the L_2 norm. Although the collection (β_k) of the local atoms is not an orthonormal basis, it forms a Riesz basis under weak conditions [26,27]. This means that (β_k) is the image of an orthonormal basis under a bounded invertible linear operator. This guarantees that the synthesis operator T is a bounded linear bijection from $\ell_2(\mathbb{Z})$ to $\mathcal{H} \subset L_2(\mathbb{R}^d)$ and allows us to define the condition number of T , referred to as the Riesz ratio in the sequel. The Riesz basis is a standard requirement in many signal-processing theories and finite-elements methods [28–32] within a broad spectrum of applications. To the best of our knowledge, the exact Riesz bounds of linear box splines are not known in high dimensions and the case of arbitrary triangulations has not been addressed in full generality so far.

In the context of the nonlocal representation, since the basis functions do not belong to $L_2(\mathbb{R}^d)$, the synthesis operator $T: \Theta \rightarrow \mathcal{H}$ is ill-conditioned, in the sense that a small change of $\boldsymbol{\theta}$ can lead to arbitrarily large changes in $\|T\{\boldsymbol{\theta}\}\|_{L_2}$. In other words, it is not possible to form a Riesz basis with nonlocal CPWL atoms. Consequently, the associated guarantees on stability disappear.

Since the condition number of the interpolation problem is always 1 for the local parameterization, we shall only concentrate on the Riesz ratio in the remainder of the paper, as the latter requires a more involved analysis.

Lipschitz constant: Beyond the stability of the synthesis operator, in many learning applications, it is also highly desirable to control the stability of the input-output mapping. For instance, neural networks with controlled Lipschitz constants tend to generalize better [33–35], to be more robust against adversarial attacks [36–39], and to be more interpretable [38,40]. Despite many recent advances to control the Lipschitz constant of deep models [41–44], it remains very challenging to learn compositional models under a Lipschitz constraint [45]. This results from the fact that the determination of the Lipschitz constant of a ReLU network is already NP-hard as soon as there is one hidden layer [46].

1.4. Contributions and outline

In this paper, we propose to investigate in any dimension the stability of the local parameterization of CPWL functions with the hope to bring detailed results against which other parameterizations could be compared. We define the Riesz ratio in Section 2 and establish a few preliminary results in Section 3. We

then derive in Section 4 an upper bound on the Riesz ratio that is applicable to any triangulation and that considers the worst-case scenario. We also propose a complementary stochastic framework that addresses better an average behavior instead. In Section 5, we show that our upper bound on the Riesz ratio is exact for linear box splines. Finally, in Section 6, we show that the local parameterization gives access to the Lipschitz constant of the map and we provide novel fast-to-evaluate upper and lower bounds on the Lipschitz constant.

2. Mathematical preliminaries

2.1. Simplicial continuous and piecewise-linear functions

Definition 1. A function $f: \mathbb{R}^d \rightarrow \mathbb{R}$ is continuous and piecewise-linear (CPWL) if it is continuous and if there exist distinct affine functions f_1, f_2, \dots, f_M and subsets R_1, R_2, \dots, R_M of \mathbb{R}^d such that

- (i) each R_m is closed with nonempty interior;
- (ii) for $n \neq m$, R_n and R_m have disjoint interiors;
- (iii) the space is partitioned as $\bigcup_{m=1}^M R_m = \mathbb{R}^d$;
- (iv) the function f agrees with f_m on R_m .

We extend this definition to compact input domains of \mathbb{R}^d and to any function $f: \mathbb{R}^d \rightarrow \mathbb{R}$ whose restriction to any compact set is CPWL (sometimes referred to as locally piecewise-affine functions [47]).

Definition 2. A set $\mathcal{V} \subset \mathbb{R}^d$ is locally finite if its intersection with any compact set of \mathbb{R}^d is finite.¹

In the sequel, \mathcal{V} will always denote a locally finite set of \mathbb{R}^d indexed by $I \subset \mathbb{N}$ that is not contained in any $(d - 1)$ -dimensional affine subspace of \mathbb{R}^d . To each point $\mathbf{v}_k \in \mathcal{V}$ we associate a target value $y_k \in \mathbb{R}$. Under Definition 1, it is not clear how to construct a CPWL function f that satisfies $f(\mathbf{v}_k) = y_k$. The local representation offers a more systematic way to address this problem. This requires first to form a triangulation of the set \mathcal{V} .

2.1.1. Partition of the input domain into simplices

A polyhedron is the intersection of finitely many half spaces. A polytope is a bounded polyhedron. Simplices are the polytopes that have the fewest number of faces; in growing number of dimensions $d = 0, \dots, 3$ they include points, segments, triangles, and tetrahedrons. Formally, a d -simplex s of \mathbb{R}^d is the convex hull of $(d + 1)$ affinely independent vertices

$$s = \text{conv}(\mathbf{v}_1, \dots, \mathbf{v}_{d+1}) = \left\{ \sum_{k=1}^{d+1} \lambda_k \mathbf{v}_k : \lambda_k \geq 0, \sum_{k=1}^{d+1} \lambda_k = 1 \right\}. \tag{3}$$

A k -face of a simplex is the convex hull of $(k + 1)$ of its vertices, which is a k -simplex embedded in \mathbb{R}^d . The volume of a simplex admits the explicit form $\text{Vol}(s) = \frac{1}{d!} |\det(\mathbf{v}_2 - \mathbf{v}_1, \dots, \mathbf{v}_{d+1} - \mathbf{v}_1)|$.

Definition 3 (adapted from [48]). A triangulation of a locally finite set $\mathcal{V} \subset \mathbb{R}^d$ of points is a collection \mathcal{S} of d -simplices whose vertices are points in \mathcal{V} and such that

- (i) the union of all the simplices equals $\text{conv}(\mathcal{V})$ (union property);

¹ Note that the meaning of the term “locally finite” depends on the mathematical field.

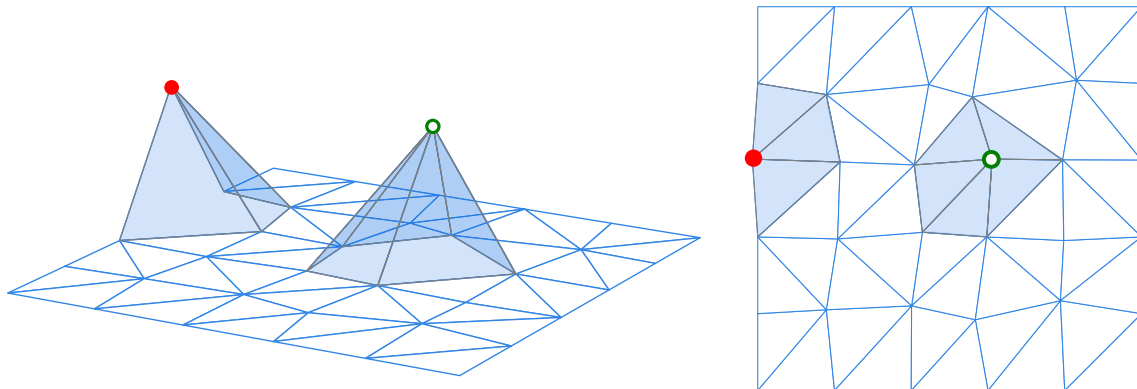


Fig. 3. Left: 3D view of two hat functions on a finite triangulation. Note that, although the left hat function seems discontinuous, it is continuous on the triangulation. Right: star of two vertices of the triangulation (filled area).

(ii) any pair of simplices intersects in a (possibly empty) common face (intersection property).

A triangulation of \mathcal{V} is said to be *full* if all the points of \mathcal{V} are vertices of it, a property that we shall always assume to hold in the sequel. For a given locally finite set $\mathcal{V} \subset \mathbb{R}^d$ of vertices of dimension d , the existence of a triangulation is granted but in general not unique. In practice, most applications fall into one of the two.

- **Regular Sampling Locations:** The vertices coincide with the sites of a lattice, for which explicit triangulations are known (*e.g.*, the Kuhn triangulation [49,50]).
- **Irregular (Random) Sampling Locations:** A Delaunay triangulation always exists in any number of dimensions for a finite set \mathcal{V} , and there exist algorithms to compute it [51,52].

Definition 4. Let \mathcal{S} be a triangulation of a locally finite set $\mathcal{V} \subset \mathbb{R}^d$. The star $\text{St}(\mathbf{v})$ of a vertex $\mathbf{v} \in \mathcal{V}$ is the set of those simplices of \mathcal{S} that contain \mathbf{v} .

We define the volume of the star of a vertex as $\text{Vol}(\text{St}(\mathbf{v})) = \sum_{s \in \text{St}(\mathbf{v})} \text{Vol}(s)$. Its cardinality is denoted by $|\text{St}(\mathbf{v})|$.

2.1.2. Hat basis functions

The set of CPWL functions on a triangulation \mathcal{S} with vertices \mathcal{V} is defined as

$$\text{CPWL}(\mathcal{S}) = \{f \in \mathbb{R}^{\text{conv}(\mathcal{V})} : f \text{ is affine on any } s \in \mathcal{S} \text{ and continuous on } \text{conv}(\mathcal{V})\}. \quad (4)$$

It is said to be the space of linear simplicial splines on \mathcal{S} . Note that the functions in $\text{CPWL}(\mathcal{S})$ are only defined over the convex hull $\text{conv}(\mathcal{V})$ of \mathcal{V} , which can range from a compact set to the whole space \mathbb{R}^d .

It is known that any polyhedron can be partitioned into simplices [53]. As a result, any CPWL function can be viewed as a linear simplicial spline. Two affine functions that coincide on the vertices of a d -simplex are equal, which means that any element of $\text{CPWL}(\mathcal{S})$ is uniquely determined by the values it assumes at the vertices \mathcal{V} of \mathcal{S} . This leads to the local linear expansion

$$\forall f \in \text{CPWL}(\mathcal{S}) : f = \sum_{\mathbf{v} \in \mathcal{V}} f(\mathbf{v}) \beta_{\mathbf{v}}^{\mathcal{S}}, \quad (5)$$

where the hat functions $\beta_{\mathbf{v}}^{\mathcal{S}} \in \mathbb{R}^{\text{conv}(\mathcal{V})}$ (see Fig. 3) are defined on every simplex $s \in \mathcal{S}$ by

$$\beta_{\mathbf{v}|s}^{\mathcal{S}} = \begin{cases} \lambda_{\mathbf{v}}^s, & s \in \text{St}(\mathbf{v}) \\ 0, & \text{otherwise,} \end{cases} \tag{6}$$

where $\lambda_{\mathbf{v}}^s$ is the unique affine function that vanishes at all vertices of s but takes value 1 at vertex \mathbf{v} . In other words, $\lambda_{\mathbf{v}}^s$ outputs the barycentric coordinate of simplex s attached to vertex \mathbf{v} for a given $\mathbf{x} \in s$. Note that depending on the set of vertices, the hat basis functions might not be defined over the whole \mathbb{R}^d and, in the sequel, for any $f \in \mathbb{R}^{\text{conv}(\mathcal{V})}$ we use the notation $\|f\|_{L_p} = (\int_{\mathbf{x} \in \text{conv}(\mathcal{V})} |f(\mathbf{x})|^p d\mathbf{x})^{1/p}$. The hat basis functions have many desirable properties such as

- for $\mathbf{u}, \mathbf{v} \in \mathcal{V}, \beta_{\mathbf{v}}^{\mathcal{S}}(\mathbf{u}) = \begin{cases} 1, & \mathbf{v} = \mathbf{u} \\ 0, & \text{otherwise;} \end{cases}$
- continuity;
- compact support $\text{supp}(\beta_{\mathbf{v}}^{\mathcal{S}}) = \bigcup_{s \in \text{St}(\mathbf{v})} s$;
- minimal support among all nonzero functions of CPWL(\mathcal{S});
- ability to reproduce polynomials of degree up to 1 on $\text{conv}(\mathcal{V})$, so that

$$\forall (\mathbf{a}, b) \in \mathbb{R}^d \times \mathbb{R}, \forall \mathbf{x} \in \text{conv}(\mathcal{V}): \mathbf{a}^T \mathbf{x} + b = \sum_{v \in \mathcal{V}} (\mathbf{a}^T \mathbf{v} + b) \beta_{\mathbf{v}}^{\mathcal{S}}(\mathbf{x}), \tag{7}$$

which includes the partition-of-unity condition $\sum_{v \in \mathcal{V}} \beta_{\mathbf{v}}^{\mathcal{S}} = 1$.

When $\text{St}(\mathbf{v})$ is convex, the hat function simply reads [54]

$$\beta_{\mathbf{v}}^{\mathcal{S}} = \left(\min_{s \in \text{St}(\mathbf{v})} \lambda_{\mathbf{v}}^s \right)_+ . \tag{8}$$

2.2. Riesz basis and Riesz ratio

For a set $I \subset \mathbb{N}$, we denote by $\ell_2(I)$ the set of complex-valued sequences indexed by I with finite energy.

