Injectivity of ReLU layers: Tools from Frame Theory

Daniel Haider*, Martin Ehler†, and Peter Balazs‡

November 2024

Abstract

Injectivity is the defining property of a mapping that ensures no information is lost and any input can be perfectly reconstructed from its output. By performing hard thresholding, the ReLU function naturally interferes with this property, making the injectivity analysis of ReLU layers in neural networks a challenging yet intriguing task that has not yet been fully solved. This article establishes a frame theoretic perspective to approach this problem. The main objective is to develop a comprehensive characterization of the injectivity behavior of ReLU layers in terms of all three involved ingredients: (*i*) the weights, (*ii*) the bias, and (*iii*) the domain where the data is drawn from. Maintaining a focus on practical applications, we limit our attention to bounded domains and present two methods for numerically approximating a maximal bias for given weights and data domains. These methods provide sufficient conditions for the injectivity of a ReLU layer on those domains and yield a novel practical methodology for studying the information loss in ReLU layers. Finally, we derive explicit reconstruction formulas based on the duality concept from frame theory.

1 Introduction

The **Re**ctified **L**inear **U**nit defined as $\mathrm{ReLU}(s) = \max(0, t)$ for $t \in \mathbb{R}$ has become indispensable as a non-linear activation function in artificial neural networks [14, 21, 15, 25]. Since originally introduced as a way to regularize the gradients in deep network architectures, there have been hardly any networks that do not use ReLU activation or some derivation of it [12, 18, 23].

A *ReLU layer* $C_\alpha(x) = \mathrm{ReLU}(Cx - \alpha)$ is the composition of an affine linear map comprising the multiplication by a weight matrix $C \in \mathbb{R}^{m\times n}$ and the shift by a bias vector $\alpha \in \mathbb{R}^m$, with an entry-wise application of ReLU on its output. The injectivity of a ReLU layer, and with that, the possibility of inverting it and inferring x from $C_\alpha(x)$, is a desired property in various applications. There has been interest in building injective models and inverting

*Acoustics Research Institute, Vienna, Austria (daniel.haider@oeaw.ac.at).
†University of Vienna, Faculty of Mathematics, Vienna, Austria (martin.ehler@univie.ac.at).
‡Acoustics Research Institute, Vienna, Austria (peter.balazs@oeaw.ac.at).

them on their range to regularize ill-posed inverse problems or designing injective generative models, such as normalizing flows, for manifold learning or compressed sensing [20, 27, 6]. Generally, knowing if a layer involves a loss of information increases the interpretability of the network immensely. As such, one can use an inverse mapping to trace back each layer output to its source input, which can help to decipher the decision-making process, diagnose model behavior, identify biases, and study accountability. Although the injectivity of ReLU layers has received increasing attention in recent years, it is still not fully understood, especially when it comes to the numerical verification in practice.

The goal of this paper is to demystify the injectivity of a single ReLU layer as a deterministic non-linear map on a comprehensive level. This involves a thorough analysis of fundamental properties of ReLU layers that are relevant for applications, and different characterizations of injectivity with respect to all properties involved, namely weights, bias, and input domain. By translating selected theoretical results into algorithmic solutions we present novel ways of verifying injectivity on bounded input domains in practice. Explicit reconstruction formulas and their implementation, together with a brief local stability analysis complete the claim of the paper.

The methodology to achieve these goals is based on *frame theory*, a mathematical paradigm that deals with stable, potentially redundant, and invertible representations of functions by means of inner products [11]. To make use of this machinery and all tools that come with it, we shall consider a weight matrix $C \in \mathbb{R}^{m\times n}$ in terms of its row vectors

$$C = \begin{pmatrix} -\phi_1- \\ \vdots \\ -\phi_m- \end{pmatrix}. \tag{1}$$

If C has more rows than columns ($m \geq n$) and full rank, the collection of row vectors $(\phi_i)_{i=1}^m$ is a spanning set for the domain space $\mathbb{R}^n$. In other words, the associated linear transform $C : \mathbb{R}^n \to \mathbb{R}^m$ is injective. In the context of frame theory, we say that $(\phi_i)_{i=1}^m$ is a *frame* for $\mathbb{R}^n$ [11] and C is the associated *analysis operator*. The application of C to x is interpreted as measuring the correlation of x to all frame vectors ϕ_i via $x \mapsto (\langle x, \phi_i\rangle)_{i=1}^m$. The resulting so-called *frame coefficients* $(\langle x, \phi_i\rangle)_{i=1}^m$ give a (potentially redundant) representation of x, from which we can always infer x explicitly. Roughly speaking, frame theory is the study of "quantifying" injectivity of a redundant representation in the sense of its numerical stability and constructing recovery maps with desired properties via the concept of dual frames. In this sense, it provides exactly the right tools for the goal of the paper.

Although frame theory deals with linear representations, it has been shown to be suitable for non-linear problems as well. One example is phase-retrieval [3] where, in the real setting, one asks for the injectivity and stability of the map

$$\begin{aligned} C_{|\,\cdot\,|} : \mathbb{R}^n/\{\pm 1\} &\to \mathbb{R}^m \\ x &\mapsto \big(\big|\langle x, \phi_i\rangle\big|\big)_{i=1}^m. \end{aligned} \tag{2}$$

It is well known that $C_{|\,\cdot\,|}$ is injective if and only if for any partition of the collection $(\langle x, \phi_i\rangle)_{i=1}^m$ into two sub-collections, at least one of them is a frame [3]. Inspired by this

approach, we may write a ReLU layer analogously as the map

$$\begin{aligned} C_\alpha : \mathbb{R}^n &\to \mathbb{R}^m \\ x &\mapsto (\mathrm{ReLU}(\langle x, \phi_i \rangle - \alpha_i))_{i=1}^m . \end{aligned} \tag{3}$$

In [26] (using another terminology) it has been shown that C_α is injective if and only if for any $x \in \mathbb{R}^n$ the frame vectors that are not affected by ReLU are a frame. We shall call a frame with this property *α-rectifying* on $\mathbb{R}^n$. While this characterization forms the basis of our work, we will focus on the practical assumption that, in applications, it may not always be most informative to consider the entire $\mathbb{R}^n$ as the domain where a ReLU layer should be injective. In fact, when considering standard normalization schemes of data sets for training and testing neural networks, it seems more reasonable to study injectivity only on bounded subsets $K \subseteq \mathbb{R}^n$ where the data is assumed or processed to live in. Indeed, a ReLU layer might be injective on K but not on $\mathbb{R}^n$. In this paper, we show that the choice of the data domain can have a profound impact on the injectivity behavior of C_α, and discuss how to leverage this fact in practice. Prototypical examples of such domains include the closed ball in $\mathbb{R}^n$ of radius $r > 0$, given by $\mathbb{B}_r = \{x \in \mathbb{R}^n : \|x\| \leq r\}$, the closed donut arising by excluding small data points $\mathbb{D}_{r,s} = \overline{\mathbb{B}_r \setminus \mathbb{B}_s}$ with $s < r$, and the sphere $\mathbb{S} = \{x \in \mathbb{R}^n : \|x\| = 1\}$ [22, 19]. Furthermore, when considering two consecutive ReLU layers, we can restrict the injectivity property of the second one to domains that lie in $\mathbb{R}^n_+$, such as the non-negative closed ball $\mathbb{B}_r^+ = \mathbb{B}_r \cap \mathbb{R}^n_+$. The restriction of the domain of C_α from $\mathbb{R}^n$ to a bounded $K \subseteq \mathbb{R}^n$ increases the feasibility and applicability of the problem in practice, while also making the mathematical setting more versatile. Furthermore, it establishes a natural connection to the bias vector and provides a framework where we can control the injectivity behavior through these two ingredients. This in turn allows us to approach the injectivity analysis also algorithmically.

The main theoretical component of this paper is a comprehensive characterization of the injectivity of a ReLU layer as a deterministic map, summarized in the following theorem.

Theorem. *Let $\Phi = (\phi_i)_{i=1}^m$ be a frame for $\mathbb{R}^n$, $\alpha \in \mathbb{R}^m$, and $\emptyset \neq K \subseteq \mathbb{R}^n$. Under the assumptions that Φ includes a unique most correlated basis everywhere (Def 3.7), K is open or strictly convex, and bias-exact for Φ (Def. 3.11), the following are equivalent.*

(i) The ReLU layer C_α, associated with Φ and α, is injective on K.

(ii) The frame Φ is α-rectifying on K (Def 2.2, Thm 2.4, 2.5).

