Stable Parametrization of Continuous and Piecewise-Linear Functions∗

Alexis Goujon †1, Joaquim Campos†, and Michael Unser†

1École polytechnique fédérale de Lausanne

March 11, 2022

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. 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) 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 ℓ2 → L2 condition number of this local representation. Finally, as a special case of our framework, we focus on a systematic parametrization of Rd 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 optimal. We also relate our local representation to a nonlocal one based on shifts of a causal ReLU-like function.

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 (vk, yk) ∈ Rd × R for k = 1, . . . , K, one wants to find f : Rd → R such that f(vk) ≈ yk for k = 1, . . . , K, without overfitting. As such, the problem is ill-posed. To make it numerically tractable, a reconstruction space H is chosen as the image of a finite-dimensional parameter space Θ under a given synthesis operator T : Θ → H. This operator maps a parameter θ ∈ Θ to its continuous representation T{θ} ∈ H. A celebrated way to choose the synthesis operator is to pick a feedforward neural network architecture. Given the multidimensional parameter θ = (θ1, . . . , θL+1) ∈ Θ, we then have that

\[ T\{θ\} = (f_{θ_{L+1}} \circ σ_L \circ f_{θ_L} \circ σ_{L-1} \circ \cdots \circ σ_2 \circ f_{θ_2} \circ σ_1 \circ f_{θ_1}), \]

where L is the number of hidden layers of the neural network, \( f_{θ_k} : \mathbb{R}^{d_k} \to \mathbb{R}^{d_{k+1}} \) is an affine function parametrized by θk, and σk is an activation function that is chosen a priori. For the model to be expressive, the activation functions need to be nonaffine—otherwise the generated function would remain trivially affine. Interestingly, the pointwise rectified-linear-unit (ReLU) function x → max(x, 0), one of the simplest nonlinear functions, provides state-of-the-art performance [2, 3]. In this case, \( T\{θ\} \) is the composition of continuous piecewise-linear (CPWL) functions, which turns out to be a CPWL function as well [4]. Remarkably, the reverse also holds true: any CPWL function \( \mathbb{R}^d \to \mathbb{R} \) can be parametrized 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 a serious 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 that worsen each time a new layer is added [9, 10]. The depth is also detrimental to the interpretability of the parametrization: in deep networks, the effect of a parameter on the constructed mapping is poorly understood.

∗This work was supported by the Swiss National Science Foundation, Grant 200020_184646 / 1.
†alexis.goujon@epfl.ch
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 (Figure 1).

1.2.1 Local Representation
In dimension \( d = 1 \), any CPWL function \( f \) with knots \( v_k \) can be represented by the linear expansion \( f = \sum_k f(v_k)\beta_k \), where the \( \beta_k \) are the underlying triangular B-spline basis functions [11]. In higher dimensions, the knots are replaced by what is called a simplicial triangulation of the set \( \{v_k\} \) of vertices, which partitions the input domain into simplices. With the appropriate triangulation, a CPWL function can be represented by the explicit linear simplicial spline expansion \( f = \sum_k f(v_k)\beta_k \), where the \( \beta_k \) now denote the nodal basis functions or hat functions [12] (see Section 2.1 and Figure 1). These functions satisfy \( \beta_k(v_q) = \delta_{kq} \) while being affine on the simplices and compactly 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 single linear box splines [13, 14].

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 \( f(x) = a_0 + a_1(x - v_1) + \sum_{k=2}^{K-1} a_k \text{ReLU}(x - v_k) \) [15]. Note that this representation has many equivalent variations such as \( f(x) = c_0 + c_1(x - v_1) + \sum_{k=2}^{K-1} c_k |x - v_k| \). The generalization to any dimension is due to Wang and Sun [16]. Their generalized hinging-hyperplanes (GHH) model can represent any CPWL function as \( f(x) = \sum_k \epsilon_k \max(f_k^1(x),\ldots,f_k^{m_k}(x)) \), where \( f_k^1,\ldots,f_k^{m_k} \) are affine functions, \( \epsilon_k = \pm 1 \), and \( m_k \leq d + 1 \). This expansion has also different variations and can be recast as \( f(x) = \sum_k \epsilon_k \max(g_k^1(x),\ldots,g_k^{m_k-1}(x)) \), where \( g_k^1,\ldots,g_k^{m_k-1} \) 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, 17, 18, 19, 20].

1.3 On the Stability of Parametrizations
While powerful in learning applications, nonlocal basis functions raise concerns about the stability of the model. Since the nonlocal atoms do not belong to any of the \( \mathcal{L}_p(\mathbb{R}) \) Lebesgue spaces, the synthesis operator \( T : \Theta \rightarrow \mathcal{H} \) is ill-conditioned. A small change in a parameter can lead to tremendous changes in the generated function. Nonlocal representations also lack stability when a CPWL function is used to interpolate data (Figure 2).
show in Appendix A that, in the one-dimensional case, the corresponding condition number for $K$ points is at least $O(K^{3/2})$. Conversely, the local representation $f = \sum_k f(v_k)\beta_k$ offers an explicit solution with, as result, a condition number equal to 1. To further examine the stability of the local parametrization, we consider the parameter space $\ell_2(\mathbb{Z})$ of finite-energy sequences and equip the function space with the $L_2$ norm. Ideally, the collection $(\beta_k)$ of functions would form an orthonormal basis. However, this does not hold for hat functions. The next best thing one can hope for is that the generating functions form a Riesz basis, which means that $(\beta_k)$ 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 $\Theta$ to $H$. This strong property has emerged as a standard requirement in many signal-processing theories and finite-elements methods [21, 22, 23, 24, 25] within a broad spectrum of applications. It is known that uniform B-splines of any degree and their high-dimensional box-spline extensions (with suitable directions) generate Riesz bases [26, 27]. Yet, to the best of our knowledge, the exact Riesz bounds of linear box splines are not known in high dimensions and the case of irregular triangulations has not been addressed in full generality so far.

In this paper, we propose to investigate precisely and in any dimension the stability of the local parametrization of CPWL functions with the hope to bring detailed results against which other parametrizations could be compared.

The paper is organized as follows: In Section 2, we present the relevant mathematical concepts. We then restrict our investigation to affine functions on simplices in Section 3 and extend it to CPWL functions on any triangulation in Section 4. Finally, we discuss the uniform-grid setting in Section 5.

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 $f_1, f_2, \ldots, f_p$ and subsets $R_1, R_2, \ldots, R_p$ of $\mathbb{R}^d$ such that

(i) each $R_k$ is closed with nonempty interior;

(ii) for $k \neq q$, $R_k$ and $R_q$ have disjoint interiors;

(iii) the space is partitioned as $\bigcup_{k=1}^p R_k = \mathbb{R}^d$;

(iv) the function $f$ agrees with $f_k$ on $R_k$.