Definition 5. Let \mathcal{H} be a separable Hilbert space over \mathbb{C} and $I \subset \mathbb{N}$. A collection of functions $\{\varphi_k\}_{k \in I}$ in \mathcal{H} is a Riesz basis if

- (i) $\overline{\text{Span}(\{\varphi_k\}_{k \in I})} = \mathcal{H}$ (completeness),
- (ii) there exist $0 < A \leq B < +\infty$ such that, for any $c \in \ell_2(I)$,

$$A\|c\|_{\ell_2} \leq \left\| \sum_{k \in I} c_k \varphi_k \right\|_{L_2} \leq B\|c\|_{\ell_2} \text{ (Riesz-sequence property),} \tag{9}$$

where $\|c\|_{\ell_2} = \left(\sum_{k \in I} |c_k|^2 \right)^{1/2}$.

The tightest constants A and B that satisfy (9) are called the *Riesz bounds*.

Riesz ratio: Consider the linear synthesis operator $T: c \mapsto \sum_{k \in I} c_k \varphi_k$, where $\{\varphi_k\}_{k \in I}$ is a Riesz basis. The synthesis operator T is a bounded linear bijection, which means that there is a unique and stable link between the parameters and the functions being generated. The condition number of the linear operator T is defined as

$$\kappa(T) = \sup_{\mathbf{c}_1, \mathbf{c}_2 \neq \mathbf{0}} \frac{\|T\{\mathbf{c}_1\}\|_{L_2} / \|T\{\mathbf{c}_2\}\|_{L_2}}{\|\mathbf{c}_1\|_{\ell_2} / \|\mathbf{c}_2\|_{\ell_2}} . \tag{10}$$

It follows from (9) that $\kappa(T)$ is the ratio B/A of the Riesz bounds of $\{\varphi_k\}_{k \in I}$. Hence, we refer to this number as the Riesz ratio. The Riesz ratio is 1 if and only if the collection of functions $(\varphi_k)_{k \in I}$ forms an orthonormal basis (up to a scaling factor).

When the collection of functions is formed by the multiindex shifts of a single generating function ($\{\varphi_{\mathbf{k}} = \varphi(\cdot - \mathbf{k}) : \mathbf{k} \in \mathbb{Z}^d\}$), the Riesz-sequence property is characterized via the discrete-time Fourier transform \widehat{g} of the sampled autocorrelation of φ , as given by

$$\widehat{g}: \boldsymbol{\omega} \mapsto \sum_{\mathbf{k} \in \mathbb{Z}^d} \langle \varphi, \varphi(\cdot - \mathbf{k}) \rangle e^{-i\mathbf{k}^T \boldsymbol{\omega}}. \quad (11)$$

In this uniform scenario, the Fourier equivalent of the Riesz-sequence condition is [29,55]

$$0 < A^2 = \operatorname{ess\,inf}_{\boldsymbol{\omega} \in [0, 2\pi]^d} \widehat{g}(\boldsymbol{\omega}) \leq B^2 = \operatorname{ess\,sup}_{\boldsymbol{\omega} \in [0, 2\pi]^d} \widehat{g}(\boldsymbol{\omega}) < +\infty. \quad (12)$$

2.3. Lipschitz constant

A function $f: \mathbb{R}^d \rightarrow \mathbb{R}$ is L_p -Lipschitz, with $L_p \in \mathbb{R}_+$ and $p \geq 1$, if, for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, it holds that

$$|f(\mathbf{x}) - f(\mathbf{y})| \leq L_p \|\mathbf{x} - \mathbf{y}\|_p, \quad (13)$$

and $\operatorname{Lip}_p(f) \in \mathbb{R}_+$ is the smallest constant L_p for which f is L_p -Lipschitz. It is well-known that a CPWL function with finitely many affine pieces is Lipschitz continuous. In particular, the Lipschitz constant of the CPWL function f as defined in Definition 1 is given by the maximum Lipschitz constant of its affine pieces, so that

$$\operatorname{Lip}_p(f) = \max_{m=1, \dots, M} \operatorname{Lip}_p(f_m). \quad (14)$$

3. Affine functions on simplices

Considerations on affine functions on simplices will help us lay the foundations of the analysis of the stability of the local parameterization on simplicial partitions (Sections 4 and 5).

Proposition 1. *Let $f: \mathbb{R}^d \rightarrow \mathbb{C}$ be an affine function and $s = \operatorname{conv}(\mathbf{v}_1, \dots, \mathbf{v}_{d+1})$ a d -simplex. Then*

$$\frac{\operatorname{Vol}(s)}{(d+2)(d+1)} \sum_{k=1}^{d+1} |f(\mathbf{v}_k)|^2 \leq \int_{\mathbf{x} \in s} |f(\mathbf{x})|^2 d\mathbf{x} \leq \frac{\operatorname{Vol}(s)}{(d+1)} \sum_{k=1}^{d+1} |f(\mathbf{v}_k)|^2. \quad (15)$$

Prior to proving Proposition 1, we provide a series of useful results regarding the computation of some integrals of affine functions over simplices.

For a linear function $f: \mathbb{R}^d \rightarrow \mathbb{R}$, an integer $p \in \mathbb{N}$, and a d -simplex $s = \operatorname{conv}(\mathbf{v}_1, \dots, \mathbf{v}_{d+1})$, it is known that [56,57]

$$\int_s f(\mathbf{x})^p d\mathbf{x} = \operatorname{Vol}(s) \binom{p+d}{d}^{-1} \sum_{\mathbf{k} \in \mathbb{N}^{d+1}, |\mathbf{k}|=p} f(\mathbf{v}_1)^{k_1} \cdots f(\mathbf{v}_{d+1})^{k_{d+1}}, \quad (16)$$

where we use the notation $\mathbf{k} = (k_1, \dots, k_{d+1})$ and $|\mathbf{k}| := k_1 + \cdots + k_{d+1}$. We can extend this to affine functions.

Lemma 1. Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be an affine function and $s = \text{conv}(\mathbf{v}_1, \dots, \mathbf{v}_{d+1})$ a d -simplex. For any $p \in \mathbb{N}$, we have that

$$\int_s f(\mathbf{x})^p d\mathbf{x} = \text{Vol}(s) \binom{p+d}{d}^{-1} \sum_{\mathbf{k} \in \mathbb{N}^{d+1}, |\mathbf{k}|=p} f(\mathbf{v}_1)^{k_1} \dots f(\mathbf{v}_{d+1})^{k_{d+1}}. \tag{17}$$

Proof. If the affine function f is not constant, then it can be written as $f(\mathbf{x}) = \mathbf{a}^T(\mathbf{x} - \mathbf{x}_0)$. Equation (16) can be applied after a change of variable, as in

$$\int_s f(\mathbf{x})^p d\mathbf{x} = \int_s (\mathbf{a}^T(\mathbf{x} - \mathbf{x}_0))^p d\mathbf{x} = \int_{s-\mathbf{x}_0} (\mathbf{a}^T \mathbf{y})^p d\mathbf{y} \tag{18}$$

$$= \text{Vol}(s - \mathbf{x}_0) \binom{p+d}{d}^{-1} \sum_{\mathbf{k} \in \mathbb{N}^{d+1}, |\mathbf{k}|=p} (\mathbf{a}^T(\mathbf{v}_1 - \mathbf{x}_0))^{k_1} \dots (\mathbf{a}^T(\mathbf{v}_{d+1} - \mathbf{x}_0))^{k_{d+1}} \tag{19}$$

$$= \text{Vol}(s) \binom{p+d}{d}^{-1} \sum_{\mathbf{k} \in \mathbb{N}^{d+1}, |\mathbf{k}|=p} f(\mathbf{v}_1)^{k_1} \dots f(\mathbf{v}_{d+1})^{k_{d+1}}, \tag{20}$$

where $s - \mathbf{x}_0 := \{\mathbf{x} - \mathbf{x}_0 : \mathbf{x} \in s\}$. If now the affine function f is constant with $f(\mathbf{x}) = b$, then $\int_s f(\mathbf{x})^p d\mathbf{x} = \text{Vol}(s)b^p$. We also have that

$$\text{Vol}(s) \binom{p+d}{d}^{-1} \sum_{\mathbf{k} \in \mathbb{N}^{d+1}, |\mathbf{k}|=p} f(\mathbf{v}_1)^{k_1} \dots f(\mathbf{v}_{d+1})^{k_{d+1}} = \text{Vol}(s) \binom{p+d}{d}^{-1} \sum_{\mathbf{k} \in \mathbb{N}^{d+1}, |\mathbf{k}|=p} b^p = \text{Vol}(s)b^p, \tag{21}$$

where we have used that $\sum_{\mathbf{k} \in \mathbb{N}^{d+1}, |\mathbf{k}|=p} 1 = \binom{p+d}{d}$. This number is known in combinatorics as the combinations with replacement [58]. \square

We can now deduce an important property of the hat functions.

Proposition 2. The L_p norm of the hat function $\beta_{\mathbf{v}}^{\mathcal{S}}$ only depends on the dimension d and the volume of its support. It reads

$$\|\beta_{\mathbf{v}}^{\mathcal{S}}\|_{L_p} = \left(\binom{p+d}{d}^{-1} \text{Vol}(\text{St}(\mathbf{v})) \right)^{1/p}. \tag{22}$$

Proof. We split the integral over the simplices of the support of $\beta_{\mathbf{v}}^{\mathcal{S}}$ and apply Lemma 1, which reads

$$\begin{aligned} \|\beta_{\mathbf{v}}^{\mathcal{S}}\|_{L_p}^p &= \int_{\text{conv}(\mathcal{V})} |\beta_{\mathbf{v}}^{\mathcal{S}}(\mathbf{x})|^p d\mathbf{x} = \int_{\text{conv}(\mathcal{V})} \beta_{\mathbf{v}}^{\mathcal{S}}(\mathbf{x})^p d\mathbf{x} \\ &= \sum_{s \in \mathcal{S}} \int_s \beta_{\mathbf{v}}^{\mathcal{S}}(\mathbf{x})^p d\mathbf{x} = \sum_{s \in \text{St}(\mathbf{v})} \int_s \beta_{\mathbf{v}}^{\mathcal{S}}(\mathbf{x})^p d\mathbf{x} \\ &= \sum_{s \in \text{St}(\mathbf{v})} \binom{p+d}{d}^{-1} \text{Vol}(s) = \binom{p+d}{d}^{-1} \text{Vol}(\text{St}(\mathbf{v})). \quad \square \end{aligned} \tag{23}$$

In the sequel, we shall make use of Proposition 2 through the two following relations:

- the L_2 norm

$$\|\beta_{\mathbf{v}}^S\|_{L_2}^2 = \frac{2\text{Vol}(\text{St}(\mathbf{v}))}{(d+1)(d+2)}; \tag{24}$$

- the inner-product relation

$$\langle \beta_{\mathbf{v}}^S, \sum_{\mathbf{u} \in \mathcal{V}} \beta_{\mathbf{u}}^S \rangle_{\text{conv}(\mathcal{V})} = \langle \beta_{\mathbf{v}}^S, 1 \rangle_{\text{conv}(\mathcal{V})} = \|\beta_{\mathbf{v}}^S\|_{L_1} = \frac{\text{Vol}(\text{St}(\mathbf{v}))}{(d+1)}, \tag{25}$$

where $\langle f, h \rangle_{\text{conv}(\mathcal{V})} = \int_{\mathbf{x} \in \text{conv}(\mathcal{V})} \overline{f(\mathbf{x})}h(\mathbf{x})d\mathbf{x}$ and where the first equality results from the partition of unity of the hat functions.

When $p = 2$, the integral in Lemma 1 is a quadratic form of the value of the function on the vertices and admits the matrix form shown in Lemma 2.

Lemma 2. *Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be an affine function, $s = \text{conv}(\mathbf{v}_1, \dots, \mathbf{v}_{d+1})$ a d -simplex, and $\mathbf{f}_s = (f(\mathbf{v}_1), \dots, f(\mathbf{v}_{d+1})) \in \mathbb{R}^{d+1}$. Then,*

$$\int_s f(\mathbf{x})^2 d\mathbf{x} = \frac{\text{Vol}(s)}{(d+1)(d+2)} \mathbf{f}_s^T \mathbf{P}_{d+1} \mathbf{f}_s, \tag{26}$$

where

$$\mathbf{P}_{d+1} = \mathbf{1}_{d+1} + \mathbf{I}_{d+1} = \begin{bmatrix} 2 & 1 & \dots & 1 \\ 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 1 & \dots & 1 & 2 \end{bmatrix} \in \mathbb{R}^{(d+1) \times (d+1)}, \tag{27}$$

and where $\mathbf{1}_{d+1} \in \mathbb{R}^{(d+1) \times (d+1)}$ is the matrix of ones and $\mathbf{I}_{d+1} \in \mathbb{R}^{(d+1) \times (d+1)}$ is the identity matrix.