(iii) The domain K lies in the maximal domain $\mathcal{K}_\alpha^*$ (Thm 3.4).

(iv) The values of the bias α do not exceed the values of the maximal bias $\alpha_K^\sharp$ (Thm 3.12).

For any bias α the maximal domain $\mathcal{K}_\alpha^*$ can be constructed explicitly as the union of intersections of closed affine half-spaces. For any domain K the maximal bias $\alpha_K^\sharp$ can be approximated numerically via sampling or via the inscribing polytope associated with Φ.

The main practical component of the paper comprises two algorithmic constructions of biases that approximate the maximal bias $\alpha_K^\sharp$ from the theorem above in different situations.

These can be used to study and effectively control the injectivity behavior of a ReLU layer in practice. Moreover, using the duality concept from frame theory we derive inversion formulas for injective ReLU layers that can be implemented easily as locally linear operators.

Related work

The approach in classical phase-retrieval in $\mathbb{R}^n$ by Balan et al. in [3] was decisive for the idea of characterizing the injectivity of a ReLU layer in terms of a property of the associated frame. The same approach is taken by Alharbi et al. to study the recovery of vectors from saturated inner measurements [2]. In a machine learning context, Puthawala et al. have introduced the notion of *directed spanning sets* in [26] as an equivalent concept to the *admissibility* condition by Bruna et al. in [8] to characterize a ReLU layer to be injective on $\mathbb{R}^n$. While the primary goal in [26] is the study of globally injective ReLU-networks on $\mathbb{R}^n$, and the one in [8] is a Lipschitz stability analysis of ReLU layers, our goal is to demystify the injectivity of a single ReLU layer in a more realistic setting, namely with any given weights and biases on bounded input *data* domains $K \subseteq \mathbb{R}^n$, and to provide methods of verifying injectivity in practice. This extends the ideas and methods by Haider et al. introduced in [17] significantly. Further related preprints are by Behrmann et al., who study the pre-images of ReLU layers from a geometric point of view [5] and by Maillard et al., which focus on injectivity of ReLU layers with random weights [24].

Outline

The paper is divided into four sections. In Section 2, we introduce *α-rectifying frames* as a characterizing family of frames that are associated with ReLU layers that are injective on a given input domain and discuss fundamental properties that are crucial for applications. In Section 3, we study the interplay of input domain and bias vector and derive a maximal domain and a maximal bias such that the associated ReLU layers become critically injective. This leads to two further characterizations of injectivity. Moreover, we present two methods to approximate the maximal bias and provide algorithmic solutions to apply them in practice. Section 4 is dedicated to explicit reconstruction formulas for injective ReLU layers and a brief local stability analysis of the recovery map.

2 Frames and the Injectivity of ReLU Layers

When applying a matrix $C \in \mathbb{R}^{m\times n}$ to a vector $x \in \mathbb{R}^n$ we can reconstruct x if and only if the collection of row vectors of C spans $\mathbb{R}^n$. Frame theory offers a definition of a spanning set that allows us to quantify "how good" it spans $\mathbb{R}^n$. So, throughout this paper, we denote a collection of m vectors by $\Phi = (\phi_i)_{i\in I} \subset \mathbb{R}^n$, using index sets I with $|I| = m \geq n$. Then, Φ is a *frame* for $\mathbb{R}^n$ if there are constants $0 < A \leq B$, called the *frame bounds* of Φ, such that

$$A \cdot \|x\|^2 \leq \sum_{i\in I} |\langle x, \phi_i\rangle|^2 \leq B \cdot \|x\|^2 \tag{4}$$

for all $x \in \mathbb{R}^n$ [10]. For $J \subseteq I$, we denote by $\Phi_J = (\phi_i)_{i\in J}$ the sub-collection of Φ with respect to the index set J. Any frame with $m = n$ is a basis and we will always assume

that Φ does not contain the zero vector. It is easy to see that (4) is equivalent to Φ being a spanning set and that the frame bounds A, B reflect the numerical stability properties of the representation of x under Φ. The *analysis operator* associated with Φ is given as

$$\begin{aligned} C : \mathbb{R}^n &\to \mathbb{R}^m \\ x &\mapsto (\langle x, \phi_i \rangle)_{i \in I} \end{aligned} \tag{5}$$

mapping a vector x to its *frame coefficients* as discussed in the introduction. So finally, we have that Φ is a frame, if and only if C is injective. Together with the assumption on the input domain mentioned in the introduction, this motivates to define a ReLU layer as follows.

Definition 2.1 (ReLU layer)**.** *A ReLU layer associated with a collection of weight vectors $\Phi = (\phi_i)_{i \in I} \subset \mathbb{R}^n$, a bias vector $\alpha = (\alpha_1, ..., \alpha_m)^\top \in \mathbb{R}^m$ and an input domain $K \subseteq \mathbb{R}^n$ is defined as the non-linear map given by*

$$\begin{aligned} C_\alpha : K &\to \mathbb{R}^m \\ x &\mapsto (\mathrm{ReLU}(\langle x, \phi_i \rangle - \alpha_i))_{i=1}^m . \end{aligned}$$

To encode the injectivity of C_α directly in terms of Φ we introduce a family of frames called *α-rectifying frames.*

2.1 Alpha-rectifying frames

For any given $x \in K$ the shift by the bias and the application of the ReLU function on the frame coefficients act as a thresholding mechanism that neglects all frame elements ϕ_i where $\langle x, \phi_i \rangle < \alpha_i$, rendering them *inactive*. According to this observation, for $x \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}^m$ we are interested in the index set associated with those frame elements that are *active* for x and α. We shall denote it by

$$I_x^\alpha = \{i \in I : \langle x, \phi_i \rangle \geq \alpha_i\}. \tag{6}$$

Dual to this notion, for $i \in I$ and $\alpha \in \mathbb{R}^m$ we denote the closed affine half-space of points where the frame element ϕ_i is active by

$$\Omega_i^\alpha = \{x \in \mathbb{R}^n : \langle x, \phi_i \rangle \geq \alpha_i\}. \tag{7}$$

The following definition gives the frame theoretic perspective to [26, Definition 1].

Definition 2.2 (α-rectifying frames)**.** *The collection $\Phi = (\phi_i)_{i \in I} \subset \mathbb{R}^n$ is called α-rectifying on $K \subseteq \mathbb{R}^n$ for $\alpha \in \mathbb{R}^m$ if for all $x \in K$ the sub-collection of active frame elements $\Phi_{I_x^\alpha} = (\phi_i)_{i \in I_x^\alpha}$ is a frame for $\mathbb{R}^n$.*

Figure 1 illustrates the two notions in (6) and (7) on the closed unit ball $\mathbb{B}$ in $\mathbb{R}^2$ (left, mid), and shows a simple example of an α-rectifying frame (right).

While many papers only consider ReLU layers without bias vectors and with input on whole $\mathbb{R}^n$ [26, 24], for our approach, bias vectors and the input domain play important roles, and the following basic properties of ReLU layers become crucial. Throughout the paper, for $\alpha, \alpha' \in \mathbb{R}^m$ we shall use the notation $\alpha \leq \alpha'$ for $\alpha_i \leq \alpha_i'$ for all $i \in I$, and $\alpha < \alpha'$ for $\alpha_i < \alpha_i'$ for all $i \in I$.

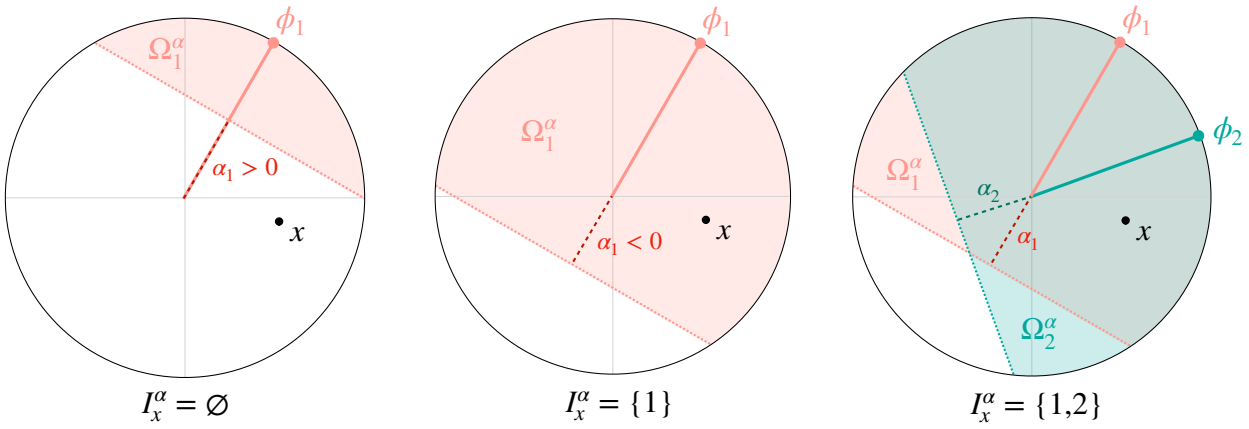

Figure 1: Illustrations of the notions I_x^α and Ω_i^α related to active frame elements on $\mathbb{B}$. The frame (ϕ_1, ϕ_2) in the most right example is α-rectifying on K if $K \subseteq (\Omega_1^\alpha \cap \Omega_2^\alpha)$.