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 [28]).

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 $v_k \in \mathcal{V}$ we associate a target value $y_k \in \mathbb{R}$. Under Definition 1 it is not obvious to find a CPWL function $f$ satisfies $f(v_k) = y_k$. The local representation, on the contrary, offers a more systematic way to address this problem. This requires first to form a triangulation of the set $\mathcal{V}$.

\(^1\)Note that the meaning of the term “locally finite” depends on the mathematical field.
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 dimension \(d = 0, \ldots, 3\) 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

\[
s = \text{conv}(v_1, \ldots, v_{d+1}) = \left\{ \sum_{k=1}^{d+1} \lambda_k v_k : \lambda_k \geq 0, \sum_{k=1}^{d+1} \lambda_k = 1 \right\}.
\]  

(2)

A \(k\)-face of a simplex is the convex hull of \((k + 1)\) of its vertices, which is a \(k\)-simplex embeded in \(\mathbb{R}^d\). The volume of a simplex admits the explicit form \(\text{Vol}(s) = \frac{1}{d!} |\det(v_2 - v_1, \ldots, v_{d+1} - v_1)|\).

**Definition 3** (adapted from [29]). 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);

(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 true 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 [30, 31]).

- **Irregular (Random) Sampling Locations**: A Delaunay triangulation always exists in any number of dimensions for a finite set \(\mathcal{V}\), and there are efficient algorithms to compute it [32, 33].

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

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

![Figure 3](image_url)

Figure 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).

2.1.2 Hat Basis Functions

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

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

(3)

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 [34]. 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 CPWL$(S)$ is uniquely determined by the values it assumes at the vertices $V$ of $S$. This leads to the local linear expansion

$$\forall f \in \text{CPWL}(S): f = \sum_{v \in V} f(v) \beta^S_v,$$

where the hat functions $\beta^S_v \in \mathbb{R}^{\text{conv}(V)}$ (see Figure 3) are defined on every simplex $s \in S$ by

$$\beta^S_v|_s = \begin{cases} \lambda^*_v, & s \in \text{St}(v) \\ 0, & \text{otherwise,} \end{cases}$$

where $\lambda^*_v$ is the unique affine function that is vanishes at all vertices of $s$ but takes value 1 at vertex $v$. In other words, $\lambda^*_v$ outputs the barycentric coordinate of simplex $s$ attached to vertex $v$ for a given $x \in s$. 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 $f \in \mathbb{R}^{\text{conv}(V)}$ we use the notation $\|f\|_{L_p} = \left( \int_{x \in \text{conv}(V)} |f(x)|^p dx \right)^{1/p}$. The hat basis functions have many desirable properties, such as,

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

$$\forall (a, b) \in \mathbb{R}^d \times \mathbb{R}, \forall x \in \text{conv}(V): a^T x + b = \sum_{v \in V} (a^T v + b) \beta^S_v(x),$$

which includes the partition-of-unity condition $\sum_{v \in V} \beta^S_v = 1$.

When $\text{St}(v)$ is convex, the hat function simply reads [35]

$$\beta^S_v = \left( \min_{s \in \text{St}(v)} \lambda^*_v \right)_+.$$  

2.2 Riesz Bases

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) $\text{Span}(\{\varphi_k\}_{k \in I}) = \mathcal{H}$ (completeness),

(ii) there exist $0 < A \leq B < +\infty$ such that, for any $c \in \ell_2(I)$,

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

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

The tightest constants $A$ and $B$ that satisfy (8) are called the Riesz bounds. We call the ratio $B/A$ the Riesz condition number. It is indeed the $\ell_2 \rightarrow L_2$ condition number of the synthesis operator $T: c \mapsto \sum_{k \in I} c_k \varphi_k$. The Riesz-basis property guarantees that $T$ is a bounded linear bijection, which means that there is a unique and stable
link between the parameters and the functions being generated. Note that the condition number 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 multi-index shifts of a single generating function \(\{\varphi_k = \varphi(\cdot - k) : k \in \mathbb{Z}^d\}\), the Riesz-sequence property is well characterized via the discrete-time Fourier transform \(\hat{g}\) of the sampled autocorrelation of \(\varphi\), as given by

\[
\hat{g}: \omega \mapsto \sum_{k \in \mathbb{Z}^d} \langle \varphi, \varphi(\cdot - k) \rangle e^{-i k^T \omega}. \tag{9}
\]

In this uniform scenario, the Fourier equivalent of the Riesz sequence condition is \([22, 36]\)

\[
0 < A^2 = \operatorname{ess inf}_{\omega \in [0,2\pi]^d} \hat{g}(\omega) \leq B^2 = \operatorname{ess sup}_{\omega \in [0,2\pi]^d} \hat{g}(\omega) < +\infty. \tag{10}
\]

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 parametrization on simplicial partitions (Sections 4 and 5).

Proposition 1. Let \(f : \mathbb{R}^d \to \mathbb{C}\) be an affine function and \(s = \text{conv}(v_1, \ldots, v_{d+1})\) a \(d\)-simplex. Then

\[
\frac{\operatorname{Vol}(s)}{(d+2)(d+1)} \sum_{k=1}^{d+1} |f(v_k)|^2 \leq \int_{x \in s} |f(x)|^2 \, dx \leq \frac{\operatorname{Vol}(s)}{(d+1)} \sum_{k=1}^{d+1} |f(v_k)|^2, \tag{11}
\]

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 \(f : \mathbb{R}^d \to \mathbb{R}\), an integer \(p \in \mathbb{N}\), and a \(d\)-simplex \(s = \text{conv}(v_1, \ldots, v_{d+1})\), it is known that \([37, 38]\)

\[
\int_s f(x)^p \, dx = \operatorname{Vol}(s) d \sum_{k \in \mathbb{N}^{d+1}, |k| = p} f(v_1)^{k_1} \cdots f(v_{d+1})^{k_{d+1}}, \tag{12}
\]

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

Lemma 1. Let \(f : \mathbb{R}^d \to \mathbb{R}\) be an affine function and \(s = \text{conv}(v_1, \ldots, v_{d+1})\) a \(d\)-simplex. For any \(p \in \mathbb{N}\), we have that

\[
\int_s f(x)^p \, dx = \operatorname{Vol}(s) d \sum_{k \in \mathbb{N}^{d+1}, |k| = p} f(v_1)^{k_1} \cdots f(v_{d+1})^{k_{d+1}}. \tag{13}
\]

Proof. If the affine function \(f\) is not constant, then it can be written as \(f(x) = a^T (x - x_0)\). Equation (12) can be applied after a change of variable, as in

\[
\int_s f(x)^p \, dx = \int_s (a^T (x - x_0))^p \, dx = \int_{s - x_0} (a^T y)^p \, dy \tag{14}
\]

\[
= \operatorname{Vol}(s - x_0) \sum_{k \in \mathbb{N}^{d+1}, |k| = p} (a^T (v_1 - x_0))^{k_1} \cdots (a^T (v_{d+1} - x_0))^{k_{d+1}} \tag{15}
\]

\[
= \operatorname{Vol}(s) \sum_{k \in \mathbb{N}^{d+1}, |k| = p} f(v_1)^{k_1} \cdots f(v_{d+1})^{k_{d+1}} \tag{16}
\]

where \(s - x_0 := \{x - x_0 : x \in s\}\). If now the affine function \(f\) is constant with \(f(x) = b\), then \(\int_s f(x)^p \, dx = \operatorname{Vol}(s) b^p\).