Proof. Following Lemma 1, on one hand we have that

$$\begin{aligned} \int_s f(\mathbf{x})^2 d\mathbf{x} &= \text{Vol}(s) \binom{2+d}{d}^{-1} \sum_{\mathbf{k} \in \mathbb{N}^{d+1}, |\mathbf{k}|=2} f(\mathbf{v}_1)^{k_1} \dots f(\mathbf{v}_{d+1})^{k_{d+1}} \\ &= \frac{2\text{Vol}(s)}{(d+1)(d+2)} 1/2 \left(\sum_{p,q=1}^{d+1} f(\mathbf{v}_p)f(\mathbf{v}_q) + \sum_{p=1}^{d+1} f(\mathbf{v}_p)^2 \right) \\ &= \frac{\text{Vol}(s)}{(d+1)(d+2)} \left(\left(\sum_{p=1}^{d+1} f(\mathbf{v}_p) \right)^2 + \sum_{p=1}^{d+1} f(\mathbf{v}_p)^2 \right). \end{aligned} \tag{28}$$

On the other hand, we have that

$$\mathbf{f}_s^T \mathbf{P}_{d+1} \mathbf{f}_s = \sum_{p=1}^{d+1} f(\mathbf{v}_p) \left(\sum_{q=1}^{d+1} f(\mathbf{v}_q) + f(\mathbf{v}_p) \right) \tag{29}$$

$$= \left(\sum_{p=1}^{d+1} f(\mathbf{v}_p) \right)^2 + \sum_{p=1}^{d+1} f(\mathbf{v}_p)^2. \quad \square \tag{30}$$

Lemma 2 can be extended to pairs of complex-valued functions.

Lemma 3. Let $f, g : \mathbb{R}^d \rightarrow \mathbb{C}$ be affine functions, $s = \text{conv}(\mathbf{v}_1, \dots, \mathbf{v}_{d+1})$ a d -simplex, $\mathbf{f}_s = (f(\mathbf{v}_1), \dots, f(\mathbf{v}_{d+1})) \in \mathbb{R}^{d+1}$, and $\mathbf{g}_s = (g(\mathbf{v}_1), \dots, g(\mathbf{v}_{d+1})) \in \mathbb{R}^{d+1}$. It holds that

$$\int_s \overline{f(\mathbf{x})}g(\mathbf{x})d\mathbf{x} = \frac{\text{Vol}(s)}{(d+1)(d+2)}\mathbf{f}_s^H \mathbf{P}_{d+1}\mathbf{g}_s, \tag{31}$$

where

$$\mathbf{P}_{d+1} = \mathbf{1}_{d+1} + \mathbf{I}_{d+1} \in \mathbb{R}^{(d+1) \times (d+1)}. \tag{32}$$

To prove Lemma 3, we first consider real-valued functions f and g and apply Lemma 2 to the left-hand side of the equality $2fg = ((f + g)^2 - f^2 - g^2)$. The announced result is then reached because \mathbf{P}_{d+1} is a symmetric matrix. The generalization to complex-valued functions is directly obtained via the decomposition of f and g into their real and imaginary parts.

Proof of Proposition 1. The matrix \mathbf{P}_{d+1} defined in Lemma 2 is a symmetric circulant matrix generated by the vector $(2, 1, \dots, 1)$. Its eigenvalues are known to be [59]

$$\lambda_m = 2 + \sum_{n=1}^d \zeta_{d+1}^{mn} \quad \text{with } m = 1, \dots, d+1 \text{ and where } \zeta_{d+1} = e^{i\frac{2\pi}{d+1}}. \tag{33}$$

These expressions are further simplified to

$$\lambda_m = 1 + \sum_{n=1}^{d+1} \zeta_{d+1}^{mn} = \begin{cases} d+2, & m = d+1 \\ 1, & \text{otherwise,} \end{cases} \tag{34}$$

which shows that $\min_{m \in \{1, \dots, d+1\}}(\lambda_m) = 1$ and $\max_{m \in \{1, \dots, d+1\}}(\lambda_m) = (d+2)$. As a result, for any $\mathbf{c} \in \mathbb{C}^{d+1}$

$$\|\mathbf{c}\|_2^2 \leq \mathbf{c}^H \mathbf{P}_{d+1} \mathbf{c} \leq (d+2)\|\mathbf{c}\|_2^2. \tag{35}$$

We now conclude by applying Lemma 3. \square

The condition number $\sqrt{d+2}$ given by the inequalities in Proposition 1 depends on the dimension d . However, the eigenvalues of \mathbf{P}_{d+1} are all 1 except for the largest one, given by $(d+2)$. This fact is used to derive Lemma 4, which offers a stochastic view on the problem.

Lemma 4. Let $s = \text{conv}(\mathbf{v}_1, \dots, \mathbf{v}_{d+1})$ be a d -simplex, $\mathbf{C} = (C_1, \dots, C_{d+1})$ a random vector of \mathbb{R}^{d+1} with independent zero-mean components and with $\mathbb{E}(\|\mathbf{C}\|_2^2) \leq +\infty$, and $f_{\mathbf{C}} : \mathbb{R}^d \rightarrow \mathbb{R}$ the unique affine function such that $f_{\mathbf{C}}(\mathbf{v}_k) = C_k$ for $k = 1, \dots, d+1$. Then,

$$\mathbb{E} \left(\int_s f(\mathbf{x})^2 d\mathbf{x} \right) = \frac{2\text{Vol}(s)}{(d+1)(d+2)} \mathbb{E}(\|\mathbf{C}\|_2^2). \tag{36}$$

Proof. Given the matrix \mathbf{P}_{d+1} defined in Lemma 2, it holds that $\mathbf{C}^T \mathbf{P}_{d+1} \mathbf{C} = \sum_{1 \leq k, l \leq d+1} (P_{d+1})_{kl} C_l C_k$. Using the independence of the zero-mean entries of \mathbf{C} , it follows that $\mathbb{E}(\mathbf{C}^T \mathbf{P}_{d+1} \mathbf{C}) = \sum_{k=1}^{d+1} (P_{d+1})_{kk} \mathbb{E}(C_k^2) = 2\mathbb{E}(\|\mathbf{C}\|_2^2)$. The conclusion follows from the formula given in Lemma 2. \square

4. Riesz ratio on arbitrary triangulations

4.1. Triangulations with any number of vertices

Triangulations in high dimensions have complex combinatorial structures that can induce a wide range of behaviors with the usual descriptors, for instance, shape of the simplices, degree of the vertices, number of simplices shared by $2, \dots, d$ vertices. Fortunately, the necessary and sufficient condition that a triangulation must satisfy for the hat functions to form a Riesz basis only relies on the volume of the star of the vertices, as outlined in Theorem 1. This result can be seen as an extension of [60, Theorem 3.1], which gives a similar bound for the condition number of the mass matrix for the hat functions on triangulations with finitely many vertices. The present result applies to any triangulation, including those with infinitely many vertices. This will be used to determine the exact Riesz bounds for linear box splines in Section 5.

Theorem 1. *Let \mathcal{S} be a triangulation of a locally finite set $\mathcal{V} = \{\mathbf{v}_k\}_{k \in I}$ of vertices in \mathbb{R}^d and let $(\beta_{\mathbf{v}}^{\mathcal{S}})_{\mathbf{v} \in \mathcal{V}}$ be the corresponding hat functions. Then, the following statements are equivalent:*

- (i) *the collection of functions $(\beta_{\mathbf{v}}^{\mathcal{S}})_{\mathbf{v} \in \mathcal{V}}$ forms a Riesz basis of the space $\text{CPWL}(\mathcal{S}) \cap L_2(\mathbb{R}^d)$;*
- (ii)
$$\begin{cases} V_{\text{inf}}^{\text{St}} = \inf_{\mathbf{v} \in \mathcal{V}} \text{Vol}(\text{St}(\mathbf{v})) > 0 \\ V_{\text{sup}}^{\text{St}} = \sup_{\mathbf{v} \in \mathcal{V}} \text{Vol}(\text{St}(\mathbf{v})) < +\infty; \end{cases}$$
- (iii)
$$\begin{cases} \inf_{\mathbf{v} \in \mathcal{V}} \|\beta_{\mathbf{v}}^{\mathcal{S}}\|_{L_2} > 0 \\ \sup_{\mathbf{v} \in \mathcal{V}} \|\beta_{\mathbf{v}}^{\mathcal{S}}\|_{L_2} < +\infty. \end{cases}$$

When these statements hold, for any $c \in \ell_2(I)$ we have that

$$\sqrt{\frac{V_{\text{inf}}^{\text{St}}}{(d+1)(d+2)}} \|c\|_{\ell_2} \leq \left\| \sum_{v \in \mathcal{V}} c_v \beta_v^{\mathcal{S}} \right\|_{L_2} \leq \sqrt{\frac{V_{\text{sup}}^{\text{St}}}{(d+1)}} \|c\|_{\ell_2} \tag{37}$$

or, equivalently, that

$$\frac{1}{\sqrt{2}} \inf_{\mathbf{v} \in \mathcal{V}} \|\beta_{\mathbf{v}}^{\mathcal{S}}\|_{L_2} \|c\|_{\ell_2} \leq \left\| \sum_{v \in \mathcal{V}} c_v \beta_v^{\mathcal{S}} \right\|_{L_2} \leq \sqrt{\frac{d+2}{2}} \sup_{\mathbf{v} \in \mathcal{V}} \|\beta_{\mathbf{v}}^{\mathcal{S}}\|_{L_2} \|c\|_{\ell_2}. \tag{38}$$

The Riesz ratio r satisfies

$$r \leq \sqrt{d+2} \sqrt{\frac{V_{\text{sup}}^{\text{St}}}{V_{\text{inf}}^{\text{St}}}} = \sqrt{d+2} \times \frac{\sup_{\mathbf{v} \in \mathcal{V}} \|\beta_{\mathbf{v}}^{\mathcal{S}}\|_{L_2}}{\inf_{\mathbf{v} \in \mathcal{V}} \|\beta_{\mathbf{v}}^{\mathcal{S}}\|_{L_2}}. \tag{39}$$

Proof. The equivalence (ii) \Leftrightarrow (iii) is a direct consequence of (24).

We now show that (i) \Rightarrow (iii). We consider sequences that are zero everywhere, but at one location, and use the Riesz-sequence property to deduce that, for any $\mathbf{v} \in \mathcal{V}$,

$$A \leq \|\beta_{\mathbf{v}}^{\mathcal{S}}\|_{L_2} \leq B, \tag{40}$$

where $0 \leq A \leq B \leq +\infty$ are the Riesz bounds.

We now prove (ii) \Rightarrow (i). Let $c \in \ell_2(I)$, $f = \sum_{\mathbf{v} \in \mathcal{V}} c_{\mathbf{v}} \beta_{\mathbf{v}}^{\mathcal{S}}$. We have that

$$\|f\|_{L_2}^2 = \int_{\text{conv}(\mathcal{V})} |f(\mathbf{x})|^2 d\mathbf{x} = \sum_{s \in \mathcal{S}} \int_s |f(\mathbf{x})|^2 d\mathbf{x}$$

$$\begin{aligned}
 &\leq \frac{1}{(d+1)} \sum_{s \in \mathcal{S}} \sum_{\mathbf{v} \in s \cap \mathcal{V}} \text{Vol}(s) |f(\mathbf{v})|^2 \\
 &= \frac{1}{(d+1)} \sum_{\mathbf{v} \in \mathcal{V}} \sum_{s \in \text{St}(\mathbf{v})} \text{Vol}(s) |f(\mathbf{v})|^2 \\
 &= \frac{1}{(d+1)} \sum_{\mathbf{v} \in \mathcal{V}} |f(\mathbf{v})|^2 \sum_{s \in \text{St}(\mathbf{v})} \text{Vol}(s) \\
 &= \frac{1}{(d+1)} \sum_{\mathbf{v} \in \mathcal{V}} \text{Vol}(\text{St}(\mathbf{v})) |f(\mathbf{v})|^2 \\
 &\leq \frac{\sup_{\mathbf{v} \in \mathcal{V}} (\text{Vol}(\text{St}(\mathbf{v})))}{(d+1)} \|c\|_{\ell_2}^2,
 \end{aligned} \tag{41}$$

where we have applied Proposition 1 and interchanged the order of the double summation with positive arguments (special case of Tonelli’s theorem). Similarly,

$$\begin{aligned}
 \|f\|_{L_2}^2 &= \int_{\text{conv}(\mathcal{V})} |f(\mathbf{x})|^2 d\mathbf{x} = \sum_{s \in \mathcal{S}} \int_s |f(\mathbf{x})|^2 d\mathbf{x} \\
 &\geq \frac{1}{(d+1)(d+2)} \sum_{s \in \mathcal{S}} \sum_{\mathbf{v} \in s \cap \mathcal{V}} \text{Vol}(s) |f(\mathbf{v})|^2 \\
 &= \frac{1}{(d+1)(d+2)} \sum_{\mathbf{v} \in \mathcal{V}} \sum_{s \in \text{St}(\mathbf{v})} \text{Vol}(s) |f(\mathbf{v})|^2 \\
 &= \frac{1}{(d+1)(d+2)} \sum_{\mathbf{v} \in \mathcal{V}} \text{Vol}(\text{St}(\mathbf{v})) |f(\mathbf{v})|^2 \\
 &\geq \frac{\inf_{\mathbf{v} \in \mathcal{V}} (\text{Vol}(\text{St}(\mathbf{v})))}{(d+1)(d+2)} \|c\|_{\ell_2}^2. \quad \square
 \end{aligned} \tag{42}$$

Theorem 1 provides a quantitative way to compare the stability of triangulations. In particular, a triangulation is good when the ratio $\frac{V_{\text{sup}}^{\text{St}}}{V_{\text{inf}}^{\text{St}}}$ is close to 1, which is an indicator of how uniform the triangulation is.