Proposition 2.3. *Let Φ be α'-rectifying on K'. The following holds.*

(i) Φ is α'-rectifying on K for every $K \subseteq K'$.

(ii) Φ is α-rectifying on K' for every $\alpha \leq \alpha'$.

Hence, we are naturally interested in knowing the largest possible domains and biases that allow the α-rectifying property. In Section 3 we prove that indeed one obtains full characterizations of the injectivity of a ReLU layer via a *maximal domain* and a *maximal bias*.

We now present the fundamental connection between the α-rectifying property and the injectivity of C_α on K, forming the backbone of this paper. Theorems 2.4 and 2.5 address the two directions separately, each focusing on specific properties of K, thereby generalizing [26, Theorem 2].

Theorem 2.4 (Injectivity of ReLU layers I)**.** *Given $\Phi = (\phi_i)_{i\in I} \subset \mathbb{R}^n$, $\alpha \in \mathbb{R}^m$, and $\emptyset \neq K \subseteq \mathbb{R}^n$. If Φ is α-rectifying on K, then C_β is injective on K for all $\beta < \alpha$. Moreover, if K is open or convex, C_α is injective on K.*

Recall that a set U is called strictly convex if for all $x, y \in U$ and $\lambda \in (0,1)$, $x_\lambda := (1-\lambda)x - \lambda y \in \mathring{U}$, where $\mathring{U}$ is the interior of U. In particular, $\mathring{U} \neq \emptyset$.

Theorem 2.5 (Injectivity of ReLU layers II)**.** *Given $\Phi = (\phi_i)_{i\in I} \subset \mathbb{R}^n$, $\alpha \in \mathbb{R}^m$, and let $\emptyset \neq K \subseteq \mathbb{R}^n$ be open or strictly convex. If C_α is injective on K, then Φ is α-rectifying on K.*

Proof of Theorem 2.4. Let $x, y \in K$ and assume $C_\beta x = C_\beta y$. Clearly,

$$\langle x, \phi_i \rangle > \beta_i \text{ if and only if } \langle y, \phi_i \rangle > \beta_i, \tag{8}$$

from which we can deduce that $\langle x, \phi_i \rangle = \langle y, \phi_i \rangle$ for all $i \in I_x^\beta$. By assumption that Φ is α-rectifying, $\Phi_{I_x^\alpha}$ is a frame, and since $\beta < \alpha$ we have that $\Phi_{I_x^\beta}$ is one, too. It follows that $x = y$ and therefore that C_β is injective on K.

To show the moreover part, let $C_\alpha x = C_\alpha y$, for $x, y \in K$. Clearly,

$$\langle x, \phi_i \rangle = \langle y, \phi_i \rangle \tag{9}$$

for all $i \in I_x^\alpha \cap I_y^\alpha$. We will show that if K is open or convex, then $\Phi_{I_x^\alpha \cap I_y^\alpha}$ is a frame.
Let us first consider the case where K is open. We may choose $\varepsilon > 0$ such that the open ball around x, denoted by $B_\varepsilon^\circ(x)$, is contained in K. When assuming that $x \neq y$ then there is $\delta < 1$ with $0 < \delta < \varepsilon \cdot \|x - y\|^{-1}$ such that

$$x_\delta := (1 - \delta)x + \delta y \in B_\varepsilon^\circ(x). \tag{10}$$

Now let $i \in I_{x_\delta}^\alpha$. By the linearity of the inner product, we have the following.

$$\text{If } \langle x_\delta, \phi_i \rangle > \alpha_i \text{ then } \langle x, \phi_i \rangle > \alpha_i \text{ and } \langle y, \phi_i \rangle > \alpha_i. \tag{11}$$

$$\text{If } \langle x_\delta, \phi_i \rangle = \alpha_i \text{ then } \langle x, \phi_i \rangle = \alpha_i \text{ and } \langle y, \phi_i \rangle = \alpha_i. \tag{12}$$

Therefore, $I_{x_\delta} \subseteq I_x^\alpha \cap I_y^\alpha$. By (10), we have that $x_\delta \in K$ and since we assumed $\Phi_{I_{x_\delta}}$ to be a frame, so is $\Phi_{I_x^\alpha \cap I_y^\alpha}$.
Now let us assume K to be convex. For $\lambda \in (0, 1)$,

$$x_\lambda := (1 - \lambda)x + \lambda y \in K. \tag{13}$$

By the same arguments as above, (11) and (12) hold for $\langle x_\lambda, \phi_i \rangle$, hence $I_{x_\lambda} \subseteq I_x^\alpha \cap I_y^\alpha$. By assumption, $\Phi_{I_{x_\lambda}}$ is a frame and thereby, $\Phi_{I_x^\alpha \cap I_y^\alpha}$ is a frame. So for both cases, we can deduce that $x = y$, hence C_α is injective. □

Proof of Theorem 2.5. We prove the claim by counterposition. Assume that Φ is not α-rectifying. Then there is $x \in K$ such that $(\phi_i)_{i \in I_x^\alpha}$ is not a frame. Hence, there is

$$0 \neq r \in \operatorname{span}(\phi_i)_{i \in I_x^\alpha}^\perp. \tag{14}$$

If K is open, for all sufficiently small $\varepsilon > 0$ we have that

$$y_\pm := x \pm \varepsilon r \in K.$$

For $i \in I_x^\alpha$, (14) implies that $\langle y_+, \phi_i \rangle = \langle y_-, \phi_i \rangle$, leading to

$$\max(0, \langle y_+, \phi_i \rangle - \alpha_i) = \langle x, \phi_i \rangle - \alpha_i = \max(0, \langle y_-, \phi_i \rangle - \alpha_i). \tag{15}$$

If $I_x^\alpha = I$, then (15) already implies $C_\alpha y_+ = C_\alpha y_-$, so that C_α is not injective on K.
To address the case $I_x^\alpha \subset I$, we recall that $\langle x, \phi_i \rangle < \alpha_i$ holds for all $i \in I \setminus I_x^\alpha$. Therefore, we may choose ε sufficiently small, such that $y_\pm \in K$ and

$$\langle y_\pm, \phi_i \rangle < \alpha_i, \qquad i \in I \setminus I_x^\alpha. \tag{16}$$

Observation (16) leads to

$$0 = \max(0, \langle y_\pm, \phi_i \rangle - \alpha_i), \qquad i \in I \setminus I_x^\alpha. \tag{17}$$

According to (15) and (17), we derive $C_\alpha y_+ = C_\alpha y_-$. Thus, C_α is not injective on K.

Note that if K is strictly convex it contains more than one element and for every $z \in K$ where $z \neq x$ and $\lambda \in (0,1)$ with the same assumptions on x as above,

$$x_\lambda := (1-\lambda)x - \lambda z \in \mathring{K}.$$

Using the linearity of the inner product, there is λ sufficiently small such that

$$\langle x_\lambda, \phi_i \rangle < \alpha_i, \qquad i \in I \setminus I_x^\alpha. \tag{18}$$

Since $\mathring{K}$ is open and not empty, we can apply the same argument as above with x_λ to show that C_α is not injective on K. □

This shows that the injectivity of a ReLU layer is contingent upon topological properties of the domain from which the data is drawn. For the cases $\mathbb{R}^n$ and $\mathbb{B}_r$, we have established equivalence, where the former case corresponds to [26, Theorem 2]. Since the non-negative ball $\mathbb{B}_r^+ = \mathbb{B}_r \cap \mathbb{R}_+^n$ is convex but not open or strictly convex, only the direction in Theorem 2.4 holds. In the case of the donut $\mathbb{D}_{r,s} = \overline{\mathbb{B}_r \setminus \mathbb{B}_s}$, which is not open and not convex, only the first part of Theorem 2.4 holds, i.e., injectivity for strictly smaller biases. In Section 3.1, we give an example of a similar scenario where injectivity fails.