We also have that

\[
\operatorname{Vol}(s) \sum_{k \in \mathbb{N}^{d+1}, |k| = p} f(v_1)^{k_1} \cdots f(v_{d+1})^{k_{d+1}} = \operatorname{Vol}(s) b^p, \tag{17}
\]

where we have used that \(\sum_{k \in \mathbb{N}^{d+1}, |k| = p} 1 = \binom{p+d}{d}\). This number is known in combinatorics as the combinations with replacement [39].
We can now deduce an important property of the hat functions.

**Proposition 2.** The $L_p$ norm of the hat function $\beta^S_v$ only depends on the dimension $d$ and the volume of its support.

It reads

$$\|\beta^S_v\|_{L_p} = \left(\frac{(p + d)^{-1}}{d} \text{Vol}(v)\right)^{1/p}. \tag{18}$$

**Proof.** We split the integral over the simplices of the support of $\beta^S_v$ and apply Lemma 1, which leads

$$\|\beta^S_v\|_{L_p}^p = \int_{\text{conv}(v)} |\beta^S_v(x)|^p \, dx = \int_{\text{conv}(v)} \beta_v(x)^p \, dx$$

$$= \sum_{s \in S} \int_{s} \beta^S_v(x)^p \, dx = \sum_{s \in \text{St}(v)} \int_s \beta^S_v(x)^p \, dx$$

$$= \sum_{s \in \text{St}(v)} \left(\frac{p + d}{d}\right)^{-1} \text{Vol}(s) = \left(\frac{p + d}{d}\right)^{-1} \text{Vol}(\text{St}(v)). \tag{19}$$

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

- the $L_2$ norm

$$\|\beta^S_v\|_{L_2}^2 = \frac{2\text{Vol}(\text{St}(v))}{(d + 1)(d + 2)}; \tag{20}$$

- the inner-product relation

$$\langle \beta^S_v, \sum_{u \in V} \beta^S_u \rangle_{\text{conv}(v)} = \langle \beta^S_v, 1 \rangle_{\text{conv}(v)} = \|\beta^S_v\|_{L_1} = \frac{\text{Vol}(\text{St}(v))}{(d + 1)}; \tag{21}$$

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

Interestingly, 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 $f : \mathbb{R}^d \to \mathbb{R}$ be an affine function, $s = \text{conv}(v_1, \ldots, v_{d+1})$ a $d$-simplex, and $f_s = (f(v_1), \ldots, f(v_{d+1})) \in \mathbb{R}^{d+1}$. Then,

$$\int_s f(x)^2 \, dx = \frac{\text{Vol}(s)}{(d + 1)(d + 2)} f_s^T P_{d+1} f_s, \tag{22}$$

where

$$P_{d+1} = I_{d+1} + I_{d+1} = \begin{bmatrix} 2 & 1 & \cdots & 1 \\ 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 1 & \cdots & 1 & 2 \end{bmatrix} \in \mathbb{R}^{(d+1) \times (d+1)}. \tag{23}$$

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

$$\int_s f(x)^2 \, dx = \text{Vol}(s) \left(\frac{2 + d}{d}\right)^{-1} \sum_{k \in \mathbb{N}^{d+1}, |k| = 2} f(v_1)^{k_1} \cdots f(v_{d+1})^{k_{d+1}}$$

$$= \frac{2\text{Vol}(s)}{(d + 1)(d + 2)} 1/2 \left(\sum_{p,q=1}^{d+1} f(v_p)f(v_q) + \sum_{p=1}^{d+1} f(v_p)^2\right)$$

$$= \frac{\text{Vol}(s)}{(d + 1)(d + 2)} \left(\sum_{p=1}^{d+1} f(v_q)^2 \right)^2 + \sum_{p=1}^{d+1} f(v_p)^2. \tag{24}$$
On the other hand, we have that

\[
f_s^T P_{d+1} f_s = \sum_{p=1}^{d+1} f(v_p) \left( \sum_{q=1}^{d+1} f(v_q) + f(v_p) \right) = \left( \sum_{p=1}^{d+1} f(v_p) \right)^2 + \sum_{p=1}^{d+1} f(v_p)^2. \tag{26}
\]

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

**Lemma 3.** Let \( f, g : \mathbb{R}^d \to \mathbb{C} \) be affine functions, \( s = \text{conv}(v_1, \ldots, v_{d+1}) \) a \( d \)-simplex, \( f_s = (f(v_1), \ldots, f(v_{d+1})) \in \mathbb{R}^{d+1} \), and \( g_s = (g(v_1), \ldots, g(v_{d+1})) \in \mathbb{R}^{d+1} \). It holds that

\[
\int_s f(x)g(x)dx = \frac{\text{Vol}(s)}{(d+1)(d+2)} f_s^H P_{d+1} g_s,
\]

where

\[
P_{d+1} = I_{d+1} + I_{d+1} \in \mathbb{R}^{(d+1) \times (d+1)}.
\]

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 \( P_{d+1} \) 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 \( P_{d+1} \) defined in Lemma 2 is a symmetric circulant matrix generated by the vector \((2, 1, \ldots, 1)\). Its eigenvalues are known to be

\[
\lambda_m = 2 + \sum_{n=1}^{d} \zeta_{d+1}^n \quad \text{with} \quad m = 1, \ldots, d + 1 \quad \text{and where} \quad \zeta_{d+1} = e^{\frac{2\pi i}{d+1}}.
\]

These expressions are further simplified to

\[
\lambda_m = 1 + \sum_{n=1}^{d} \zeta_{d+1}^n = \begin{cases} d + 2, & m = d + 1 \\ 1, & \text{otherwise} \end{cases}
\]

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

\[
\|c\|^2 \leq c^H P_{d+1} c \leq (d+2)\|c\|^2.
\]

We now conclude by applying Lemma 3.