From Theorem 1, we deduce a stronger condition that is sufficient for the Riesz property to hold and that gives further insight into the problem. Since $\text{Vol}(\text{St}(\mathbf{v})) = \sum_{s \in \text{St}(\mathbf{v})} \text{Vol}(s)$, we have that

$$\inf_{s \in \mathcal{S}} \text{Vol}(s) \times \inf_{\mathbf{v} \in \mathcal{V}} |\text{St}(\mathbf{v})| \leq V_{\text{inf}}^{\text{St}} \leq V_{\text{sup}}^{\text{St}} \leq \sup_{s \in \mathcal{S}} \text{Vol}(s) \times \sup_{\mathbf{v} \in \mathcal{V}} |\text{St}(\mathbf{v})|. \tag{43}$$

This means that the hat functions form a Riesz basis whenever the degree of the vertices is upper bounded and the volume of the simplices is upper and lower bounded. This condition, however, is not necessary.

Theorem 1 does not contain direct information on the tightness of the bounds. Yet, we shall prove in Section 5 that, when the hat functions are shifts of a linear box spline, the given bounds are optimal.

The upper bound on the Riesz ratio given in Theorem 1 behaves as $\mathcal{O}(\sqrt{d})$. In Corollary 1, we propose a stochastic perspective that conveys the intuition that the upper Riesz bound B only accounts for very rare behaviors, all the more in high dimensions.

Corollary 1. *Let \mathcal{S} be a triangulation of a locally finite set $\mathcal{V} = \{\mathbf{v}_k\}_{k \in I}$ of vertices in \mathbb{R}^d and let $(\beta_{\mathbf{v}}^{\mathcal{S}})_{\mathbf{v} \in \mathcal{V}}$ be the corresponding hat functions. In addition, suppose that*

$$\begin{cases} V_{\text{inf}}^{\text{St}} = \inf_{\mathbf{v} \in \mathcal{V}} \text{Vol}(\text{St}(\mathbf{v})) > 0 \\ V_{\text{sup}}^{\text{St}} = \sup_{\mathbf{v} \in \mathcal{V}} \text{Vol}(\text{St}(\mathbf{v})) < +\infty. \end{cases} \tag{44}$$

Let $C \in \ell_2(I)$ be a random sequence with independent zero-mean entries and such that $\mathbb{E}(\|C\|_{\ell_2}^2) \leq +\infty$. Then,

$$A^2 \mathbb{E}(\|C\|_{\ell_2}^2) \leq \mathbb{E} \left(\left\| \sum_{\mathbf{v} \in \mathcal{V}} C_{\mathbf{v}} \beta_{\mathbf{v}}^{\mathcal{S}} \right\|_{L_2}^2 \right) \leq 2A^2 \left(\frac{V_{\text{sup}}^{\text{St}}}{V_{\text{inf}}^{\text{St}}} \right) \mathbb{E}(\|C\|_{\ell_2}^2), \tag{45}$$

where $A = \sqrt{\frac{V_{\text{inf}}^{\text{St}}}{(d+1)(d+2)}}$ is the lower Riesz bound of $(\beta_{\mathbf{v}}^{\mathcal{S}})_{\mathbf{v} \in \mathcal{V}}$ given in Theorem 1.

Proof. The inequalities in (45) follow from Lemma 4 combined with the proof of Theorem 1. \square

Within the setting of Corollary 1, it can be inferred from (45) that the standard deviation of the random variable $\| \sum_{\mathbf{v} \in \mathcal{V}} C_{\mathbf{v}} \beta_{\mathbf{v}}^{\mathcal{S}} \|_{L_2}$ is upper bounded by a quantity that varies with d in a similar way as the lower Riesz bound A given in Theorem 1. This yields the intuition that, in this random setting, the typical behavior is better represented by the deterministic lower bound, and that the deterministic upper Riesz bound is only rarely approached in high-dimensional input spaces.

4.2. Triangulations with finitely many vertices

For a triangulation \mathcal{S} of a finite set \mathcal{V} of vertices in d dimensions, the hat functions always form a Riesz basis. Indeed, since we assumed that \mathcal{V} cannot be contained in any $(d - 1)$ -dimensional affine subspace of \mathbb{R}^d , all simplices of \mathcal{S} are not degenerated and the conditions of Theorem 1 are fulfilled. We now apply our formalism to recover a classical result from the finite-element literature, for example [60, Theorem 3.1].

Theorem 2 (see also [60, Theorem 3.1.]). *Let \mathcal{S} be a triangulation of a finite set \mathcal{V} of dimension d of vertices. The corresponding hat functions form a Riesz basis of $\text{CPWL}(\mathcal{S})$ with Riesz bounds*

$$A = \sqrt{\frac{\lambda_{\min}(\mathbf{M})}{(d+1)(d+2)}} \quad \text{and} \quad B = \sqrt{\frac{\lambda_{\max}(\mathbf{M})}{(d+1)(d+2)}}, \tag{46}$$

where the matrix $\mathbf{M} \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{V}|}$ is symmetric and defined as

$$[\mathbf{M}]_{pq} = \begin{cases} 2\text{Vol}(\text{St}(\mathbf{v}_p)), & p = q \\ \text{Vol}(\text{St}(\mathbf{v}_p) \cap \text{St}(\mathbf{v}_q)), & \text{otherwise,} \end{cases} \tag{47}$$

and where $\lambda_{\min}(\mathbf{M})$ and $\lambda_{\max}(\mathbf{M})$ are the smallest and largest eigenvalues of \mathbf{M} .

Proof. Let $f = \sum_{\mathbf{v} \in \mathcal{V}} c_{\mathbf{v}} \beta_{\mathbf{v}}^{\mathcal{S}}$. The coefficients $c_{\mathbf{v}}$ are ordered to form a vector $\mathbf{c} \in \mathbb{R}^{|\mathcal{V}|}$. To each $s \in \mathcal{S}$ we associate a matrix $\mathbf{L}_s \in \mathbb{R}^{(d+1) \times |\mathcal{V}|}$ such that $\mathbf{L}_s \mathbf{c} \in \mathbb{R}^{d+1}$ contains the coefficients associated with the vertices of s . By invoking Lemma 2, we obtain

$$\begin{aligned} \|f\|_{L_2}^2 &= \int_{\text{conv}(\mathcal{V})} |f(\mathbf{x})|^2 d\mathbf{x} = \sum_{s \in \mathcal{S}} \int_s |f(\mathbf{x})|^2 d\mathbf{x} \\ &= \frac{1}{(d+1)(d+2)} \sum_{s \in \mathcal{S}} \text{Vol}(s) (\mathbf{L}_s \mathbf{c})^H \mathbf{P}_{d+1} (\mathbf{L}_s \mathbf{c}) \end{aligned}$$

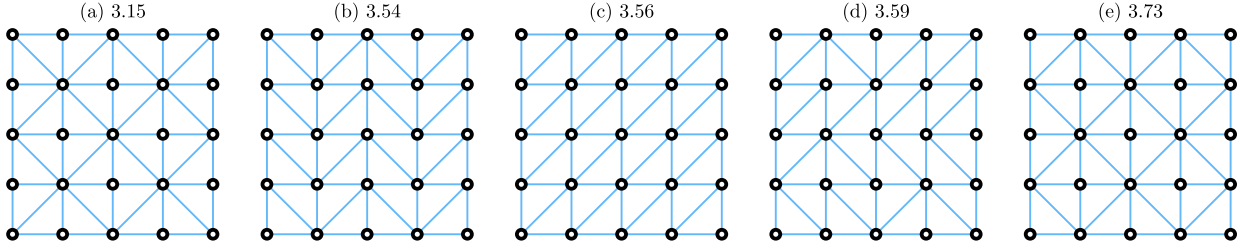


Fig. 4. Riesz ratio of the local parameterization of CPWL functions on various Delaunay triangulations for the same set of vertices. The results stem from Theorem 2 and were obtained numerically. In a) and e), the pattern is similar but the Riesz ratios differ significantly. This comes mainly from the behavior on the border since the smallest star has two simplices for a) while it has a single one for e).

$$\begin{aligned}
 &= \frac{1}{(d+1)(d+2)} \mathbf{c}^H \left(\sum_{s \in \mathcal{S}} \text{Vol}(s) \mathbf{L}_s^T \mathbf{P}_{d+1} \mathbf{L}_s \right) \mathbf{c} \\
 &= \frac{1}{(d+1)(d+2)} \mathbf{c}^H \mathbf{M} \mathbf{c},
 \end{aligned} \tag{48}$$

where $\mathbf{M} = \sum_{s \in \mathcal{S}} \text{Vol}(s) \mathbf{L}_s^T \mathbf{P}_{d+1} \mathbf{L}_s$. Let $1 \leq p, q \leq |\mathcal{V}|$. For $p \neq q$, each entry (p, q) of $\text{Vol}(s) \mathbf{L}_s^T \mathbf{P}_{d+1} \mathbf{L}_s$ is given by

- $\text{Vol}(s)$, if and only if \mathbf{v}_p and \mathbf{v}_q are in s or, equivalently, $s \in \text{St}(\mathbf{v}_p) \cap \text{St}(\mathbf{v}_q)$;
- 0, otherwise.

Now, for $p = q$, each term $\text{Vol}(s) \mathbf{L}_s^T \mathbf{P}_{d+1} \mathbf{L}_s$ of the sum has the entry (p, p) be

- $2\text{Vol}(s)$, if and only if \mathbf{v}_p is in s or, equivalently $s \in \text{St}(\mathbf{v}_p)$;
- 0, otherwise.

This shows that \mathbf{M} is the matrix given in Theorem 2. \square

In the previously cited work [60, Theorem 3.1], the basis functions are attached to interior vertices, which corresponds to zero boundary conditions in our framework. For finite elements, the matrix \mathbf{M} is usually referred to as the mass matrix: it is the Gram matrix of the set of basis functions. For hat basis functions, this matrix has the direct geometric interpretation given in Theorem 2.

For a finite set \mathcal{V} of vertices, one may wonder which triangulation yields the most stable CPWL model in the L_2 sense (*i.e.*, the smallest Riesz ratio). The Delaunay triangulation is known to be optimal in several ways (*e.g.*, in the plane it maximizes the minimum angle of all the angles of the triangles in the triangulation), but it does not necessarily give the smallest Riesz ratio. When the Delaunay triangulation is not unique, the choice can be guided by the related Riesz ratio, as detailed in Fig. 4.

5. Exact Riesz bounds for linear box splines

Throughout this section, we assume that the vertices coincide with the sites of a lattice Λ . This is relevant in some applications, such as image processing [61], but can also be taken advantage of in low-dimensional learning problems [18]. While a uniform grid constrains the model and thereby reduces its expressivity, it significantly improves the computational performance. In this setting, the parameterization of CPWL functions involves shifts of a single linear box spline B for the hat basis functions. The generated space $\text{CPWL}(\mathcal{S}) = \{ \sum_{\mathbf{k} \in \Lambda} c_{\mathbf{k}} B(\cdot - \mathbf{k}) : c_{\mathbf{k}} \in \mathbb{C} \}$ is now integer-shift-invariant and lends itself to the tools of

Fourier analysis [62]. Note that, for any $\mathbf{x} \in \mathbb{R}^d$, the sum $\sum_{\mathbf{k} \in \Lambda} c_{\mathbf{k}} B(\mathbf{x} - \mathbf{k})$ has at most $(d + 1)$ nonzero terms. This follows from the short support of linear box splines.

5.1. Linear box splines

Box splines of any degree have been extensively studied. We refer the reader to the book by de Boor et al. [13] for a general theory and a more comprehensive account. In this paper, we shall concentrate solely on linear box splines, which will in return allow us to derive exact Riesz bounds, which is one of our contributions.

Consider a matrix $\Xi = [\xi_1 \cdots \xi_d] \in \mathbb{R}^{d \times d}$, where (ξ_1, \dots, ξ_d) is a collection of linearly independent vectors of \mathbb{R}^d . The matrix Ξ generates a lattice of \mathbb{R}^d whose sites are $\Xi\mathbb{Z}^d = \{\Xi\mathbf{k} : \mathbf{k} \in \mathbb{Z}^d\}$. Moreover, let $\Xi_{d+1} = [\xi_1 \cdots \xi_{d+1}] \in \mathbb{R}^{d \times (d+1)}$ with $\xi_{d+1} = \sum_{k=1}^d \xi_k$. The linear box spline $B_{\Xi_{d+1}} : \mathbb{R}^d \rightarrow \mathbb{R}$ generated by the collection of vectors $(\xi_1, \dots, \xi_{d+1})$ can be defined via its Fourier transform

$$\widehat{B}_{\Xi_{d+1}}(\omega) = |\det \Xi| \prod_{k=1}^{d+1} \frac{1 - e^{-i\xi_k^T \omega}}{i\xi_k^T \omega}. \tag{49}$$

The normalization factor $|\det \Xi|$ ensures consistency with our definition of a hat function, but is often not included in the literature. The fact that $B_{\Xi_{d+1}}$ is a CPWL function is made explicit with Proposition 3.

Proposition 3.

$$B_{\Xi_{d+1}}(\mathbf{x}) = \sum_{\epsilon \in \{0,1\}^{d+1}} (-1)^{|\epsilon|} \min(\Xi^{-1}(\mathbf{x} - \Xi_{d+1}\epsilon))_+, \tag{50}$$

where $|\epsilon| = \sum_{k=1}^{d+1} \epsilon_k$.