While the α-rectifying property makes the injectivity of a ReLU layer more accessible, it remains challenging to verify it in practice. In the following, we discuss how specific properties of the frame influence its α-rectifying property, and how to leverage them.

2.2 Normalized frames

It is a simple, yet, crucial observation that we may restrict the α-rectifying property to frames with unit norm vectors by scaling the bias with the norms of the frame elements. In the context of neural networks, normalizing the underlying frame (or the weight matrix in a row-wise manner) is a standard normalization technique [30]. We write $\Phi \subset \mathbb{S}$.

Lemma 2.6. *A frame $\Phi = (\phi_i)_{i\in I} \subset \mathbb{R}^n$ is α-rectifying on K if and only if the normalized frame $\Phi' = (\phi_i \cdot \|\phi_i\|^{-1})_{i\in I} \subset \mathbb{S}$ is α'-rectifying on K, where $\alpha_i' = \alpha_i \cdot \|\phi_i\|^{-1}$.*

The statement follows from the fact that $\langle x, \phi_i \rangle \geq \alpha_i$ is equivalent to $\langle x, \phi_i \cdot \|\phi_i\|^{-1} \rangle \geq \alpha_i \cdot \|\phi_i\|^{-1}$ for all $x \in K$. Therefore, we may always assume $\|\phi_i\| = 1$ for all $i \in I$, which simplifies the problem setting substantially. The norms can be reintroduced at any stage of processing or analysis.

2.3 Full-spark frames

The effect of ReLU can be interpreted as introducing input-dependent erasures in the underlying frame, i.e., losing certain coefficients [16]. In this context, the *spark* of a frame has been shown to be a useful concept. It is defined as the smallest number $s \geq n+1$ of linearly dependent frame elements that one can choose from Φ. In other words, any sub-collection with $s-1$ frame elements from Φ is a frame. Frames with $s = n+1$ are called *full-spark frames*.

This family of frames has shown to be maximally robust to erasures [1] which makes the full-spark property interesting for injective ReLU layers in particular. For phase retrieval, it is known that if a full-spark frame has $m \geq 2n-1$ elements then the phase-retrieval operator (2) is injective [3]. For the ReLU case, knowing the spark of Φ relaxes the condition for a frame to be α-rectifying to a counting argument.

Corollary 2.7. *Let Φ be a frame with spark s then Φ is α-rectifying on K if and only if $|I_x^\alpha| \geq s-1$ for all $x \in K$.*

Although it is an NP-hard problem to verify if a given frame is full-spark, in a numerical setting it is a mild condition that is almost surely satisfied in the presence of randomness. Indeed, if the entries of the frame elements are i.i.d. samples from an absolutely continuous probability distribution, then the associated random frame is full-spark with probability one [1]. Since most initialization methods in neural networks are based on i.i.d. sampling schemes, full-spark frames appear naturally in the context of deep learning.

2.4 Perturbed frames

For a bounded domain K the α-rectifying property is robust to perturbation. In particular, small perturbations of an α-rectifying frame result in an α'-rectifying frame where α' is close to α.

Lemma 2.8. *Let Φ be α-rectifying on a bounded domain K with $M = \sup_{x\in K} \|x\|$. For $\varepsilon > 0$, a perturbed frame $\Phi' = (\phi_i')_{i\in I}$ satisfying $\|\phi_i - \phi_i'\| < \varepsilon$ for all $i \in I$ is α'-rectifying on K with $\alpha_i' = \alpha_i - \varepsilon M$, $i \in I$.*

Proof. Let $x \in K$, then for any $i \in I_x^\alpha$ it holds that

$$\langle x, \phi_i' \rangle = \langle x, \phi_i \rangle - \langle x, \phi_i - \phi_i' \rangle > \alpha_i - \varepsilon \|x\| \geq \alpha_i - \varepsilon M.$$

□

It is known that for any frame Φ there is an arbitrarily small perturbation such that the resulting perturbed frame is full-spark [10]. Therefore, we may interpret Lemma 2.8 in the sense that for any α-rectifying frame, there is an arbitrarily close full-spark frame that is α'-rectifying with α' being arbitrarily close to α.

2.5 Redundancy

One of the central properties of a frame is its redundancy, i.e., the ratio $q = \frac{m}{n} \geq 1$. For random ReLU layers with i.i.d. Gaussian entries and no bias, it has been studied at which redundancy they become injective on $\mathbb{R}^n$ asymptotically [26, 24]. Let $p_{m,n}$ denote the probability that C_α with m frame elements in $\mathbb{R}^n$ and $\alpha = \mathbf{0}$ is injective then it has been proven that $q \leq 3.3$ implies that $\lim_{n\to\infty} p_{m,n} = 0$ and $q \geq 9.091$ implies that $\lim_{n\to\infty} p_{m,n} = 1$. Furthermore, the authors in [24] state the conjecture that there exists a redundancy $q \in (6.6979, 6.6981)$ where the transition from non-injectivity to injectivity happens. We will revisit this conjecture in a non-asymptotic setting in the experimental part later in the paper (Section 3.3).

In a non-random setting, a trivial leverage of redundancy is considering the collection $\Psi = (\Phi, -\Phi)$ for any given frame Φ. Doubling the redundancy in this symmetric way makes Ψ become $\mathbf{0}$-rectifying on $\mathbb{R}^n$ by construction (see Figure 2 left) [8, 26, 32]. In a general deterministic setting, however, it is difficult to establish sufficient conditions for the α-rectifying property only in terms of redundancy as it depends heavily on the geometric characteristics of the frame. As already mentioned in Section 2.3, this is in contrast to the phase-retrieval setting, where a redundancy of $q \geq \frac{2n-1}{n}$, together with a full spark assumption is already sufficient for injectivity. In the ReLU case, it is known that a redundancy of two is necessary when considering $\alpha = \mathbf{0}$ and $K = \mathbb{R}^n$ [8]. We extend this known result from $\alpha = \mathbf{0}$ to arbitrary α in the following proposition.

Proposition 2.9. *Any α-rectifying frame on $\mathbb{R}^n$ has at least redundancy two.*

Proof. By assuming that the zero vector is not a frame element, we can choose $x \in \mathbb{R}^n$ with $\langle x, \phi_i \rangle \neq 0$ for all $i \in I$. We denote

$$I_x^+ = \{i \in I : \langle x, \phi_i \rangle > 0\}, \tag{19}$$
$$I_x^- = \{i \in I : \langle x, \phi_i \rangle < 0\}. \tag{20}$$

By the choice of x, the sets I_x^+ and I_x^- form a disjoint partition of I. Let $\alpha \in \mathbb{R}^m$ and define

$$t^* := \max_{i \in I} \frac{\alpha_i}{\langle x, \phi_i \rangle} \tag{21}$$

then for all $t > t^* > 0$ we have

$$\langle tx, \phi_i \rangle > \alpha_i \text{ and } \langle -tx, \phi_i \rangle < \alpha_i \quad \text{for } i \in I_x^+ \tag{22}$$
$$\langle tx, \phi_i \rangle < \alpha_i \text{ and } \langle -tx, \phi_i \rangle > \alpha_i \quad \text{for } i \in I_x^-. \tag{23}$$

Hence, $I_{tx}^\alpha = I_x^+$ and $I_{-tx}^\alpha = I_x^-$. We found two elements $u = tx, v = -tx \in \mathbb{R}^n$ with $I_u^\alpha \cap I_v^\alpha = \emptyset$. Assuming Φ to be α-rectifying on $\mathbb{R}^n$ implies that $\Phi_{I_u^\alpha}$ and $\Phi_{I_v^\alpha}$ are frames, i.e., contain at least n elements in particular. Since $I_u^\alpha \cap I_v^\alpha = \emptyset$, it must hold that $m \geq 2n$. □

Assuming that the input for a ReLU layer is contained in $\mathbb{B}_r$, we find that the necessary redundancy-two condition from Proposition 2.5 breaks. We use boldface notation for bias vectors with constant entries, i.e., $\mathbf{r} \in \mathbb{R}^m$ denotes the vector with entries $r \in \mathbb{R}$.