Unfortunately, the condition number \( \sqrt{d+2} \) given by the inequalities in Proposition 1 depends on the dimension \( d \). However, the eigenvalues of \( P_{d+1} \) are all \( 1 \) except for one that is \( (d+2) \). This means that, in general, \( \frac{\sqrt{C^H P_{d+1} C}}{\|C\|_F} \) is very heavily distributed toward 1 and only rarely approaches the upper bound \( \sqrt{d+2} \), especially in high dimensions, as quantified in Lemma 3. One should recall that, even though the dimension worsens the condition number, the effect is stochastically negligible.

**Proposition 3.** Let \( P_{d+1} = I_{d+1} + I_{d+1} \in \mathbb{R}^{(d+1) \times (d+1)} \), \( C \) a random unit vector of \( \mathbb{C}^{d+1} \) whose distribution is uniform on the unit sphere. Then, the first two moments of the random variable \( \frac{\sqrt{C^H P_{d+1} C}}{\|C\|_2} \) can be bounded by quantities that do not depend on the dimension \( d \), as expressed by

\[
\mathbb{E} \left( \frac{\sqrt{C^H P_{d+1} C}}{\|C\|_2} \right) \leq \sqrt{2} \quad \text{and} \quad \mathbb{E} \left( \frac{C^H P_{d+1} C}{\|C\|_2^2} \right) = 2.
\]

\[
E \left( \frac{\sqrt{C^H P_{d+1} C}}{\|C\|_2} \right) \leq \sqrt{2} \quad \text{and} \quad E \left( \frac{C^H P_{d+1} C}{\|C\|_2^2} \right) = 2.
\]
Proof. Knowing that \( P_{d+1} \) is a real and symmetric matrix, there exists an orthonormal basis \((u_k)_{k=1}^{d+1}\) made of its eigenvectors. They can be chosen so that the associated eigenvalues are \(1, \ldots, (d+2)\) (see proof of Proposition \ref{pro:sec2}). The random vector \( C \) can be decomposed as \( C = \sum_{k=1}^{d+1} A_k u_k \), where the coordinates \( A_k \) are random scalar variables that follow the same distribution and satisfy \( \sum_{k=1}^{d+1} |A_k|^2 = 1 \). We directly deduce that \( E(|A_k|^2) = \frac{1}{d+1} \). In addition, \( C^H P_{d+1} C = \sum_{k=1}^{d} |A_k|^2 + (d+2)|A_{d+1}|^2 = 1 + (d+1)|A_{d+1}|^2 \), which yields that \( E \left( \frac{C^H P_{d+1} C}{\|C\|_2^2} \right) = 2 \) since \( C^H C = 1 \). The conclusion follows from the well known inequality \( E(X)^2 \leq E(X^2) \) true for any random variable \( X \). \qed

4 Stability of the Local Parametrization on Irregular 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 \(2, \ldots, d\) vertices. Fortunately, the necessary and sufficient condition that a triangulation must satisfy for the local hat functions to form a Riesz basis only relies on a simple quantity: the volume of the star of the vertices.

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

(i) the collection of functions \((\beta^S_v)_{v \in V}\) forms a Riesz basis of the space \(\text{CPWL}(S) \cap L_2(\mathbb{R}^d)\);

(ii) \[
\begin{align*}
\inf_{v \in V} V_{\text{inf}}^S &= \inf_{v \in V} \text{Vol(St(v))} > 0 \\
\sup_{v \in V} V_{\text{sup}}^S &= \sup_{v \in V} \text{Vol(St(v))} < +\infty;
\end{align*}
\]

(iii) \[
\begin{align*}
\inf_{v \in V} \|\beta^S_v\|_{L_2} > 0 \\
\sup_{v \in V} \|\beta^S_v\|_{L_2} < +\infty.
\end{align*}
\]

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

\[
\sqrt{\frac{\sum_{v \in V} c_v \beta^S_v}{(d+1)(d+2)}} \|c\|_{\ell_2} \leq \| c \|_{\ell_2} \leq \sqrt{\frac{\sum_{v \in V} c_v \beta^S_v}{(d+1)(d+2)}} \|c\|_{\ell_2}
\]

or, equivalently, that

\[
\frac{1}{\sqrt{2}} \inf_{v \in V} \|\beta^S_v\|_{L_2} \|c\|_{\ell_2} \leq \| c \|_{\ell_2} \leq \sqrt{\frac{d+2}{2}} \sup_{v \in V} \|\beta^S_v\|_{L_2} \|c\|_{\ell_2}.
\]

The Riesz condition number \( r \) satisfies

\[
r \leq \sqrt{d+2} \sqrt{\frac{\inf_{v \in V} V_{\text{inf}}^S}{\sup_{v \in V} V_{\text{sup}}^S}} = \sqrt{d+2} \times \frac{\sup_{v \in V} \|\beta^S_v\|_{L_2}}{\inf_{v \in V} \|\beta^S_v\|_{L_2}}.
\]

**Proof.** The equivalence \((ii) \Leftrightarrow (iii)\) is a direct consequence of \([20]\).

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 \( v \in V \),

\[
A \leq \|\beta^S_v\|_{L_2} \leq B,
\]

where \(0 \leq A \leq B \leq +\infty\) are the Riesz bounds.
We now prove \((ii) \Rightarrow (i)\). Let \(c \in \ell_2(I)\), \(f = \sum_{v \in V} c_v \beta^S_v\). We have that

\[
\|f\|_{L_2}^2 = \int_{\text{conv}(V)} |f(x)|^2 \, dx = \sum_{S} \int_{s} |f(x)|^2 \, dx \\
\leq \frac{1}{(d+1)} \sum_{s \in S} \sum_{v \in s \cap V} \text{Vol}(s) |f(v)|^2 \\
= \frac{1}{(d+1)} \sum_{v \in V} \sum_{s \in \text{St}(v)} \text{Vol}(s) |f(v)|^2 \\
= \frac{1}{(d+1)} \sum_{v \in V} \text{Vol}(\text{St}(v)) |f(v)|^2 \\
\leq \frac{\sup_{v \in V} (\text{Vol}(\text{St}(v)))}{(d+1)} \|c\|_{\ell_2}^2,
\]

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

\[
\|f\|_{L_2}^2 = \int_{\text{conv}(V)} |f(x)|^2 \, dx = \sum_{s \in S} \int_{s} |f(x)|^2 \, dx \\
\geq \frac{1}{(d+1)(d+2)} \sum_{s \in S} \sum_{v \in s \cap V} \text{Vol}(s) |f(v)|^2 \\
= \frac{1}{(d+1)(d+2)} \sum_{v \in V} \sum_{s \in \text{St}(v)} \text{Vol}(s) |f(v)|^2 \\
= \frac{1}{(d+1)(d+2)} \sum_{v \in V} \text{Vol}(\text{St}(v)) |f(v)|^2 \\
\geq \frac{\inf_{v \in V} (\text{Vol}(\text{St}(v)))}{(d+1)(d+2)} \|c\|_{\ell_2}^2.
\]