Proof. The product in (49) can be expanded as

$$\widehat{B}_{\Xi_{d+1}}(\omega) = |\det \Xi| \left(\prod_{k=1}^{d+1} \frac{1}{i\xi_k^T \omega} \right) \times \sum_{\epsilon \in \{0,1\}^{d+1}} (-1)^{|\epsilon|} e^{-i(\Xi_{d+1}\epsilon)^T \omega}. \tag{51}$$

By invoking Lemma 7 (Appendix B) and making use of the general Fourier-stretch theorem, we get that

$$\min(\Xi^{-1}\mathbf{x})_+ \xrightarrow{\mathcal{F}} |\det \Xi| \prod_{k=1}^{d+1} \left(\frac{1}{i\xi_k^T \omega} + \pi \delta(\xi_k^T \omega) \right), \tag{52}$$

where \mathbf{x} is the space variable, ω is the pulsation variable and δ is the Dirac distribution. Knowing that $(1 - e^{-i\xi_k^T \omega})\delta(\xi_k^T \omega) = 0$, we observe that (49) has the equivalent form

$$\widehat{B}_{\Xi_{d+1}}(\omega) = |\det \Xi| \left(\prod_{k=1}^{d+1} \frac{1}{i\xi_k^T \omega} + \pi \delta(\xi_k^T \omega) \right) \sum_{\epsilon \in \{0,1\}^{d+1}} (-1)^{|\epsilon|} e^{-i\epsilon^T \Xi_{d+1}^T \omega}. \tag{53}$$

We conclude by taking the inverse Fourier transform on both sides of (53). \square

Proposition 3 is illustrated in dimension $d = 2$ in Fig. 5. This expansion gives a way to prove that the GHH model [21] can represent any linear box spline. In addition, it provides a relation between the local and nonlocal representations, with explicit generalized hinging hyperplanes (namely, the shifts of a single

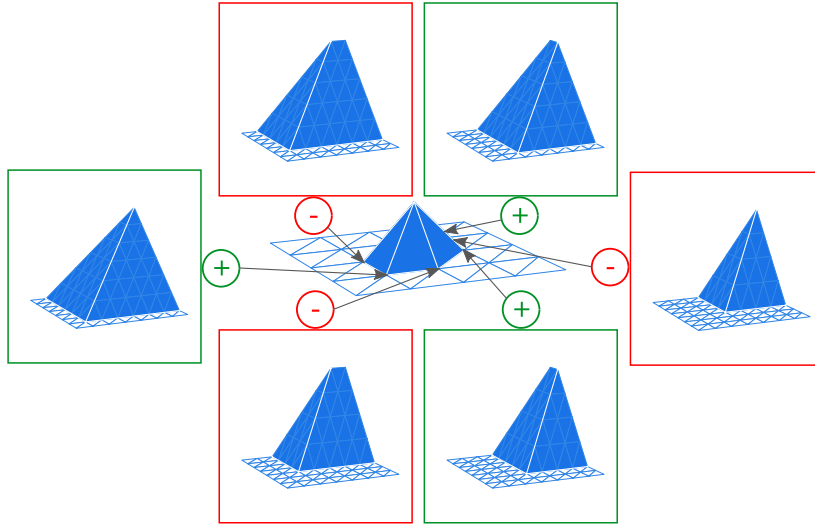


Fig. 5. Linear decomposition of the 2D linear box spline with translates of the nonlocal function $(x, y) \mapsto \min(x, y)_+$. The six basis functions are all translates of $(x, y) \mapsto \min(x, y)_+$. They are only represented here over a finite-size grid, but they are not compactly supported. The location of the corner of each hinge in the central scene is given by the gray arrow.

generalized hinging hyperplane). Our result is close to [63], where a similar formula is proven for three directional box splines of any degree, but only in dimension $d = 2$. In any dimension and for box splines of any degree, a comparable decomposition is given in [64]. Nonetheless, when applied to linear box splines, the expansion in [64] is made of discontinuous Green functions and contains more terms.

The key properties of the linear box spline $B_{\Xi_{d+1}}$ are

- continuity and piecewise linearity (obvious from Proposition 3 since $\mathbf{x} \mapsto \min(\mathbf{x})_+$ is CPWL);
- compact support;
- approximation power of order 2 [13], which means that the reconstruction error of sufficiently smooth decaying functions decreases with the square of the grid size.

5.2. Derivation of the exact Riesz-basis bounds

Prior to giving the general Riesz-basis bounds of linear box splines (Theorem 3), we show an invariance result: the Riesz ratio associated to linear box splines does not depend on the directions (ξ_1, \dots, ξ_d) .

Proposition 4. *Let (ξ_1, \dots, ξ_d) be a family of linearly independent vectors of \mathbb{R}^d , $\xi_{d+1} = \sum_{k=1}^d \xi_k$, $\Xi_{d+1} = [\xi_1 \cdots \xi_{d+1}]$ and $\mathbf{U} \in \mathbb{R}^{d \times d}$ an invertible matrix. Then, the Riesz bounds of $B_{\mathbf{U}\Xi_{d+1}}$ are the ones of $B_{\Xi_{d+1}}$ scaled by $\sqrt{|\det(\mathbf{U})|}$.*

Proof. From Proposition 3, we infer that $B_{\mathbf{U}\Xi_{d+1}} = B_{\Xi_{d+1}} \circ \mathbf{U}^{-1}$. (The generalized version of this relation $M_{\mathbf{U}\Xi} = |\det \mathbf{U}^{-1}| M_{\Xi} \circ \mathbf{U}^{-1}$ is given in [13] and holds true for unnormalized box splines of any degree.) In addition, for any $f \in L_2(\mathbb{R}^d)$, a change of variable allows one to show that $\|f \circ \mathbf{U}^{-1}\|_{L_2}^2 = |\det(\mathbf{U})| \|f\|_{L_2}^2$, which means that, for $c \in \ell_2(\mathbb{Z})$, it holds that

$$\left\| \sum_{\mathbf{k} \in \mathbb{Z}^d} c_{\mathbf{k}} B_{\mathbf{U}\Xi_{d+1}}(\cdot - \mathbf{U}\Xi\mathbf{k}) \right\|_{L_2}^2 = \left\| \sum_{\mathbf{k} \in \mathbb{Z}^d} c_{\mathbf{k}} B_{\Xi_{d+1}} \circ \mathbf{U}^{-1}(\cdot - \mathbf{U}\Xi\mathbf{k}) \right\|_{L_2}^2 = \left\| \sum_{\mathbf{k} \in \mathbb{Z}^d} c_{\mathbf{k}} B_{\Xi_{d+1}}(\mathbf{U}^{-1} \cdot - \Xi\mathbf{k}) \right\|_{L_2}^2 \tag{54}$$

$$= \left\| \left(\sum_{\mathbf{k} \in \mathbb{Z}^d} c_{\mathbf{k}} B_{\Xi_{d+1}}(\cdot - \Xi\mathbf{k}) \right) \circ \mathbf{U}^{-1} \right\|_{L_2}^2 = |\det(\mathbf{U})| \left\| \sum_{\mathbf{k} \in \mathbb{Z}^d} c_{\mathbf{k}} B_{\Xi_{d+1}}(\cdot - \Xi\mathbf{k}) \right\|_{L_2}^2, \tag{55}$$

which allows us to conclude. \square

The fact that $\{B_{\Xi_{d+1}}(\cdot - \Xi \mathbf{k}) : \mathbf{k} \in \mathbb{Z}^d\}$ forms a Riesz basis can be inferred directly from established criterion in the literature [65,66,26,27]. Here, we go further and provide the upper and lower Riesz bounds in any number of dimensions (Theorem 3), which is a new result to the best of our knowledge.

Theorem 3. *Let (ξ_1, \dots, ξ_d) be a basis of \mathbb{R}^d , $\xi_{d+1} = \sum_{k=1}^d \xi_k$, and $\Xi_{d+1} = [\xi_1 \cdots \xi_{d+1}]$. Then, the collection $\{B_{\Xi_{d+1}}(\cdot - \Xi \mathbf{k}) : \mathbf{k} \in \mathbb{Z}^d\}$ of linear box splines forms a Riesz basis with Riesz bounds*

$$A = \sqrt{\frac{|\det \Xi|}{(d+2)}} \quad \text{and} \quad B = \sqrt{|\det \Xi|}. \tag{56}$$

The associated Riesz ratio is $\sqrt{d+2}$.

Proof. Following Proposition 4, we focus solely on the Cartesian lattice \mathbb{Z}^d and denote the corresponding linear box spline by $D = B_{\Xi_{d+1}}$, where Ξ_d is the identity matrix of \mathbb{R}^d . In this case, it is known that the simplices form a Kuhn/Freudenthal triangulation [14]. The vertices of any simplex $s = \text{conv}(\mathbf{v}_0, \dots, \mathbf{v}_d)$ of the triangulation take the form

$$\mathbf{v}_k = \mathbf{v}_0 + \sum_{p=1}^k \mathbf{e}_{\sigma(p)}, \tag{57}$$

where σ is a permutation of the set $\{0, \dots, d\}$, $k = 1, \dots, d$ and \mathbf{e}_i denotes the i th element of the canonical basis of \mathbb{R}^d . Correspondingly, for $p = 0, \dots, d$, there is a unique vertex \mathbf{v}_{k_p} in each simplex such that $|\mathbf{v}_{k_p}| \equiv p \pmod{d+1}$, where for any $\mathbf{x} = (x_1, \dots, x_{d+1}) \in \mathbb{R}^{d+1}$, we use the notation $|\mathbf{x}| = \sum_{k=1}^{d+1} x_k$. Indeed, (57) yields that $|\mathbf{v}_k| = |\mathbf{v}_0| + k$. Then, we use the Fourier characterization (12) to find the Riesz bounds. Specifically, we have to find the essential extrema of

$$\widehat{g}: \boldsymbol{\omega} \mapsto \sum_{\mathbf{k} \in \mathbb{Z}^d} \langle D, D(\cdot - \mathbf{k}) \rangle e^{-i\boldsymbol{\omega}^T \mathbf{k}}. \tag{58}$$

Since the basis function D is compactly supported, the sum is finite and the function $\widehat{g}: \mathbb{R}^d \rightarrow \mathbb{C}$ is continuous and 2π -periodic with respect to each coordinate. We can therefore simply look for the maximum and the minimum of \widehat{g} in the hypercube $[0, 2\pi)^d$. We apply the triangular inequality, use the nonnegativity of D and invoke the partition of unity of the linear box spline to conclude that the maximum of \widehat{g} is attained for $\boldsymbol{\omega} = \mathbf{0}$. With (25) we find the upper bound $B^2 = \frac{\text{Vol}(\text{St}(D))}{d+1}$, where $\text{St}(D)$ is the support of D or, equivalently, the star of the vertex located at $\mathbf{1}$. Now, for the minimum, we evaluate \widehat{g} at $\boldsymbol{\omega}_0 = (-\frac{2\pi}{d+1} \mathbf{1})$ and get

$$\begin{aligned} \sum_{\mathbf{k} \in \mathbb{Z}^d} \langle D, D(\cdot - \mathbf{k}) \rangle e^{-i\boldsymbol{\omega}_0^T \mathbf{k}} &= \langle D, \sum_{\mathbf{k} \in \mathbb{Z}^d} \zeta_{d+1}^{|\mathbf{k}|} D(\cdot - \mathbf{k}) \rangle \\ &= \sum_{s \in \text{St}(D)} \langle D, \sum_{\mathbf{k} \in \mathbb{Z}^d} \zeta_{d+1}^{|\mathbf{k}|} D(\cdot - \mathbf{k}) \rangle_s \\ &= \sum_{s \in \text{St}(D)} \langle D, \sum_{(\mathbf{k}-1) \in \mathbb{Z}^d \cap s} \zeta_{d+1}^{|\mathbf{k}|} D(\cdot - \mathbf{k}) \rangle_s, \end{aligned} \tag{59}$$

where $\langle f, h \rangle_s = \int_{\mathbf{x} \in s} \overline{f(\mathbf{x})} h(\mathbf{x}) d\mathbf{x}$. On one hand, we apply Lemma 3 and use the inherent structure of the Kuhn triangulation displayed in (57) to deduce that

$$\begin{aligned} \min_{\boldsymbol{\omega} \in [0, 2\pi]^d} \widehat{g}(\boldsymbol{\omega}) \leq \widehat{g}(\boldsymbol{\omega}^*) &= \sum_{s \in \text{St}(D)} \frac{1}{(d+1)(d+2)} \text{Vol}(s) [1 \ 0 \ \dots \ 0] \begin{bmatrix} 2 & 1 & \dots & 1 \\ 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 1 & \dots & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ \zeta_{d+1} \\ \vdots \\ \zeta_{d+1}^d \end{bmatrix} \\ &= \frac{\text{Vol}(\text{St}(D))}{(d+1)(d+2)}. \end{aligned} \tag{60}$$

On the other hand, Theorem 1 implies that $\min_{\boldsymbol{\omega} \in [0, 2\pi]^d} \widehat{g}(\boldsymbol{\omega}) \geq \frac{\text{Vol}(\text{St}(D))}{(d+1)(d+2)}$, from which we infer that $A^2 = \frac{\text{Vol}(\text{St}(D))}{(d+1)(d+2)}$. Moreover, we have that $\int_{\mathbb{R}^d} D(\mathbf{x})d\mathbf{x} = \widehat{D}(\mathbf{0}) = \det(\mathbf{I}) = 1$ (from (49)) and, since the box spline D is nonnegative, $\int_{\mathbb{R}^d} D(\mathbf{x})d\mathbf{x} = \|D\|_{L_1} = \frac{\text{Vol}(\text{St}(D))}{(d+1)}$ (from (25)). In short, $\text{Vol}(\text{St}(D)) = (d+1)$, which allows us to derive the result for the Cartesian lattice. The extension to any lattice then follows from Proposition 4. \square

Theorem 1 and 3 yield the same bounds for linear box splines, which shows the relevance of the bounds provided for arbitrary triangulations.