Lemma 2.10. *Any normalized frame is $(-\mathbf{r})$-rectifying on $\mathbb{B}_r$. If it is additionally a basis, this is also necessary, i.e., $-r$ is the maximal value.*

Proof. Let $x \in \mathbb{B}_r$. Since $\langle x, \phi_i \rangle = \|x\| \langle \frac{x}{\|x\|}, \phi_i \rangle \geq -\|x\| \geq -r$ the first statement follows. For the second, let Φ be a basis, then Φ is α-rectifying on $\mathbb{B}_r$ if and only if for every $x \in \mathbb{B}_r$ it holds that $I = I_x^\alpha$. In particular, since $-r \cdot \Phi \subset \mathbb{B}_r$ and for every $i \in I$, the maximal choice of the α_i is determined by the fact that $\langle -r \cdot \phi_i, \phi_i \rangle = -r$. □

This emphasizes that the choice of the input domain can have a significant impact on the α-rectifying property. The following section will examine this interaction between domain and bias in greater detail and explain how it can be leveraged.

3 Interplay of Domain and Bias

In the context of applications, we may find ourselves in a situation where we have provided a trained ReLU layer and wish to ascertain whether it is injective for a specific data set. One way to address this is to verify that the data set in question is contained within a set where we have already established that the ReLU layer is injective. This leads to the following natural question.

Q1: *Given α, what is the largest domain K such that Φ is α-rectifying on K?*

An alternative approach, building upon the inclusiveness property of ReLU layers (Prop. 2.3), is to ascertain that the values of the given bias do not exceed the values of a bias for which we already know that the corresponding ReLU layer is injective. This leads to the dual question to the one above.

Q2: *Given K, what is the largest bias α such that Φ is α-rectifying on K?*

Answering these questions will provide us with two further characterizations of the α-rectifying property in terms of domain and bias, respectively. With this, we obtain alternative ways of verifying the injectivity of the associated ReLU layer. To better understand how bias and domain interact, we point out some basic scaling relations.

Lemma 3.1. *Let Φ be α-rectifying on K. The following holds.*

(i) Φ is $(r \cdot \alpha)$-rectifying on $r \cdot K$ for any $r > 0$.

(ii) If $\alpha \geq 0$, then Φ is α-rectifying on $r \cdot K$ with $r \geq 1$.

(iii) If $0 \in K$, then at least n bias values are non-positive.

Proof. All properties are easy to see.

(i) For $r > 0$, $\langle x, \phi_i \rangle \geq \alpha_i$ if and only if $\langle r \cdot x, \phi_i \rangle \geq r \cdot \alpha_i$.

(ii) For $r \geq 1$, $\langle x, \phi_i \rangle \geq \alpha_i$ implies $\langle r \cdot x, \phi_i \rangle \geq r \cdot \alpha_i \geq \alpha_i$.

(iii) For $x = 0$, we have that $\Phi_{I_x^\alpha}$ is a frame and $0 = \langle 0, \phi_i \rangle \geq \alpha_i$ holds for all $i \in I_x^\alpha$. Since $|I_x^\alpha| \geq n$, the claim follows.

□

Consequently, by either scaling the data or the bias, it may be possible to compensate for situations where a frame is not α-rectifying on K but is on $K' \subsetneq K$. The following examples show such compensation through restrictions other than scaling.

Example 3.2. *A basis can never be α-rectifying on $\mathbb{R}^n$ for any α. However, the standard basis for $\mathbb{R}^n$ is $\mathbf{0}$-rectifying on $\mathbb{R}^n_+$ (Figure 2 mid).*

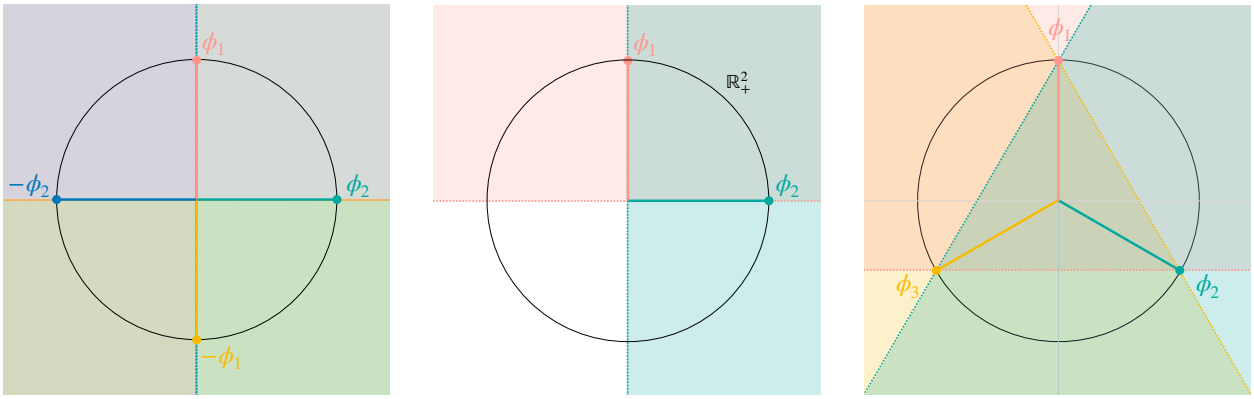

Figure 2: Left: The frame composed of the standard basis and its negative elements is $\mathbf{0}$-rectifying on $\mathbb{R}^2$. Mid: The standard basis is $\mathbf{0}$-rectifying on $\mathbb{R}^2_+$ and $(-\mathbf{1})$-rectifying on $\mathbb{B}$. Right: The triangle frame is $(-\frac{\mathbf{1}}{\mathbf{2}})$-rectifying on $\mathbb{B}$, but never on $\mathbb{R}^2$ since there will always be cones where only one element is active (lighter areas).

Example 3.3. *The frame*

$$\Phi_3 = \left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -\sqrt{3}/2 \\ -1/2 \end{pmatrix}, \begin{pmatrix} \sqrt{3}/2 \\ -1/2 \end{pmatrix} \right)$$

is not α-rectifying on $\mathbb{R}^n$ for any α since $m = 3 < 4 = 2n$. However, by a geometric argument (see Figure 2 right), it is easy to see that Φ is $(-\frac{\mathbf{1}}{\mathbf{2}})$-rectifying on $\mathbb{B}$. Note that $-\frac{1}{2}$ is the largest possible value here.

This makes clear that it is essential to select the domain carefully if we want to effectively study the injectivity behavior of the associated ReLU layer. This leads us to Question ***Q1***.

3.1 Maximal domain

We aim to identify the maximal domain K for a frame Φ and a bias α. This provides a characterization of the α-rectifying property from a geometric point of view. Recall that for $i \in I$ and $\alpha \in \mathbb{R}^m$ we denote the closed affine half-space where the frame element ϕ_i is active for α by

$$\Omega_i^\alpha = \{x \in \mathbb{R}^n : \langle x, \phi_i \rangle \geq \alpha_i\}. \tag{24}$$

Extending the intuition from the example in Figure 1 (right), we find that any frame is α-rectifying on the intersection of sufficiently many Ω_i^α's. Restricting to minimal frames, i.e., basis, reveals the characterization.

Theorem 3.4 (Maximal domain)**.** *Let $\Phi \subset \mathbb{R}^n$ be a frame and $\alpha \in \mathbb{R}^m$. The maximal domain where Φ is α-rectifying is given by*

$$\mathcal{K}_\alpha^* = \bigcup_{\substack{J \subseteq I \\ \Phi_J \text{ basis}}} \bigcap_{i \in J} \Omega_i^\alpha. \tag{25}$$

In other words, Φ is α-rectifying on K if and only if $K \subseteq \mathcal{K}_\alpha^*$.