(37)

(38)

Theorem 1 provides a quantitative way to compare the stability of triangulations. In particular, a triangulation is good when the ratio \(\frac{\text{Vol}(\text{St}(v))}{\text{Vol}(\text{St}(v))}\) 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 \(\text{Vol}(\text{St}(v)) = \sum_{s \in \text{St}(v)} \text{Vol}(s)\), we have that

\[
\inf_{s \in S} \text{Vol}(s) \times \inf_{v \in V} |\text{St}(v)| \leq V^\text{St}_{\inf} \leq V^\text{St}_{\sup} \leq \sup_{s \in S} \text{Vol}(s) \times \sup_{v \in V} |\text{St}(v)|.
\]

(39)

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 quality 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.

### 4.2 Triangulations with Finitely Many Vertices

For a triangulation \(S\) of a finite set \(V\) of vertices in \(d\) dimensions, the hat functions always form a Riesz basis. Indeed, since we assumed that \(V\) cannot be contained in any \((d-1)\)-dimensional affine subspace of \(\mathbb{R}^d\), all simplices of \(S\) are not degenerated and the conditions of Theorem 1 are fulfilled. The exact bounds can be computed in accordance with Theorem 2.
Lemma 2. We obtain a matrix $L$ if one chooses shifts of a single linear box spline to improve the computational performance. In this setting, the parametrization of CPWL functions is naturally suitable for some applications, such as image processing [41], but can also be a sensible choice in low dimensional learning problems [42]. While a uniform grid constrains the model and thereby reduces its expressivity, it significantly improves the computational performance. Throughout this section, we assume that the vertices coincide with the sites of a lattice $\Lambda$. This is natural in some applications, such as image processing [41], but can also be a sensible choice in low dimensional learning problems [42]. While a uniform grid constrains the model and thereby reduces its expressivity, it significantly improves the computational performance. In this setting, the parametrization of CPWL functions is naturally handled if one chooses shifts of a single linear box spline $B$ for the hat basis functions. The generated space $\text{CPWL}(\mathcal{S}) = \{ \sum_{k \in \Lambda} c_k B(x - k) : c_k \in \mathbb{C} \}$ is now shift-invariant and lends itself to the powerful tools of Fourier analysis [43]. Note that, for any $x \in \mathbb{R}^d$, the sum $\sum_{k \in \Lambda} c_k B(x - k)$ has at most $(d + 1)$ nonzero arguments. This follows from the short support of linear box splines (we provide more details on the properties of linear box splines in Section 5.1).

5 Uniform Setting

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 $\ell_2 \rightarrow L_2$ condition number). 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 condition number. When the Delaunay triangulation is not unique, the choice can be guided by the related condition number, as detailed in Figure 4.
The matrix \( \Xi \) splines, which will in return allow us to derive precise specific results.

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 precise specific results.

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

\[
\hat{B}_{\Xi_{d+1}}(\omega) = |\det \Xi| \prod_{k=1}^{d+1} \frac{1 - e^{-i \xi_k^T \omega}}{i \xi_k^T \omega}.
\]

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 4.

Proposition 4.

\[
B_{\Xi_{d+1}}(x) = \sum_{\epsilon \in \{0, 1\}^{d+1}} (-1)^{|\epsilon|} \min (\Xi^{-1}(x - \Xi_{d+1} \epsilon))_+.
\]

Proof. The product in (43) can be expanded as

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

By invoking Lemma 5 (Appendix B) and making use of the general Fourier stretch theorem, we get that

\[
\min(\Xi^{-1}x)_+ \overset{\mathcal{F}}{\rightarrow} |\det \Xi| \prod_{k=1}^{d+1} \left( \frac{1}{i \xi_k^T \omega} + \pi \delta(\xi_k^T \omega) \right),
\]

where \( x \) is the space variable, \( \omega \) is the pulsation variable and \( \delta \) is the Dirac distribution. Knowing that \( (1 - e^{-i \xi_k^T \omega}) \delta(\xi_k^T \omega) = 0 \), we observe that (43) has the equivalent form

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

We conclude by taking the inverse Fourier transform on both sides of (47). \( \square \)

Proposition 4 is illustrated in dimension \( d = 2 \) in Figure 5. This expansion gives a way to prove that the GHH model [16] can represent any linear box spline. More interestingly, it provides a precise relation between the local and nonlocal representations, with explicit generalized hinging hyperplanes (namely, the shifts of a single generalized hinging hyperplane). Our result is close to [14], where a similar formula is proven for three directional...
Figure 5: Linear decomposition of the 2D linear box spline in translates of the causal hinge functions \((x, y) \mapsto \min(x, y)_+\).

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 [13]. Nonetheless, when applied to linear box splines, the expansion in [13] is made of discontinuous Green functions and is less compact.

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

- continuity and piecewise linearity (obvious from Proposition 4 since \(x \mapsto \min(x)_+\) is CPWL);
- compact support since the support is the Minkowski sum of \((\xi_1, \ldots, \xi_{d+1})\), which is a zonotope that is symmetric with respect to its center \(\xi_{d+1} = \sum_{k=1}^{d} \xi_k\);
- \(B_{\Xi_{d+1}}(\xi_{d+1}) = 1\) and \(B_{\Xi_{d+1}}(\Xi_k) = 0\) for \(k \in \mathbb{Z}^d \setminus \{1\}\);
- 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 present a simple but powerful result.

Proposition 5. Let \((\xi_1, \ldots, \xi_d)\) 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 \(U \in \mathbb{R}^{d \times d}\) an invertible matrix. Then, the Riesz bounds of \(B_{U\Xi_{d+1}}\) are the ones of \(B_{\Xi_{d+1}}\) scaled by \(\sqrt{|\det(U)|}\).

Proof. From Proposition 4 we infer that \(B_{U\Xi_{d+1}} = B_{\Xi_{d+1}} \circ U^{-1}\). (The generalized version of this relation \(M_{U\Xi} = |\det U^{-1}| M_{\Xi} \circ U^{-1}\) 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 \(\|f \circ U^{-1}\|_{L^2}^2 = |\det(U)| \|f\|_{L^2}^2\), which means that, for \(c \in \ell_2(\mathbb{Z})\), it holds that

\[
\| \sum_{k \in \mathbb{Z}^d} c_k B_{U\Xi_{d+1}}(\cdots - k) \|_{L^2}^2 = \| (\sum_{k \in \mathbb{Z}^d} c_k B_{\Xi_{d+1}}) \circ U^{-1}(\cdots - k) \|_{L^2}^2 = |\det(U)| \| \sum_{k \in \mathbb{Z}^d} c_k B_{\Xi_{d+1}}(\cdots - k) \|_{L^2}^2,
\]

which allows us to conclude.