In the proof of Theorem 3 we showed, on one hand, that the volume of the star of a vertex in the Kuhn triangulation is $(d+1)$. On the other hand, the volume of the simplices of this triangulation can be readily computed and amounts to $\frac{1}{d!}$. We deduce that the linear box spline is made of $(d+1)!$ nonzero affine pieces.

From Theorem 3, the Riesz ratio of the linear box-spline parameterization is $\sqrt{d+2}$, which grows with the dimension. However, this metric only reflects extreme cases. A good estimate of the average behavior of the parameterization is given by the mean of the function \widehat{g} (defined in the proof of Theorem 3) over $[0, 2\pi]^d$, computed as

$$\frac{1}{(2\pi)^d} \int_{\boldsymbol{\omega} \in [0, 2\pi]^d} \sum_{\mathbf{k} \in \mathbb{Z}^d} \langle D, D(\cdot - \mathbf{k}) \rangle e^{-i\boldsymbol{\omega}^T \mathbf{k}} d\boldsymbol{\omega} = \langle D, D \rangle = \|D\|_{L_2}^2 = 2 \frac{\text{Vol}(\text{St}(D))}{(d+2)(d+1)} = 2 \min_{\boldsymbol{\omega} \in [0, 2\pi]^d} \widehat{g}(\boldsymbol{\omega}). \tag{61}$$

We infer that \widehat{g} rarely takes values close to its upper bound, especially in high dimensions, which was observed for affine functions on a simplex in Lemma 4. In short, although the Riesz ratio scales badly with the dimension, most linear combinations of box splines will behave in a similar way to the lower bound and the dimension should not significantly degrade the stability of the parameterization.

6. Lipschitz bounds

While the determination of the Lipschitz constant of ReLU networks is NP-hard, the situation is much more favorable for the local parameterization.

6.1. Lipschitz properties of ReLU networks

Consider a one-hidden-layer ReLU network mapping \mathbb{R}^d to \mathbb{R} and defined by

$$h: \mathbf{x} \mapsto \mathbf{u}^T \text{ReLU}(\mathbf{W}^T \mathbf{x} + \mathbf{b}) = \sum_{k=1}^m u_k \text{ReLU}(\mathbf{w}_k^T \mathbf{x} + b_k), \tag{62}$$

where $\mathbf{W} \in \mathbb{R}^{d \times m}$, $\mathbf{u}, \mathbf{b} \in \mathbb{R}^m$ are the parameters, and \mathbf{w}_k are the columns of \mathbf{W} . This network parameterizes a CPWL function; hence, its Lipschitz constant is the maximal Lipschitz constant of its affine pieces. However, the knowledge of these affine pieces is not straightforward as it requires one to determine the linear regions of h . This amounts to finding the nonempty regions of the form

$$\bigcap_{k=1}^m \{\mathbf{x} \in \mathbb{R}^d : \epsilon_k(\mathbf{w}_k^T \mathbf{x} + b_k) \geq 0\}, \tag{63}$$

where $\epsilon_k \in \{0, 1\}$. This geometrical problem is known as “finding the chambers of a hyperplane arrangement”, and a sharp upper bound on the maximal number of these regions for given (d, m) can be found in [67]. The combinatorial aspect of the problem is such that the computation of $\text{Lip}_2(h)$ is NP-hard [46]. In addition, under the exponential-time hypothesis, there exists no polynomial-time approximation algorithm with ratio $\Omega(d^{1-c})$ for the computation of $\text{Lip}_1(h)$ and $\text{Lip}_\infty(h)$ for any constant $c > 0$ [68]. In practice, the strict control of the Lipschitz constant can be achieved by the use of fast-to-compute but loose bounds. The most-common approach relies on the submultiplicativity of the Lipschitz constant for compositions [41,69,70], which yields that

$$\text{Lip}_p(h) \leq \|\mathbf{W}^T\|_{p,p} \|\mathbf{u}\|_q, \tag{64}$$

where the matrix norm $\|\cdot\|_{\alpha,\beta}$ is defined as

$$\|\mathbf{M}\|_{\alpha,\beta} = \sup_{\|\mathbf{x}\|_\alpha=1} \|\mathbf{M}\mathbf{x}\|_\beta. \tag{65}$$

Remark. The (p, p) norm of the row matrix $\mathbf{u}^T \in \mathbb{R}^{1 \times d}$ is the q vector norm $\|\mathbf{u}\|_q$, where $1/p + 1/q = 1$.

6.2. Lipschitz properties of the local parameterization

6.2.1. Lipschitz constant

In comparison to ReLU neural networks, the local parameterization $f = \sum_{\mathbf{v} \in \mathcal{V}} c_{\mathbf{v}} \beta_{\mathbf{v}}^{\mathcal{S}}$ generates CPWL functions with known linear regions, namely, the simplices of \mathcal{S} . The Lipschitz constant of f therefore reads

$$\text{Lip}_p(f) = \sup_{s \in \mathcal{S}} (\|\nabla f|_s\|_q), \tag{66}$$

where $1/p + 1/q = 1$, and where $\nabla f|_s$ is the gradient of f in the interior of the simplex s , hence, a constant vector. Note that when (66) is not finite, then f is not a Lipschitz function. For $d > 1$, the discrete estimate

$$\sup_{\mathbf{v}, \mathbf{u} \in \mathcal{V}} \frac{|c_{\mathbf{v}} - c_{\mathbf{u}}|}{\|\mathbf{v} - \mathbf{u}\|_p} \leq \text{Lip}_p(f), \tag{67}$$

is inadequate, as illustrated in [71, Figure 2], and is generally only a lower bound. The determination of $\text{Lip}_p(f)$ requires some more involved computations. Let $s = \text{conv}(\mathbf{v}_1, \dots, \mathbf{v}_{d+1})$ be a d -simplex, and let $h: \mathbf{x} \mapsto \mathbf{g}_s^T \mathbf{x} + b_s$ be the affine function which satisfies $h(\mathbf{v}_k) = c_k$ for $k = 1, \dots, d + 1$. The gradient \mathbf{g}_s can then be computed as

$$\mathbf{g}_s = (\Delta \mathbf{V}_s^{-T}) \Delta \mathbf{c}, \tag{68}$$

where $\Delta \mathbf{V}_s = [\mathbf{v}_2 - \mathbf{v}_1 \ \dots \ \mathbf{v}_{d+1} - \mathbf{v}_1]$ and $\Delta \mathbf{c} = (c_2 - c_1, \dots, c_{d+1} - c_1)$. The matrix $\Delta \mathbf{V}_s$ is indeed invertible since we have assumed that $\text{Vol}(s) = \frac{1}{d!} |\det(\Delta \mathbf{V}_s)| > 0$. It follows from (66) and (68) that $\text{Lip}_p(f)$ can always be computed in polynomial time for finite partitions.

6.2.2. Lipschitz constant of the hat function

We propose a geometrical interpretation of the Lipschitz constant of the hat basis function, which will then lead us to a practical Lipschitz bound in Proposition 5.

Lemma 5. Let $\beta_{\mathbf{v}}^{\mathcal{S}}$ be the hat basis function attached to the vertex \mathbf{v} . For $s \in \text{St}(\mathbf{v})$, let $f_{s,\mathbf{v}}$ be the facet of s opposed to \mathbf{v} , let $\mathbf{n}_{s,\mathbf{v}}$ be the normal vector of $f_{s,\mathbf{v}}$ pointing toward the interior of s , and let $h_{s,\mathbf{v},p}$ be the p -norm length of the height of s with apex \mathbf{v} . Then,

$$\nabla \beta_{\mathbf{v}}^{\mathcal{S}}|_s = (1/h_{s,\mathbf{v},2})\mathbf{n}_{s,\mathbf{v}}, \text{ and } \text{Lip}_p(\beta_{\mathbf{v}}^{\mathcal{S}}) = \max_{s \in \text{St}(\mathbf{v})} (1/h_{s,\mathbf{v},q}), \tag{69}$$

where $1/p + 1/q = 1$.

Proof. Let $s \in \text{St}(\mathbf{v})$. To simplify the proof, we write $s = \text{conv}(\mathbf{v}_1, \dots, \mathbf{v}_{d+1})$, where $\mathbf{v}_1 = \mathbf{v}$. Following (68), we have that

$$\nabla \beta_{\mathbf{v}}^{\mathcal{S}}|_s = -(\Delta \mathbf{V}_s^{-T})\mathbf{1}. \tag{70}$$

In addition, we have that

$$\Delta \mathbf{V}_s \mathbf{e}_i = \mathbf{v}_{i+1}, \tag{71}$$

where \mathbf{e}_i is the i th element of the canonical basis. Finally, we obtain for any $1 \leq i < j \leq d$

$$\nabla \beta_{\mathbf{v}}^{\mathcal{S}}|_s^T (\mathbf{v}_{i+1} - \mathbf{v}_{j+1}) = -\mathbf{1}^T \Delta \mathbf{V}_s^{-1} \Delta \mathbf{V}_s (\mathbf{e}_i - \mathbf{e}_j) = 0, \tag{72}$$

which means that $\nabla \beta_{\mathbf{v}}^{\mathcal{S}}|_s$ is orthogonal to the facet of s opposed to \mathbf{v} . Let \mathbf{r} be the orthogonal projection of \mathbf{v} onto the facet of s opposed to \mathbf{v} . On the one hand, $\beta_{\mathbf{v}}^{\mathcal{S}}|_s(\mathbf{v}) = 1$ and $\beta_{\mathbf{v}}^{\mathcal{S}}|_s(\mathbf{r}) = 0$. On the other hand, $\mathbf{v} - \mathbf{r}$ is parallel to $\nabla \beta_{\mathbf{v}}^{\mathcal{S}}|_s$, which yields

$$\nabla \beta_{\mathbf{v}}^{\mathcal{S}}|_s = \mathbf{v} - \mathbf{r}, \tag{73}$$

and allows us to conclude the proof. \square

6.2.3. A Lipschitz bound with finite differences and geometrical descriptors

To compute the Lipschitz constant of a CPWL function expressed with the local representation, one has to compute the gradients of the function over each simplex. This requires solving a set of linear problems that are specific to both the geometry (triangulation) and the sequence of coefficients (\mathbf{c}) , as shown in (68). We now propose simpler-to-compute upper and lower bounds. They require solving linear problems that depend only on the geometry and, hence, can be solved offline for a given triangulation and then used for any sequence of coefficients \mathbf{c} (Proposition 5). The bounds highlight the role of the finite differences of the coefficients and the geometry of the simplices.

Proposition 5. Consider the same setting as in Theorem 1. Then,

$$\sup_{\mathbf{v}, \mathbf{u} \in \mathcal{V}} \frac{|c_{\mathbf{v}} - c_{\mathbf{u}}|}{\|\mathbf{v} - \mathbf{u}\|_2} \leq \text{Lip}_2 \left(\sum_{\mathbf{v} \in \mathcal{V}} c_{\mathbf{v}} \beta_{\mathbf{v}}^{\mathcal{S}} \right) \leq \sup_{s \in \mathcal{S}} \sum_{\mathbf{v} \in s \cap \mathcal{V}} \frac{|c_{\mathbf{v}} - \alpha_s|}{h_{s,\mathbf{v},2}}, \tag{74}$$

for any $\alpha_s \in \mathbb{R}$ for $s \in \mathcal{S}$, and where $h_{s,\mathbf{v},2}$ is the height of simplex s with apex \mathbf{v} .