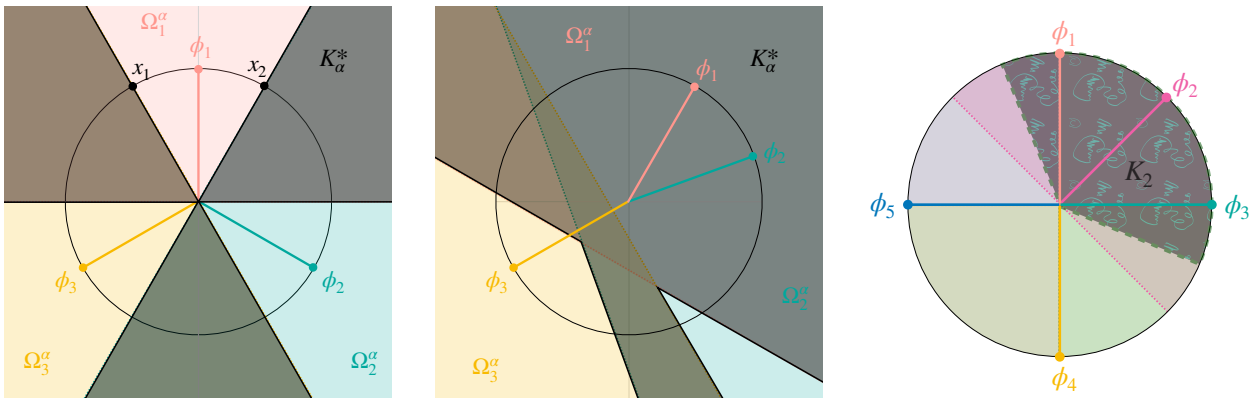

Figure 3: The dark areas in the left and mid picture indicate the maximal domains $\mathcal{K}_\alpha^*$ for the triangle frame with zero bias (left), and a normalized random frame with random bias (mid). The right illustration corresponds to Example 3.10. We point out how $K_2 = \{x \in K : 2 \in J^*(x)\}$ looks like, where $J^*(x)$ is the most correlated basis for x, see Definition 3.7.

Proof. Let $x \in K$. Assuming Φ to be α-rectifying on K, then $\Phi_{I_x^\alpha}$ is a frame. Since (in $\mathbb{R}^n$) every frame contains a basis, there is $L \subseteq I_x^\alpha$ such that the sub-collection Φ_L is a basis. Clearly, $\langle x, \phi_i \rangle \geq \alpha_i$ still holds for all $i \in L$, hence, $x \in \mathcal{K}_\alpha^*$.

For the converse direction, by definition of $\mathcal{K}_\alpha^*$ for all $x \in \mathcal{K}_\alpha^*$ there is $M \subseteq I$ with $\langle x, \phi_i \rangle \geq \alpha_i$ for all $i \in M$ such that Φ_M is a basis. Since $M \subseteq I_x^\alpha$, it follows that $\Phi_{I_x^\alpha}$ is a frame. □

Using Theorem 3.4 we find that any normalized frame $\Phi \subset \mathbb{S}$ is α-rectifying on the closed ball $\mathbb{B}_r$ if and only if

$$r \leq \inf_{x \in \mathbb{R}^n \setminus \mathcal{K}_\alpha^*} \|x\|.$$

Another consequence of the theorem, together with (ii) of Lemma 3.1 is that Φ is $\mathbf{0}$-rectifying on $\mathbb{B}_r$ if and only if Φ is $\mathbf{0}$-rectifying on $\mathbb{R}^n$. Hence, in this setting, checking a small neighborhood around the origin is already sufficient for the entire space. Note that the implication does not hold for $\alpha < \mathbf{0}$. We refer to Example 3.3 for an example.

The characterization in Theorem 3.4 further allows extending the implication from the α-rectifying property to the injectivity of C_α (Theorem 2.4) to domains that are not open or convex, such as the sphere $\mathbb{S}$, the donut $\mathbb{D}_{r,s}$, and discrete data sets.

Corollary 3.5. *Let $\Phi = (\phi_i)_{i \in I} \subset \mathbb{R}^n$, $\alpha \in \mathbb{R}^m$ and $\emptyset \neq K \subseteq O \subseteq \mathcal{K}_\alpha^*$ for O open or convex. If Φ is α-rectifying on K, then C_α is injective on K.*

Proof. By Theorem 3.4, Φ is α-rectifying on O. Since O is open or convex, by Theorem 2.5, C_α is injective on O, hence also on $K \subseteq O$. □

Note that, in general, the set $\mathcal{K}_\alpha^*$ is neither open nor convex. We can use this to demonstrate that a violation of the assumptions on K in Theorem 2.4 (i.e., not open and not convex) indeed leads to the conclusion that C_α is not injective.

Example 3.6. *Consider the frame Φ_3 defined in Example 3.3, and look at*

$$x_1 = \begin{pmatrix} 1/2 \\ \sqrt{3}/2 \end{pmatrix}, \; x_2 = \begin{pmatrix} -1/2 \\ \sqrt{3}/2 \end{pmatrix},$$

(see Figure 3 left). For $\alpha = \mathbf{0}$ we have that $x_1, x_2 \in \mathcal{K}_\alpha^*$ but also that $C_\alpha x_1 = C_\alpha x_2$. Hence, by Theorem 3.4, Φ is α-rectifying on $\mathcal{K}_\alpha^*$ but C_α is not injective on $\mathcal{K}_\alpha^*$.

A similar example can be constructed for the standard basis in $\mathbb{R}^n$, $\alpha = \mathbf{0}$, and $K = \mathbb{B}_r^+$. See Figure 2 (mid) for an illustration in $\mathbb{R}^2$. The geometric intuition from the construction of the maximal domain reveals a natural trade-off to the bias vector, where

$$\alpha' \geq \alpha \quad \Rightarrow \quad \mathcal{K}_{\alpha'}^* \subseteq \mathcal{K}_\alpha^*.$$

This should serve as the linking idea to the fact that finding a maximal bias for the α-rectifying property can reveal another perspective to Theorem 3.4. With this, we proceed to answer Question ***Q2***.

3.2 Maximal bias

We aim to construct a maximal bias for a given frame Φ and domain K. This provides a characterization of the α-rectifying property which is particularly suitable for verifying it in applications as it is straightforward to implement numerically. Our approach to this is to decompose a frame Φ into sub-frames with highly correlated frame elements and identify the smallest analysis coefficients among all points $x \in K$ associated with these sub-frames. We present two approaches for such a decomposition. Approach A is based on finding the n most correlated elements of Φ for each $x \in K$. It allows us to identify the maximal bias and with this the characterization of the α-rectifying property of Φ under reasonable assumptions. Approach B, first introduced in [17], is based on the vertex-facet configuration of the inscribing polytope associated with Φ. It gives a geometrically intuitive sufficient condition for the α-rectifying property of Φ but yields the maximal bias only in special situations. Algorithmic solutions are provided along with the theoretical results.

Approach A: Most correlated bases

The construction of the maximal bias is based on the idea of finding the least correlated frame element in the *most correlated basis* among all $x \in K$.

Definition 3.7. *Let Φ be a frame and $x \in K$. We call $\Phi_{J^*(x)}$ a most correlated basis for x if $J^*(x) \subseteq I$ satisfies that for all $J \subseteq I$ such that Φ_J is a basis it holds that*

$$\min_{j \in J} \langle x, \phi_j \rangle \leq \min_{j \in J^*(x)} \langle x, \phi_j \rangle. \tag{26}$$

We say that Φ includes a unique most correlated basis everywhere if $J^*(x)$ is unique for every $x \in K$.

To give an illustration, in the setting of Example 3.6 we have that $J^*(x_1) = \{1, 3\}$ and $J^*(x_2) = \{1, 2\}$. Alternatively, we may interpret the condition in (26) in the sense that $J^*(x)$ maximizes the functional

$$\alpha(x) = \max_{\substack{J \subseteq I \\ \Phi_J \text{ basis}}} \min_{j \in J} \langle x, \phi_j \rangle. \tag{27}$$

As a preliminary stage, we construct a maximal *constant* bias. This construction is similar to the one for the critical saturation level in [2].

Proposition 3.8. *Let Φ be a frame and $K \subseteq \mathbb{R}^n$. The maximal constant bias for Φ and K is given by $\boldsymbol{\alpha_c}$ with*

$$\alpha_c = \inf_{x \in K} \max_{\substack{J \subseteq I \\ \Phi_J \text{ basis}}} \min_{j \in J} \langle x, \phi_j \rangle = \inf_{x \in K} \min_{j \in J^*(x)} \langle x, \phi_j \rangle. \tag{28}$$

In other words, Φ is $\boldsymbol{r}$-rectifying on K if and only if $r \leq \alpha_c$.

Proof. First, we show that Φ is $\boldsymbol{\alpha_c}$-rectifying on K. Let $x \in K$ then there is a basis $\Phi_{J(x)}$ such that for all $j \in J(x)$

$$\langle x, \phi_j \rangle \geq \min_{j \in J(x)} \langle x, \phi_j \rangle \geq \alpha_c.$$

Since $\Phi_{J(x)}$ is a frame, Φ is $\boldsymbol{\alpha_c}$-rectifying on K.