Theorem 3. Let \((\xi_1, \ldots, \xi_d)\) be a family of \(d\) free vectors 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 \(\{B_{\Xi_{d+1}}(\cdots - \Xi k) : k \in \mathbb{Z}^d\}\) of linear box splines forms a Riesz basis with bounds

\[
A = \sqrt{\frac{|\det \Xi|}{(d+2)}} \quad \text{and} \quad B = \sqrt{|\det \Xi|}.
\]
The associated Riesz condition number is \( \sqrt{d+2} \).

**Proof.** Following Proposition 5, we focus solely on the cartesian lattice and denote the Cartesian linear box spline by \( D \). In this case, it is known that the simplices form a Kuhn/Freudenthal triangulation \( [14] \). The vertices of any simplex \( s = \text{conv}(v_0, \ldots, v_d) \) of the triangulation take the form

\[
\mathbf{v}_k = \mathbf{v}_0 + \sum_{p=1}^k \mathbf{e}_{\sigma(p)}, \tag{50}
\]

where \( \sigma \) is a permutation of the set \( \{0, \ldots, d\} \) and \( k = 1, \ldots, d \). Correspondingly, for \( p = 0, \ldots, d \), there is a unique vertex \( \mathbf{v}_{k_p} \) in each simplex such that \( |v_{k_p}| \equiv p \mod (d+1) \). Indeed, \( (50) \) yields that \( |\mathbf{v}_k| = |\mathbf{v}_0| + k \). Then, we use the Fourier characterization \( [10] \) to find the Riesz bounds. Specifically, we have to find the essential extrema of

\[
\hat{g}: \omega \mapsto \sum_{\mathbf{k} \in \mathbb{Z}^d} \langle D, D(\cdot - \mathbf{k}) \rangle e^{-i\omega^T \mathbf{k}}. \tag{51}
\]

Since the basis function \( D \) is compactly supported, the sum is finite and the function \( \hat{g}: \mathbb{R}^d \to \mathbb{C} \) is continuous and 2\( \pi \)-periodic with respect to each coordinate. We can therefore simply look for the maximum and the minimum of \( \hat{g} \) in the hypercube \([0, 2\pi]^d\). We apply the triangular inequality, use the non negativity of \( D \) and invoke the partition of unity of the linear box spline to conclude that the maximum of \( \hat{g} \) is attained for \( \omega = 0 \). With \( [21] \) 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 \( 1 \).

Now, for the minimum, we evaluate \( \hat{g} \) at \( \omega_0 = (-\frac{2\pi}{d+1}, \ldots, -\frac{2\pi}{d+1}) \) and get

\[
\sum_{\mathbf{k} \in \mathbb{Z}^d} \langle D, D(\cdot - \mathbf{k}) \rangle e^{-i\omega_0^T \mathbf{k}} = \langle D, \sum_{\mathbf{k} \in \mathbb{Z}^d} \zeta_{d+1}^{[\mathbf{k}]_D} D(\cdot - \mathbf{k}) \rangle = \sum_{s \in \text{St}(D)} \langle D, \sum_{\mathbf{k} \in \mathbb{Z}^d} \zeta_{d+1}^{[\mathbf{k}]_D} D(\cdot - \mathbf{k}) \rangle_s = \sum_{s \in \text{St}(D)} \langle D, \sum_{(k-1) \in \mathbb{N}^d} \zeta_{d+1}^{[\mathbf{k}]_D} D(\cdot - \mathbf{k}) \rangle_s, \tag{52}
\]

where \( \langle f, h \rangle_s = \int_{x \in s} \overline{f(x)} h(x) dx \). On one hand, we apply Lemma 3 and use the inherent structure of the Kuhn triangulation displayed in \( (50) \) to deduce that

\[
\min_{\omega \in [0, 2\pi]^d} \hat{g}(\omega) \leq \hat{g}(\omega^*) = \sum_{s \in \text{St}_D} \frac{1}{(d+1)(d+2)} \text{Vol}(s) \begin{bmatrix} 2 & 1 & \ldots & 1 \\ 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 1 & \ldots & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ \zeta_{d+1} \\ \vdots \\ \zeta_{d+1}^d \end{bmatrix} = \frac{\text{Vol}(\text{St}(D))}{(d+1)(d+2)}. \tag{53}
\]

On the other hand, Theorem 1 implies that \( \min_{\omega \in [0, 2\pi]^d} \hat{g}(\omega) \geq \frac{\text{Vol}(\text{St}(D))}{(d+1)(d+2)} \), from which we infer that \( A^2 = \frac{\text{Vol}(\text{St}(D))}{(d+1)(d+2)} \) and, since the box spline \( D \) is nonnegative, \( \int_{\mathbb{R}^d} D(x) dx = \|D\|_{L_1} = \frac{\text{Vol}(\text{St}(D))}{(d+1)} \) (from \( [21] \)). 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 5.

Theorem 1 and 3 yield the same bounds for linear box splines, which confirms the good quality of the bounds provided for irregular 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 \( \frac{1}{d+1} \). We deduce that the linear box spline is made of \( (d+1)! \) nonzero affine pieces.

From Theorem 3 the Riesz condition number of the linear box-spline parametrization 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 parametrization is given by the mean of the function $\hat{g}$ (defined in the proof of Theorem 3) over $[0, 2\pi]^d$, compute as
\[
\frac{1}{(2\pi)^d} \int_{\omega \in [0, 2\pi]^d} \sum_{k \in \mathbb{Z}^d} \langle D, D(-k) \rangle e^{-i\omega^{\top}k} \, d\omega = \langle D, D \rangle = \|D\|_{L_2}^2 = 2 \frac{\text{Vol}(\text{St}(D))}{(d + 2)(d + 1)} = 2 \min_{\omega \in [0, 2\pi]^d} \hat{g}(\omega). \tag{54}
\]

We infer that $\hat{g}$ rarely takes values close to its upper bound, especially in high dimensions, which was observed for affine functions on a simplex in Lemma 3. In short, although the condition number scales badly with the dimension, most linear combinations of box splines will behave closely to the lower bound and the dimension should not be significantly detrimental to the stability of the parametrization.