Proof. Let $f = \sum_{\mathbf{v} \in \mathcal{V}} c_{\mathbf{v}} \beta_{\mathbf{v}}^{\mathcal{S}}$. For any $s \in \mathcal{S}$, it holds that

$$\nabla f|_s = \sum_{\mathbf{v} \in s \cap \mathcal{V}} c_{\mathbf{v}} \nabla \beta_{\mathbf{v}}^{\mathcal{S}}|_s. \tag{75}$$

Hence,

$$\|\nabla f|_s\|_2 \leq \sum_{\mathbf{v} \in s \cap \mathcal{V}} |c_{\mathbf{v}}| \|\nabla \beta_{\mathbf{v}}^S|_s\|_2 = \sum_{\mathbf{v} \in s \cap \mathcal{V}} \frac{|c_{\mathbf{v}}|}{h_{s,\mathbf{v},2}}. \quad (76)$$

We now note that

$$\nabla \left(\sum_{\mathbf{v} \in s \cap \mathcal{V}} \beta_{\mathbf{v}}^S \right) |_s = \nabla(1)|_s = \mathbf{0}. \quad (77)$$

Hence,

$$\nabla f|_s = \sum_{\mathbf{v} \in s \cap \mathcal{V}} c_{\mathbf{v}} \nabla \beta_{\mathbf{v}}^S|_s = \sum_{\mathbf{v} \in s \cap \mathcal{V}} (c_{\mathbf{v}} - \alpha_s) \nabla \beta_{\mathbf{v}}^S|_s, \quad (78)$$

for any $\alpha_s \in \mathbb{R}$, which allows us to conclude. \square

The bound given in Proposition 5 requires one to choose a reference value α_s for each simplex s against which to compare the value of the function at the vertices. In the case where $h_{s,\mathbf{v},2}$ does not depend on \mathbf{v} , then the choice of α_s as the median of $\{c_{\mathbf{v}}\}$ for $\mathbf{v} \in s \cap \mathcal{V}$ leads to the tightest bound. More generally, a suitable choice is to pick $\alpha_s \in [\min_{\mathbf{v} \in s \cap \mathcal{V}} c_{\mathbf{v}}, \max_{\mathbf{v} \in s \cap \mathcal{V}} c_{\mathbf{v}}]$, which ensures an upper bound of 0 for constant functions. Note that this choice yields an exact bound when $d = 1$, but this is not necessarily true for $d > 1$.

7. Conclusion

We have provided a measure of the stability of the parameterization of CPWL functions with hat basis functions on any triangulation. First, we have estimated the $\ell_2 \rightarrow L_2$ condition number of the parameterization in full generality. We have found that it is mainly determined by the relative volume of the star of the vertices of the triangulation, namely, the relative size of the support of the hat functions. This result is in accordance with the literature on finite-element methods, which however only focus on the case of finitely many vertices. When the vertices lie at the sites of a lattice, we parameterize the CPWL functions with linear box splines. We have provided a formula to relate these local basis functions to nonlocal ReLU-like functions. In this uniform setting, we have proved that the exact condition number only depends on the dimension d and is $\sqrt{d+2}$. Although it increases with the dimension, we noticed that, from a stochastic point of view, the dimension might only rarely affect the stability of the local representation. Eventually, we have shown that the Lipschitz constant of a CPWL function can be easily determined when its local parameterization is known. This highly contrasts with CPWL functions defined by ReLU neural networks.

Data availability

No data was used for the research described in the article.

Appendix A. Interpolation condition number of the nonlocal parameterization

Consider the nonlocal parameterization of one-dimensional CPWL functions with knots $v_1 < \dots < v_K \in \mathbb{R}$

$$T\{\boldsymbol{\theta}\}(x) = \theta_1 + \theta_2(x - v_1) + \sum_{k=2}^{K-1} \theta_{k+1}(x - v_k)_+. \quad (79)$$

Given target values $y_1, \dots, y_K \in \mathbb{R}$, the parameters $\theta_1, \dots, \theta_K \in \mathbb{R}$ have to satisfy for $p = 1, \dots, K$ that

$$y_p = T\{\boldsymbol{\theta}\}(v_p) = \theta_1 + \theta_2(v_p - v_1) + \sum_{k=2}^{K-1} \theta_{k+1}(v_p - v_k)_+, \tag{80}$$

which, assuming a constant step size $h = (v_{k+1} - v_k)$, further simplifies in $y_p = \theta_1 + h \sum_{k=1}^{p-1} \theta_{k+1}(p - k)$. This yields the matrix equation

$$\boldsymbol{\theta} = \frac{1}{h} \begin{bmatrix} 1/h & 0 & \cdots & \cdots & 0 \\ 1/h & 1 & 0 & \cdots & 0 \\ \vdots & 2 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 1/h & (K-1) & \cdots & 2 & 1 \end{bmatrix}^{-1} \mathbf{y} = \mathbf{M}^{-1} \mathbf{y}. \tag{81}$$

To give a lower bound to the condition number of the problem, we remark that $\mathbf{M}\mathbf{e}_2 = h \sum_{k=1}^{K-1} k\mathbf{e}_{k+1}$ and $\mathbf{M}\mathbf{e}_K = h\mathbf{e}_K$, where \mathbf{e}_k are the canonical vectors of \mathbb{R}^d . We infer that $\|\mathbf{M}^{-1} \sum_{k=1}^{K-1} k\mathbf{e}_k\|_2 / \|\sum_{k=1}^{K-1} k\mathbf{e}_k\|_2 = (K(K-1)(2K-1)/6)^{-1/2}/h$ and that $\frac{\|\mathbf{M}^{-1}\mathbf{e}_K\|_2}{\|\mathbf{e}_K\|_2} = 1/h$. It implies that the ℓ_2 condition number r of the problem satisfies

$$r = \max_{\mathbf{a}, \mathbf{b} \in \mathbb{R}^d \setminus \{0\}} \left\{ \frac{\|\mathbf{M}^{-1}\mathbf{a}\|}{\|\mathbf{a}\|} \frac{\|\mathbf{b}\|}{\|\mathbf{M}^{-1}\mathbf{b}\|} \right\} \geq \sqrt{\frac{K(K-1)(2K-1)}{6}}. \tag{82}$$

Appendix B. The generalized hinging-hyperplane generating function of linear box splines

The Fourier transform \widehat{h} of the Heaviside function $h: x \mapsto \begin{cases} 1, & x \geq 0 \\ 0, & \text{otherwise,} \end{cases}$ is given by

$$\widehat{h}: \omega \mapsto \frac{1}{i\omega} + \pi\delta(\omega). \tag{83}$$

From this we infer \widehat{H} , the Fourier transform of $H: \mathbf{x} \mapsto \prod_{k=1}^d h(x_k)$, the separable version of h in d dimensions,

$$\widehat{H}: \boldsymbol{\omega} \mapsto \prod_{k=1}^d \left(\frac{1}{i\omega_k} + \pi\delta(\omega_k) \right). \tag{84}$$

We define a directional Heaviside wall function as $W: \mathbf{x} \mapsto h(x_d) \prod_{k=1}^{d-1} \delta(x_k)$ and the matrix $\mathbf{A}_d =$

$$\begin{bmatrix} 1 & -1 & & & \\ & \ddots & \ddots & & \\ & & & 1 & -1 \\ 1 & \cdots & \cdots & & 1 \end{bmatrix}.$$

Lemma 6.

$$\forall \mathbf{x} \in \mathbb{R}^d, (H * (W \circ \mathbf{A}_d))(\mathbf{x}) = \min(\mathbf{x})_+. \tag{85}$$

Proof. We have that

$$(H * (W \circ \mathbf{A}_d))(\mathbf{x}) = \int_{\mathbb{R}^d} H(\mathbf{x} - \mathbf{t})W(\mathbf{A}_d\mathbf{t})d\mathbf{t}. \tag{86}$$

In the sequel, we take advantage of two properties.

- We prove that $\det(\mathbf{A}_d) = d$ by induction using the Laplace/cofactor expansion along the last column.
- We observe that $\mathbf{A}_d \mathbf{1} = d\mathbf{e}_d$, which in turn implies that $\mathbf{A}_d^{-1} \mathbf{e}_d = d^{-1} \mathbf{1}$.

We now use the change of variable $\mathbf{y} = \mathbf{A}_d \mathbf{t}$. It yields that

$$\begin{aligned}
 (H * (W \circ \mathbf{A}_d))(\mathbf{x}) &= (1/d) \int_{\mathbb{R}^d} H(\mathbf{x} - \mathbf{A}_d^{-1} \mathbf{y}) h(y_d) \prod_{k=1}^{d-1} \delta(y_k) d\mathbf{y} \\
 &= (1/d) \int_{\mathbb{R}^d} H(\mathbf{x} - y_d \mathbf{A}_d^{-1} \mathbf{e}_d) h(y_d) \prod_{k=1}^{d-1} \delta(y_k) d\mathbf{y} \\
 &= (1/d) \int_{\mathbb{R}^d} H(\mathbf{x} - \frac{y_d}{d} \mathbf{1}) h(y_d) \prod_{k=1}^{d-1} \delta(y_k) d\mathbf{y} \\
 &= (1/d) \int_{\mathbb{R}^d} h(y_d) \left(\prod_{k=1}^d h(x_k - y_d/d) \right) \left(\prod_{k=1}^{d-1} \delta(y_k) \right) d\mathbf{y} \\
 &= (1/d) \int_{\mathbb{R}} h(y_d) \left(\prod_{k=1}^d h(x_k - y_d/d) \right) dy_d.
 \end{aligned} \tag{87}$$

The quantity $\prod_{k=1}^d h(x_k - y_d/d)h(y_d)$ is nonzero when $y_d > 0$ and $y_d < dx_k$ for $k = 1, \dots, d$, which is equivalent to $y_d > 0$ and $y_d < d \min(x_k)$. We can now conclude that $(H * (L \circ \mathbf{A}_d))(\mathbf{x}) = \min(\mathbf{x})_+$. \square

Lemma 7.

$$\min(\mathbf{x})_+ \xrightarrow{\mathcal{F}} \prod_{k=1}^{d+1} \left(\frac{1}{i\omega_k} + \pi\delta(\omega_k) \right), \tag{88}$$

where $\omega_{d+1} = \sum_{k=1}^d \omega_k$.

Proof. From Lemma 6, we have that $(H * (W \circ \mathbf{A}_d))(\mathbf{x}) = \min(\mathbf{x})_+$. The function W is separable and its Fourier transform reads $\widehat{W}(\boldsymbol{\omega}) = \frac{1}{i\omega_d} + \pi\delta(\omega_d)$. In addition, the general stretch theorem implies that

$$(W \circ \mathbf{A}_d) \xrightarrow{\mathcal{F}} \frac{1}{d} \widehat{W}(\mathbf{A}_d^{-T} \boldsymbol{\omega}) = \frac{1}{d} \left(\frac{1}{i\mathbf{e}_d^T \mathbf{A}_d^{-T} \boldsymbol{\omega}} + \pi\delta(\mathbf{e}_d^T \mathbf{A}_d^{-T} \boldsymbol{\omega}) \right). \tag{89}$$

Now, we use that $\mathbf{A}_d^{-1} \mathbf{e}_d = d^{-1} \mathbf{e}_d$ and the effect of a dilation on the Dirac distribution to conclude that

$$(W \circ \mathbf{A}_d) \xrightarrow{\mathcal{F}} \frac{1}{d} \widehat{W}(\mathbf{A}_d^{-T} \boldsymbol{\omega}) = \left(\frac{1}{i\omega_{d+1}} + \pi\delta(\omega_{d+1}) \right). \tag{90}$$

We conclude by using the Fourier transform of H and by transforming the convolution into a product in the Fourier domain. \square