Now let $r \in \mathbb{R}$ and assume that Φ is $\boldsymbol{r}$-rectifying on K. For any $x \in K$ we deduce

$$r \leq \inf_{x \in K} \min_{j \in I_x^r} \langle x, \phi_j \rangle \leq \inf_{x \in K} \min_{\substack{j \in J \subseteq I_x^r \\ \Phi_J \text{ basis}}} \langle x, \phi_j \rangle \leq \inf_{x \in K} \max_{\substack{J \subseteq I \\ \Phi_J \text{ basis}}} \min_{j \in J} \langle x, \phi_j \rangle = \alpha_c. \tag{29}$$

□

To construct a (non-constant) maximal bias vector we restrict ourselves to frames that include a unique most correlated basis everywhere. Similar to the full-spark assumption, this is a mild condition in a numerical setting since for any frame there is an arbitrarily small perturbation such that the resulting perturbed frame includes a unique most correlated basis everywhere (c.f. Lemma 2.8). If the frame is full-spark then the most correlated basis for x is given by the collection of the n frame elements which have the largest frame coefficients with x. A random frame fulfills this condition with probability one. Under this assumption, we can partition K uniquely into subsets that are associated with a frame element that belongs to a most correlated basis. For every $i \in I$ we denote the corresponding set by

$$K_i = \{x \in K : i \in J^*(x)\}. \tag{30}$$

Since every $x \in K$ has a most correlated basis $\bigcup_{i \in I} K_i = K$ indeed holds. The right picture in Figure 3 illustrates the set K_i in $\mathbb{R}^2$, and the right plot in Figure 4 illustrates the decomposition of $\mathbb{S}$ in $\mathbb{R}^3$ into K_i's. By minimizing the frame coefficients of $x \in K_i$ similar to (28) we indeed obtain a bias such that Φ possesses the α-rectifying property but it is in general not maximal.

Proposition 3.9. *Let Φ be a frame that includes a unique most correlated basis everywhere, and let $K \subseteq \mathbb{R}^n$. If $\alpha_K^\flat$ is given as*

$$\left(\alpha_K^\flat\right)_i = \inf_{x \in K_i} \langle x, \phi_i \rangle, \tag{31}$$

then Φ is $\alpha_K^\flat$-rectifying on K.

Proof. Let $x \in K$ then $\langle x, \phi_j \rangle \geq (\alpha_K^\flat)_j$ for all $j \in J^*(x)$. Since $\Phi_{J^*(x)}$ is a basis, Φ is $\alpha_K^\flat$-rectifying on K. □

Note that if K_i is empty then ϕ_i is never contained in a most correlated basis. This means that ϕ_i is irrelevant for the α-rectifying property of Φ, i.e., the corresponding bias can be chosen arbitrarily large without affecting it. We shall exclude these cases from the estimation for a maximal bias.

Moreover, although $\alpha_K^\flat$ gives a simple and intuitive indication of the critical bias, there is still room for increasing this bias while maintaining the α-rectifying property. Instead of minimizing over K_i as in Proposition 3.8, we shall minimize over the set of all points x such that $i \in J^*(x)$ and no element outside the most correlated basis is active for x and $\alpha_K^\flat$. This set is given as

$$K_i^\sharp = K_i \setminus \left(\bigcap_{y \in K_i} \bigcup_{j \notin J^*(y)} \Omega_j^{\alpha_K^\flat} \right). \tag{32}$$

Since $K_i^\sharp$ is a subset of K_i the bias values that we get from minimizing over the $K_i^\sharp$ will be larger than the ones of $\alpha_K^\flat$. However, note that if $K_i \subseteq \left(\bigcap_{y \in K_i} \bigcup_{j \notin J^*(y)} \Omega_j^{\alpha_K^\sharp} \right)$ then whenever ϕ_i belongs to the most correlated basis for x, there is additionally another active frame element from outside the most correlated basis. Similarly to when $K_i = \emptyset$, the frame element ϕ_i can then be interpreted as being redundant in the sense that it can be completely removed from Φ while preserving the α-rectifying property. We give an example of such a pathological situation.

Example 3.10. *For the frame*

$$\Phi = \left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix} \right)$$

and $K = \mathbb{B}$ we have that $\alpha_{\mathbb{B}}^\flat = \mathbf{0}$ and

$$K_2 = \left\{ x = s \cdot \begin{pmatrix} \cos t \\ \sin t \end{pmatrix} : t \in \left[-\frac{3\pi}{8}, \frac{3\pi}{8} \right], s \in [0, 1] \right\}.$$

The right picture in Figure 3 shows this setting. For every $x \in K_2$ there is an element outside the most correlated basis that is additionally active. It follows that $K_2^\sharp = \emptyset$. Hence, ϕ_2 is redundant for the α-rectifying property of Φ.

Hence, to guarantee that a maximal bias exists we shall assume $K_i^\sharp \neq \emptyset$ for all $i \in I$.

Definition 3.11. *We call $K \subseteq \mathbb{R}^n$ to be bias-exact for Φ if $K_i^\sharp \neq \emptyset$ for all $i \in I$, where $K_i^\sharp$ is defined as in* (32).

The following theorem contains the main result on the maximal bias and represents the counterpart to Theorem 3.4 on the maximal domain.

Theorem 3.12 (Maximal bias)**.** *Let Φ be a frame that includes a unique most correlated basis everywhere and $K \subseteq \mathbb{R}^n$ be bias-exact for Φ. The maximal bias for Φ and K is given by $\alpha_K^\sharp$ with*

$$\left(\alpha_K^\sharp\right)_i = \inf_{x \in K_i^\sharp} \langle x, \phi_i \rangle. \tag{33}$$

In other words, Φ is α-rectifying on K if and only if $\alpha \leq \alpha_K^\sharp$.

Proof. For the converse direction, we have that $\langle x, \phi_j \rangle \geq (\alpha_K^\sharp)_j$ for all $j \in J^*(x)$ and $x \in K_j^\sharp \neq \emptyset$. Since $\Phi_{J^*(x)}$ is a basis, Φ is $\alpha_K^\sharp$-rectifying on K.

We show the implication direction by counterposition. Let $i \in I$ and assume that Φ is α-rectifying on K where α is given by $\alpha_i = (\alpha_K^\sharp)_i + \varepsilon$ for some $\varepsilon > 0$ and $\alpha_j = (\alpha_K^\sharp)_j$ for $j \neq i$. By definition of the infimum in (33), the set $K_i^\sharp$ (32), and the construction and uniqueness of the most correlated basis (27) there is $x_0 \in K_i$ such that

$$\left(\alpha_K^\sharp\right)_i < \langle x_0, \phi_i \rangle < \left(\alpha_K^\sharp\right)_i + \varepsilon \qquad \text{and} \qquad \langle x_0, \phi_j \rangle < \left(\alpha_K^\sharp\right)_j \tag{34}$$

for all $j \notin J^*(x_0)$. This implies that the active coordinates for x_0 and α are exactly given by

$$I_{x_0}^\alpha = J^*(x_0) \setminus \{i\}.$$

As a consequence, $\Phi_{I_{x_0}^\alpha}$ is not a frame. Therefore, Φ is not α-rectifying on K_i, finishing the proof. □

Note that while assuming uniqueness of the most correlated basis everywhere is natural in the numerical setting, it excludes frames that exhibit certain symmetries, such as the frame from Example 3.10. In such a case, the computations of the biases in (31) and (33) will in general depend on the choice of the most correlated basis and therefore give ambiguous results. By including a condition that chooses one of the most correlated bases, Proposition 3.9 and Theorem 3.12 can also be formulated without the uniqueness everywhere assumption. We will not pursue this idea here.

Summarizing, Proposition 3.9 provides a simple and intuitive sufficient condition for the α-rectifying property on K via the bias vector $\alpha_K^\flat$. On the other hand, Theorem 3.12 provides a full characterization via the more complicated bias vector $\alpha_K^\sharp$ under some additional assumptions that are difficult to check in practice. In fact, except for special situations, it is unclear how to compute both of the presented bias vectors explicitly. To implement the construction of a bias for verifying injectivity in applications, we therefore present an algorithmic approach that computes $\alpha_{X_N}^\flat$ for a finite sampling set $X_N \subset K$ as an approximation for $\alpha_K^\flat$.