\section{Conclusion}

We have provided a precise measure of the stability of the parametrization of CPWL functions with hat basis functions on any triangulation. First, we have estimated the $\ell_2 \to L_2$ condition number of the parametrization 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. Then, we have proposed a method to compute the condition number for finite triangulations. It boils down to the computation of the condition number of a given matrix. When the vertices lie at the sites of a lattice, we parametrize 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 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, dimension will only rarely affect the stability of the local representation.
A Condition Number of the Nonlocal Parametrization

Consider the nonlocal parametrization of one-dimensional CPWL functions with control points \(v_1 < \ldots < v_K \in \mathbb{R}\)

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

(55)

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

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

(56)

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

\[
\begin{bmatrix}
1/h & 0 & \cdots & 0 \\
1/h & 1 & 0 & \cdots \\
\vdots & 2 & \ddots & \ddots \\
\vdots & \vdots & \ddots & 0 \\
1/h & (K-1) & \cdots & 2 & 1
\end{bmatrix}
= \begin{bmatrix}
\theta_1 \\
\theta_2 \\
\vdots \\
\cdots \\
\theta_K
\end{bmatrix},
\]

\[
y = M^{-1}y.
\]

(57)

To give a lower bound to the condition number of the problem, we remark that \(Me_2 = h \sum_{k=1}^{K-1} ke_{k+1}\) and \(Me_K = he_K\), where \(e_k\) are the canonical vectors of \(\mathbb{R}^d\). We infer that \(\|M^{-1}(\sum_{k=1}^{K-1} ke_k)\|_2/\|\sum_{k=1}^{K-1} ke_k\|_2 = (K(K-1)(2K-1)/6)^{-1/2}/h\) and that \(\|M^{-1}e_K\|_2 = 1/h\). It implies that the \(\ell_2\) condition number \(r\) of the problem satisfies

\[
r = \max_{a,b \in \mathbb{R}^d \setminus \{0\}} \left\{ \frac{\|M^{-1}a\|_2 \|b\|_2}{\|a\| \|M^{-1}b\|_2} \right\} \geq \sqrt{\frac{K(K-1)(2K-1)}{6}}.
\]

(58)

B The Generalized Hinging Hyperplane Generating Function of Linear Box Splines

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

\[
\mathcal{H}: \omega \mapsto \frac{1}{i\omega} + \pi\delta(\omega).
\]

(59)

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

\[
\hat{H}: \omega \mapsto \prod_{k=1}^d \left( \frac{1}{i\omega_k} + \pi\delta(\omega_k) \right).
\]

(60)

We define a directional Heaviside wall function as \(W: x \mapsto h(x_d)\prod_{k=1}^{d-1} \delta(x_k)\) and the matrix \(A_d = \begin{bmatrix} 1 & \cdots & 1 \\ \cdots & \ddots & \cdots \\ 1 & \cdots & 1 \end{bmatrix}\).

Lemma 4.

\[
\forall x \in \mathbb{R}^d, (H \ast (W \circ A_d))(x) = \min(x)_+.
\]

(61)
Proof. We have that
\[
(H \ast (W \circ \mathbf{A}_d))(x) = \int_{\mathbb{R}^d} H(x - t)W(\mathbf{A}_d t) dt.
\]
(62)
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 \( y = \mathbf{A}_d t \). It yields that
\[
(H \ast (W \circ \mathbf{A}_d))(x) = (1/d) \int_{\mathbb{R}^d} H(x - \mathbf{A}_d^{-1} y) h(y_d) \prod_{k=1}^{d-1} \delta(y_k) dy
\]
\[
= (1/d) \int_{\mathbb{R}^d} H(x - y_d \mathbf{A}_d^{-1} \mathbf{e}_d) h(y_d) \prod_{k=1}^{d-1} \delta(y_k) dy
\]
\[
= (1/d) \int_{\mathbb{R}^d} H(x - \frac{y_d}{d} \mathbf{1}) h(y_d) \prod_{k=1}^{d-1} \delta(y_k) dy
\]
\[
= (1/d) \int_{\mathbb{R}} h(y_d) \left( \prod_{k=1}^{d} h(x_k - y_d/d) \right) \left( \prod_{k=1}^{d-1} \delta(y_k) \right) dy
\]
\[
= (1/d) \int_{\mathbb{R}} h(y_d) \left( \prod_{k=1}^{d} h(x_k - y_d/d) \right) dy_d.
\]
(63)
The quantity \( \prod_{k=1}^{d} h(x_k - y_d/d) h(y_d) \) is nonzero when \( y_d > 0 \) and \( y_d < d x_k \) for \( k = 1, \ldots, d \), which is equivalent to \( y_d > 0 \) and \( y_d < d \min(x_k) \). We can now conclude that \( (H \ast (L \circ \mathbf{A}_d))(x) = \min(x)_+ \).

\[ \square \]

Lemma 5.
\[
\min(x)_+ \overset{x}{\mapsto} \prod_{k=1}^{d+1} \left( \frac{1}{x \omega_k} + \pi \delta(\omega_k) \right),
\]
(64)
where \( \omega_{d+1} = \sum_{k=1}^{d} \omega_k \).

Proof. From Lemma 4 we have that \( (H \ast (W \circ \mathbf{A}_d))(x) = \min(x)_+ \). The function \( W \) is separable and its Fourier transform reads \( \hat{W}(\omega) = \frac{1}{i \omega d} + \pi \delta(\omega_d) \). In addition, the general stretch theorem implies that
\[
(W \circ \mathbf{A}_d) \overset{x}{\mapsto} \frac{1}{d} \hat{W}(\mathbf{A}_d^{-T} \omega) = \frac{1}{d} \left( \frac{1}{i e_d^T \mathbf{A}_d^{-T} \omega} + \pi \delta(e_d^T \mathbf{A}_d^{-T} \omega) \right).
\]
(65)
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
\[
(W \circ \mathbf{A}_d) \overset{x}{\mapsto} \frac{1}{d} \hat{W}(\mathbf{A}_d^{-T} \omega) = \left( \frac{1}{i \omega_{d+1}} + \pi \delta(\omega_{d+1}) \right).
\]
(66)
We reach the conclusion by using the Fourier transform of \( H \) and by transforming the convolution into a product in the Fourier domain.