References

- [1] C.M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
- [2] Y. Lecun, Y. Bengio, G. Hinton, Deep learning, *Nature* 521 (7553) (2015) 436–444, <https://doi.org/10.1038/nature14539>.
- [3] X. Glorot, A. Bordes, Y. Bengio, Deep sparse rectifier neural networks, in: *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics*, vol. 15, 2011, pp. 315–323.
- [4] G.F. Montúfar, R. Pascanu, K. Cho, Y. Bengio, On the number of linear regions of deep neural networks, in: *Proceedings of the 27th Conference on Advances in Neural Information Processing Systems*, vol. 27, Montréal, Canada, 2014.
- [5] R. Arora, A. Basu, P. Mianjy, A. Mukherjee, Understanding deep neural networks with rectified linear units, in: *6th International Conference on Learning Representations, ICLR 2018*, 2018, arXiv:1611.01491.
- [6] R. Eldan, O. Shamir, The power of depth for feedforward neural networks, in: *29th Annual Conference on Learning Theory*, vol. 49, PMLR, Columbia University, New York, New York, USA, 2016, pp. 907–940, <https://proceedings.mlr.press/v49/eldan16.html>.
- [7] H. Mhaskar, T. Poggio, Deep vs. shallow networks: an approximation theory perspective, *Anal. Appl.* 14 (06) (2016) 829–848, arXiv:1608.03287.
- [8] T. Poggio, H. Mhaskar, L. Rosasco, B. Miranda, Q. Liao, Why and when can deep—but not shallow – networks avoid the curse of dimensionality: a review, *Int. J. Autom. Comput.* 14 (5) (2017) 503–519, arXiv:1611.00740.
- [9] H. Gouk, E. Frank, B. Pfahringer, M.J. Cree, Regularisation of neural networks by enforcing Lipschitz continuity, *Mach. Learn.* 110 (2) (2021) 393–416, <https://doi.org/10.1007/s10994-020-05929-w>, arXiv:1804.04368.
- [10] K. Scaman, A. Virmaux, Lipschitz regularity of deep neural networks: analysis and efficient estimation, *Adv. Neural Inf. Process. Syst.* 1 (2018) 3835–3844, arXiv:1805.10965.
- [11] C. de Boor, *Splines as Linear Combinations of B-Splines. A Survey*, *Approximation Theory*, 1976.
- [12] J. He, L. Li, J. Xu, C. Zheng, ReLU deep neural networks and linear finite elements, *J. Comput. Math.* 38 (3) (2020) 502–527, <https://doi.org/10.4208/JCM.1901-M2018-0160>.
- [13] C. De Boor, K. Höllig, S. Riemenschneider, *Box Splines*, vol. 98, Springer Science & Business Media, 1993.
- [14] M. Kim, J. Peters, Symmetric box-splines on the An^* lattice, *J. Approx. Theory* 162 (9) (2010) 1607–1630, <https://doi.org/10.1016/j.jat.2010.04.007>.
- [15] Y. Liu, G. Yin, Nonparametric functional approximation with Delaunay triangulation learner, in: *2019 IEEE International Conference on Big Knowledge (ICBK)*, 2019, pp. 167–174.
- [16] Y. Liu, G. Yin, The Delaunay triangulation learner and its ensembles, *Comput. Stat. Data Anal.* 152 (2020) 107030, <https://doi.org/10.1016/j.csda.2020.107030>, <https://www.sciencedirect.com/science/article/pii/S0167947320301213>.
- [17] J. Gu, G. Yin, Crystallization learning with the Delaunay triangulation, in: M. Meila, T. Zhang (Eds.), *Proceedings of the 38th International Conference on Machine Learning*, in: *Proceedings of Machine Learning Research*, vol. 139, PMLR, 2021, pp. 3854–3863, <https://proceedings.mlr.press/v139/gu21a.html>.
- [18] J. Campos, S. Aziznejad, M. Unser, Learning of continuous and piecewise-linear functions with Hessian total-variation regularization, *IEEE Open J. Signal Process.* 3 (2022) 36–48, <https://doi.org/10.1109/OJSP.2021.3136488>.
- [19] M. Pourya, A. Goujon, M. Unser, Delaunay-triangulation-based learning with Hessian total-variation regularization, *IEEE Open J. Signal Process.* 4 (2023) 167–178, <https://doi.org/10.1109/OJSP.2023.3250104>, <https://ieeexplore.ieee.org/abstract/document/10054490>.
- [20] C.D. Aliprantis, D. Harris, R. Tourky, Continuous piecewise linear functions, *Macrocon. Dyn.* 10 (1) (2006) 77.
- [21] S. Wang, X. Sun, Generalization of hinging hyperplanes, *IEEE Trans. Inf. Theory* 51 (12) (2005) 4425–4431, <https://doi.org/10.1109/TIT.2005.859246>.
- [22] A.L. Maas, A.Y. Hannun, A.Y. Ng, Rectifier Nonlinearities Improve Neural Network Acoustic Models, *Proceedings of the 30th International Conference on Machine Learning*, vol. 30, Citeseer, 2013, p. 3.
- [23] K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectifiers: surpassing human-level performance on ImageNet classification, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1026–1034.
- [24] W. Shang, K. Sohn, D. Almeida, H. Lee, Understanding and improving convolutional neural networks via concatenated rectified linear units, in: *33rd International Conference on Machine Learning*, 2016, pp. 2217–2225, arXiv:1603.05201.
- [25] I.J. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, Y. Bengio, Maxout networks, in: *30th International Conference on Machine Learning*, vol. 3, International Machine Learning Society (IMLS), 2013, pp. 2356–2364, arXiv:1302.4389.
- [26] W. Dahmen, C.A. Micchelli, Translates of multivariate splines, *Linear Algebra Appl.* 52 (1983) 217–234.
- [27] Y. Guan, S. Lu, Y. Tang, Characterization of compactly supported refinable splines whose shifts form a Riesz basis, *J. Approx. Theory* 133 (2) (2005) 245–250.
- [28] A. Aldroubi, M. Unser, Sampling procedures in function spaces and asymptotic equivalence with Shannon’s sampling theory, *Numer. Funct. Anal. Optim.* 15 (1–2) (1994) 1–21, <https://doi.org/10.1080/01630569408816545>.
- [29] A. Aldroubi, Oblique projections in atomic spaces, *Proc. Am. Math. Soc.* 124 (7) (1996) 2051–2060, <https://doi.org/10.1090/S0002-9939-96-03255-8>.
- [30] M. Unser, I. Daubechies, On the approximation power of convolution-based least squares versus interpolation, *IEEE Trans. Signal Process.* 45 (7) (1997) 1697–1711, <https://doi.org/10.1109/78.599940>.
- [31] R. Jia, W. Zhao, Riesz bases of wavelets and applications to numerical solutions of elliptic equations, *Math. Comput.* 80 (275) (2011) 1525–1556.
- [32] N. Fukuda, T. Kinoshita, T. Kubo, On the finite element method with Riesz bases and its applications to some partial differential equations, in: *2013 10th International Conference on Information Technology: New Generations*, 2013, pp. 761–766.
- [33] P.L. Bartlett, D.J. Foster, M. Telgarsky, Spectrally-normalized margin bounds for neural networks, in: *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS’17*, Curran Associates Inc., Red Hook, NY, USA, 2017, pp. 6241–6250.

- [34] U. von Luxburg, O. Bousquet, Distance-based classification with Lipschitz functions, *J. Mach. Learn. Res.* 5 (2004) 669–695.
- [35] J. Sokolić, R. Giryes, G. Sapiro, M.R.D. Rodrigues, Robust large margin deep neural networks, *IEEE Trans. Signal Process.* 65 (16) (2017) 4265–4280.
- [36] M. Cisse, P. Bojanowski, E. Grave, Y. Dauphin, N. Usunier, Parseval networks: improving robustness to adversarial examples, in: *International Conference on Machine Learning*, PMLR, 2017, pp. 854–863.
- [37] P. Hagemann, S. Neumayer, Stabilizing invertible neural networks using mixture models, *Inverse Probl.* 37 (8) (2021) 85002.
- [38] D. Tsipras, S. Santurkar, L. Engstrom, A. Turner, A. Madry, Robustness may be at odds with accuracy, in: *7th International Conference on Learning Representations, ICLR 2019*, New Orleans, LA, USA, May 6–9, 2019, OpenReview.net, 2019, <https://openreview.net/forum?id=SyxAb30cY7>.
- [39] Y. Tsuzuku, I. Sato, M. Sugiyama, Lipschitz-Margin Training: Scalable Certification of Perturbation Invariance for Deep Neural Networks, *Advances in Neural Information Processing Systems*, vol. 31, Curran Associates, Inc., 2018, pp. 6542–6551.
- [40] A.S. Ross, F. Doshi-Velez, Improving the adversarial robustness and interpretability of deep neural networks by regularizing their input gradients, *Tech. Rep.*, arXiv:1711.09404, Nov. 2017, <http://arxiv.org/abs/1711.09404>.
- [41] T. Miyato, T. Kataoka, M. Koyama, Y. Yoshida, Spectral normalization for generative adversarial networks, in: *International Conference on Learning Representations*, 2018, pp. 1–26, <https://openreview.net/forum?id=B1QRgziT->.
- [42] E. Ryu, J. Liu, S. Wang, X. Chen, Z. Wang, W. Yin, Plug-and-play methods provably converge with properly trained denoisers, in: *International Conference on Machine Learning*, PMLR, 2019, pp. 5546–5557.
- [43] C. Anil, J. Lucas, R. Grosse, Sorting out Lipschitz function approximation, in: *International Conference on Machine Learning*, PMLR, 2019, pp. 291–301.
- [44] D. Zou, R. Balan, M. Singh, On Lipschitz bounds of general convolutional neural networks, *IEEE Trans. Inf. Theory* 66 (3) (2020) 1738–1759, <https://doi.org/10.1109/TIT.2019.2961812>.
- [45] P. Bohra, D. Perdios, A. Goujon, S. Emery, M. Unser, Learning Lipschitz-controlled activation functions in neural networks for plug-and-play image reconstruction methods, in: *NeurIPS 2021 Workshop on Deep Learning and Inverse Problems*, 2021, <https://openreview.net/forum?id=efCsbTzQTbH>.
- [46] A. Virmaux, K. Scaman, Lipschitz regularity of deep neural networks: analysis and efficient estimation, in: S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, R. Garnett (Eds.), *Advances in Neural Information Processing Systems*, vol. 31, Curran Associates, Inc., 2018, <https://proceedings.neurips.cc/paper/2018/file/d54e99a6c03704e95e6965532dec148b-Paper.pdf>.
- [47] S. Adeb, V.G. Troitsky, Locally piecewise affine functions and their order structure, *Positivity* 21 (1) (2017) 213–221.
- [48] J. De Loera, J. Rambau, F. Santos, *Triangulations: Structures for Algorithms and Applications*, vol. 25, Springer Science & Business Media, 2010.
- [49] H.W. Kuhn, Some combinatorial lemmas in topology, *IBM J. Res. Dev.* 4 (5) (1960) 518–524, <https://doi.org/10.1147/rd.45.0518>.
- [50] E. Allgower, K. Georg, *Triangulations by Reflections with Applications to Approximation*, Birkhäuser Basel, Basel, 1978, pp. 9–32.
- [51] D.F. Watson, Computing the n-dimensional Delaunay tessellation with application to Voronoi polytopes, *Comput. J.* 24 (2) (1981) 167–172, <https://doi.org/10.1093/comjnl/24.2.167>.
- [52] V.T. Rajan, Optimality of the Delaunay triangulation in \mathbb{R}^d , *Discrete Comput. Geom.* 12 (2) (1994) 189–202, <https://doi.org/10.1007/BF02574375>.
- [53] A.L. Edmonds, Simplicial decompositions of convex polytopes, *Pi Mu Epsilon J.* 5 (3) (1970) 124–128.
- [54] J. He, L. Li, J. Xu, C. Zheng, Relu deep neural networks and linear finite elements, *J. Comput. Math.* (2020), <https://doi.org/10.4208/JCM.1901-M2018-0160>.
- [55] D. Van De Ville, T. Blu, M. Unser, W. Philips, I. Lemahieu, R. Van de Walle, Hex-splines: a novel spline family for hexagonal lattices, *IEEE Trans. Image Process.* 13 (6) (2004) 758–772, <https://doi.org/10.1109/TIP.2004.827231>.
- [56] J.B. Lasserre, K.E. Avrachenkov, The multi-dimensional version of $\int_{-1}^1 b^a x^p dx$, *Am. Math. Mon.* 108 (2) (2001) 151–154, <https://doi.org/10.1080/00029890.2001.11919735>.
- [57] V. Baldoni, N. Berline, J.A. De Loera, M. Köppe, M. Vergne, How to integrate a polynomial over a simplex, *Math. Comput.* 80 (273) (2010) 297–325, <https://doi.org/10.1090/s0025-5718-2010-02378-6>.
- [58] C. Heumann, M. Schomaker, Shalabh, *Combinatorics*, Springer International Publishing, Cham, 2016, pp. 97–107.
- [59] I. Kra, S.R. Simanca, On circulant matrices, *Not. Am. Math. Soc.* 59 (03) (2012) 368, <https://doi.org/10.1090/noti804>.
- [60] L. Kamenski, W. Huang, H. Xu, Conditioning of finite element equations with arbitrary anisotropic meshes, *Math. Comput.* 83 (289) (2014) 2187–2211.
- [61] M. Kim, A. Entezari, J. Peters, Box spline reconstruction on the face-centered cubic lattice, *IEEE Trans. Vis. Comput. Graph.* 14 (6) (2008) 1523–1530, <https://doi.org/10.1109/TVCG.2008.115>.
- [62] C. de Boor, R.A. DeVore, A. Ron, Approximation from shift-invariant subspaces of $L^2(\mathbb{R}^d)$, *Trans. Am. Math. Soc.* 341 (2) (1994) 787–806, <https://doi.org/10.2307/2154583>.
- [63] L. Condat, D. Van De Ville, Three-directional box-splines: characterization and efficient evaluation, *IEEE Signal Process. Lett.* 13 (7) (2006) 417–420, <https://doi.org/10.1109/LSP.2006.871852>.
- [64] J. Horacek, U. Alim, *Evaluating Box Splines with Reduced Complexity*, 2018.
- [65] C. de Boor, K. Höllig, *B-Splines from Parallelepipeds*, *Tech. Rep.*, Wisconsin Univ-Madison Mathematics Research Center, 1982.
- [66] W.A. Dahmen, C.A. Micchelli, On the linear independence of multivariate B-splines, I. *Triangulations of simplices*, *Tech. Rep.* 5 (1982).
- [67] T. Zaslavsky, Facing up to Arrangements: Face-Count Formulas for Partitions of Space by Hyperplanes, vol. 154, *American Mathematical Society*, 1975.

- [68] M. Jordan, A.G. Dimakis, Exactly computing the local Lipschitz constant of ReLU networks, in: H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, H. Lin (Eds.), *Advances in Neural Information Processing Systems*, vol. 33, Curran Associates, Inc., 2020, pp. 7344–7353, <https://proceedings.neurips.cc/paper/2020/file/5227fa9a19dce7ba113f50a405dcdf09-Paper.pdf>.
- [69] H. Gouk, E. Frank, B. Pfahringer, M. Cree, Regularisation of neural networks by enforcing Lipschitz continuity, *Mach. Learn.* 110 (2021) 393–416.
- [70] H. Sedghi, V. Gupta, P.M. Long, The singular values of convolutional layers, in: *International Conference on Learning Representations*, 2019, pp. 1–12, <https://openreview.net/forum?id=rJevYoA9Fm>.
- [71] S. Neumayer, A. Goujon, P. Bohra, M. Unser, Approximation of Lipschitz functions using deep spline neural networks, *SIAM J. Math. Data Sci.* 5 (2) (2023) 306–322, <https://doi.org/10.1137/22M1504573>.