[4]

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. doi:10.1038/nature14539 URL http://colah.github.io/

[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. Montufar, 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: V. Feldman, A. Rakhlin, O. Shamir (Eds.), 29th Annual Conference on Learning Theory, Vol. 49 of Proceedings of Machine Learning Research, PMLR, Columbia University, New York, New York, USA, 2016, pp. 907–940. URL https://proceedings.mlr.press/v49/eldan16.html

[7] H. Mhaskar, T. Poggio, Deep vs. Shallow Networks : An Approximation Theory Perspective, Analysis and Applications 14 (06) (2016) 829—848. arXiv:1608.03257

[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, International Journal of Automation and Computing 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 Machine Learning 110 (2) (2021) 393–416. arXiv:1804.04368 doi:10.1007/s10994-020-05929-w URL https://doi.org/10.1007/s10994-020-05929-w

[10] K. Scaman, A. Virmaux, Lipschitz Regularity of Deep Neural Networks: Analysis and Efficient Estimation, Advances in Neural Information Processing Systems (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, Journal of Computational Mathematics 38 (3) (2020) 502–527. doi: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, Journal of Approximation Theory 162 (9) (2010) 1607–1630. doi:10.1016/j.jat.2010.04.007

[15] C. D. Aliprantis, D. Harris, R. Tourky, Continuous Piecewise Linear Functions, Macroeconomic Dynamics 10 (1) (2006) 77.

[16] S. Wang, X. Sun, Generalization of Hinging Hyperplanes, IEEE Transactions on Information Theory 51 (12) (2005) 4425—4431. doi:10.1109/TIT.2005.859246

[17] A. L. Maas, A. Y. Hannun, A. Y. Ng, Rectifier Nonlinearities Improve Neural Network Acoustic Models, in: in ICML Workshop on Deep Learning for Audio, Speech and Language Processing, 2013.

[18] K. He, X. Zhang, S. Ren, J. Sun, Delving Deep into rRctifiers: Surpassing Human-Level Performance on ImageNet Classification, in: Proceedings of the IEEE international conference on computer vision, 2015, pp. 1026–1034.

[19] 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
[20] I. J. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, Y. Bengio, Maxout Networks, in: 30th International Conference on Machine Learning, no. 3, International Machine Learning Society (IMLS), 2013, pp. 2356–2364. arXiv:1302.4389

[21] A. Aldroubi, M. Unser, Sampling Procedures in Function Spaces and Asymptotic Equivalence with Shannon’s Sampling Theory, Numerical Functional Analysis and Optimization 15 (1-2) (1994) 1–21. doi:10.1080/01630569408816545

[22] A. Aldroubi, Oblique Projections in Atomic Spaces, Proceedings of the American Mathematical Society 124 (7) (1996) 2051–2060. doi:10.1090/S0002-9939-96-03255-8

[23] M. Unser, I. Daubechies, On the Approximation Power of Convolution-Based Least Squares versus Interpolation, IEEE Transactions on Signal Processing 45 (7) (1997) 1697–1711. doi:10.1109/78.599940

[24] R. Jia, W. Zhao, Riesz Bases of Wavelets and Applications to Numerical Solutions of Elliptic Equations, Mathematics of Computation 80 (275) (2011) 1525–1556.

[25] 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. doi:10.1109/ITNG.2013.121

[26] W. Dahmen, C. A. Micchelli, Translates of Multivariate Splines, Linear Algebra and its Applications 52 (1983) 217–234.

[27] Y. Guan, S. Lu, Y. Tang, Characterization of Compactly Supported Refinable Splines whose Shifts Form a Riesz Basis, Journal of Approximation Theory 133 (2) (2005) 245–250.

[28] S. Adeeb, V. G. Troitsky, Locally Piecewise Affine Functions and their Order Structure, Positivity 21 (1) (2017) 213–221.

[29] J. De Loera, J. Rambau, F. Santos, Triangulations: Structures for Algorithms and Applications, Vol. 25, Springer Science & Business Media, 2010.

[30] H. W. Kuhn, Some Combinatorial Lemmas in Topology, IBM Journal of Research and Development 4 (5) (1960) 518–524. doi:10.1147/rd.45.0518

[31] E. Allgower, K. Georg, Triangulations by Reflections with Applications to Approximation, Birkhäuser Basel, Basel, 1978, pp. 9–92. doi:10.1007/978-3-0348-6460-2_1 URL https://doi.org/10.1007/978-3-0348-6460-2_1

[32] D. F. Watson, Computing the n-Dimensional Delaunay Tessellation with Application to Voronoi Polytopes, The Computer Journal 24 (2) (1981) 167–172. doi:10.1093/comjnl/24.2.167

[33] V. T. Rajan, Optimality of the Delaunay triangulation in Rd, Discrete & Computational Geometry 12 (2) (1994) 189–202. doi:10.1007/BF02574375

[34] A. L. Edmonds, Simplicial decompositions of convex polytopes, Pi Mu Epsilon Journal 5 (3) (1970) 124–128.

[35] J. He, L. Li, J. Xu, C. Zheng, Relu deep neural networks and linear finite elements, Journal of Computational Mathematics (2020). doi:10.4208/JCM.1901-M2018-0160

[36] 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 Transactions on Image Processing 13 (6) (2004) 758–772. doi:10.1109/TIP.2004.827231

[37] J. B. Lasserre, K. E. Avrachenkov, The Multi-Dimensional Version of $\int_a^b x^p dx$, The American Mathematical Monthly 108 (2) (2001) 151–154. doi:10.1080/00029890.2001.11919735 URL https://doi.org/10.1080/00029890.2001.11919735

[38] V. Baldoni, N. Berline, J. A. De Loera, M. Köppe, M. Vergne, How to Integrate a Polynomial over a Simplex, Mathematics of Computation 80 (273) (2010) 297–325. doi:10.1090/s0025-5718-2010-02378-6
[39] C. Heumann, M. Schomaker, Shalabh, *Combinatorics*, Springer International Publishing, Cham, 2016, pp. 97–107. doi:10.1007/978-3-319-46162-5_5
URL https://doi.org/10.1007/978-3-319-46162-5_5

[40] I. Kra, S. R. Simanca, *On Circulant Matrices*, Notices of the American Mathematical Society 59 (03) (2012) 368. doi:10.1090/noti804
URL http://dx.doi.org/10.1090/noti804

[41] M. Kim, A. Entezari, J. Peters, Box Spline Reconstruction on the Face-Centered Cubic Lattice, IEEE Transactions on Visualization and Computer Graphics 14 (6) (2008) 1523–1530. doi:10.1109/TVCG.2008.115

[42] J. Campos, S. Aziznejad, M. Unser, Learning of Continuous and Piecewise-Linear Functions with Hessian Total-Variation Regularization (under review).

[43] C. de Boor, R. A. DeVore, A. Ron, Approximation from Shift-Invariant Subspaces of L2 (Rd), Transactions of the American Mathematical Society 341 (2) (1994) 787–806. doi:10.2307/2154583

[44] L. Condat, D. Van De Ville, Three-Directional Box-Splines: Characterization and Efficient Evaluation, IEEE Signal Processing Letters 13 (7) (2006) 417–420. doi:10.1109/LSP.2006.871852

[45] J. Horacek, U. Alim, Evaluating Box Splines with Reduced Complexity, 2018.