Algorithmic solution for Approach A. We show how $\alpha_K^\flat$ can be approximated numerically via sampling. Let $X_N = (x_k)_{k=1}^N \subset K$ be a sequence of N samples in K. By

iteratively updating the values of the bias estimation corresponding to the most correlated basis of the current sample as in Proposition 3.9, we obtain an approximation of $\alpha_K^\flat$ through the sampling set X_N. To measure the sampling error quantitatively, we use the Euclidean covering radius (or "mesh norm") of X_N for K [13], defined by

$$\rho(X_N; K) = \sup_{x \in K} \min_{1 \leq i \leq N} \|x - x_i\|. \tag{35}$$

Theorem 3.13 (Sampling-based Bias Estimation)**.** *Let $X_N = \{x_i\}_{i=1}^N \subset K$ and $\Phi \subset \mathbb{S}$ be a normalized full-spark frame. Choose $\alpha^{(0)} \in \mathbb{R}^m$ and iteratively define for all $1 \leq k \leq N$ and $i \in J^*(x_k)$*

$$(\alpha^{(k)})_i = \min\left\{\langle x_k, \phi_i\rangle, \alpha_i^{(k-1)}\right\}. \tag{36}$$

Then Φ is $\alpha^{(N)}$-rectifying on X_N. Moreover, Φ is $(\alpha^{(N)} - \rho(X_N; K))$-rectifying on K.

If the elements in Φ are not normalized, we can extend the statement above by including the norms $w = (\|\phi_i\|)_{i \in I}$, obtaining that Φ is $(\alpha^{(N)} - \rho(X_N; K) \cdot w)$-rectifying on K. If the frame elements of Φ lie in K, a good initialization is starting the bias estimation with the frame elements themselves as samples. Otherwise, we may set $(\alpha^{(0)})_i = \infty$ for all $i \in I$.

Proof of Theorem 3.13. Let $x \in K$, then there is $x_i \in X_N$ with $\|x - x_i\| \leq \rho(X_N; K)$. Furthermore, for every $j \in I_{x_i}^{\alpha^{(N)}}$ it holds that

$$\begin{aligned}(\alpha^{(N)})_j &\leq \langle x_i, \phi_j\rangle \\ &= \langle x_i - x, \phi_j\rangle + \langle x, \phi_j\rangle \\ &\leq \|x - x_i\| + \langle x, \phi_j\rangle \\ &\leq \rho(X_N; K) + \langle x, \phi_j\rangle.\end{aligned} \tag{37}$$

By rearranging (37) it follows that $\langle x, \phi_j\rangle \geq (\alpha^{(N)})_j - \rho(X_N; K)$. We deduce that $I_{x_i}^{\alpha^{(N)}} \subseteq I_x^{\alpha^{(N)} - \rho(X_N;K)}$ for all $x \in K$. Clearly, Φ is $\alpha^{(N)}$-rectifying on X_N by construction. Using (37), the second claim follows immediately. □

Implementing the algorithm described in Theorem 3.13 can be done via a Monte-Carlo approach, where X_N is a collection of N random samples on K w.r.t. some probability measure on K, see Appendix C1. We demonstrate numerical experiments in Section 3.3. Intuitively, we want a measure that guarantees a small covering radius $\rho(X_N; K)$ for large N, such that $\alpha^{(N)}$ converges to $\alpha_K^\flat$ in probability. The uniform distribution on $\mathbb{B}_r$ is a possible example. If X_N are i.i.d. uniform samples on $\mathbb{S}$ the expectation of $\rho(X_N; \mathbb{S})$ is given by

$$\mathbb{E}[\rho(X_N; \mathbb{S})] \asymp \left(\frac{\log(N)}{N}\right)^{\frac{1}{n}}. \tag{38}$$

The above estimate and more explicit tail-bound estimates for the probability distribution of $\rho(X_N; \mathbb{S})$ are derived in [28]. Unfortunately, this indicates that in high dimensions it becomes infeasible to handle the covering radius only by increasing the number of test samples N. Hence, it appears necessary to construct the sampling sequence X_N in a more structured way,

e.g., by quasi Monte-Carlo methods [7], or by sampling directly from the distribution of the dataset. With the latter approach, the number of sampling points for a good approximation of $\alpha_K^\flat$ can potentially be reduced, and the injectivity of the ReLU layer is ensured on a domain that is tailored to the dataset.

Approach B: Facets of the inscribing polytope

We obtain another natural decomposition of Φ into sub-frames with high correlation via the convex polytope that arises from taking the convex hull of the set of all elements in Φ [17]. This so-called *inscribing polytope* of Φ [29] is given by

$$P_\Phi = \{x \in \mathbb{R}^n : x = \sum_{i\in I} c_i \cdot \phi_i, c_i \geq 0, \sum_{i\in I} c_i = 1\}. \tag{39}$$

Any non-empty intersection of P_Φ with an affine half-space such that none of the interior points of P_Φ (w.r.t. the induced topology on P_Φ) lie on its boundary is called a face of P_Φ [33]. The 0-dimensional faces of P_Φ are known as vertices and the $(n-1)$-dimensional faces are called *facets*. Assuming normalized frames here ($\Phi \subset \mathbb{S}$) the set of vertices of P_Φ always coincides with the set of frame elements of Φ. Note that this is generally the case if the elements in Φ lie in a strictly convex set. Moreover, every facet is a convex polytope where the vertices coincide with a sub-collection of Φ, and every element occurs as a vertex at least once. We refer to Figure 5 for an illustration in $\mathbb{R}^3$. For a facet F, we denote the index set corresponding to its vertices by

$$I_F = \{i \in I : \phi_i \in F\}. \tag{40}$$

The following property is key, stated and proven in [17].

Lemma 3.14. *Let Φ be a frame and F be a facet of P_Φ. If $0 \notin F$, then Φ_{I_F} is a frame.*

This ensures that the facets of P_Φ provide a natural decomposition into sub-frames Φ_{I_F} of Φ. Moreover, for any facet F, there is $a \in \mathbb{R}^n$, $a \neq 0$ and $b \in \mathbb{R}$ such that $F = \{x \in P_\Phi : \langle a, x\rangle = b\}$, and therefore,

$$\begin{aligned} \langle a, \phi_k\rangle &= b, \text{ for } k \in I_F, \\ \langle a, \phi_\ell\rangle &< b, \text{ for } \ell \notin I_F. \end{aligned}$$

Hence, the vertices of any facet are a frame that is highly correlated to all points "close" to the facet. In particular, Φ_{I_F} is the most correlated basis for the normal vector a.

Now, analog to the decomposition of K using most correlated bases (30) we can decompose K into facet-specific subsets F_j^K, each associated with a facet F_j of P_Φ (according to some enumeration of the facets). A natural and practical approach in this setting is a decomposition into conical caps resulting in

$$F_j^K = \text{cone}(F_j) \cap K = \{x \in K : x = cy \ , y \in F_j, c \geq 0\}. \tag{41}$$

Figure 4 shows the facets F_j of the inscribing polytope (left), and the decomposition of $\mathbb{S}$ into spherical caps $F_j^{\mathbb{S}}$ (mid) for a i.i.d. randomly generated frame Φ with elements on $\mathbb{S}$.

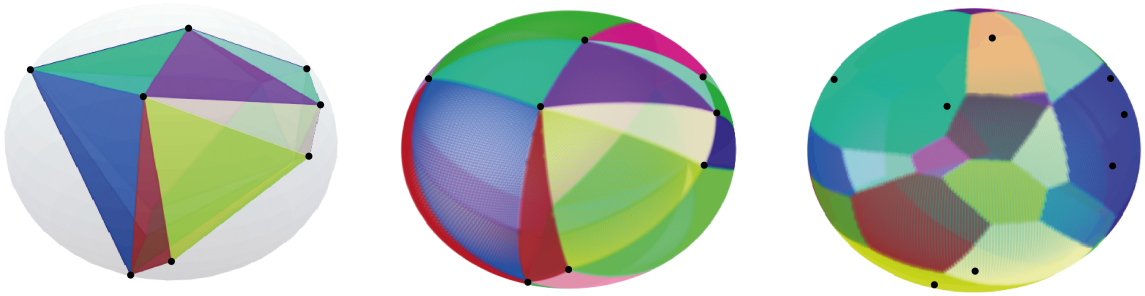

Figure 4: The three subplots show different decompositions of the sphere in $\mathbb{R}^3$. The black dots indicate the $m = 12$ frame elements of a random frame on $\mathbb{S}$. From left to right: The facets F_j of the inscribing polytope P_Φ, the associated spherical caps $F_j^{\mathbb{S}} = \operatorname{cone}(F_j) \cap \mathbb{S}$, and the spherical patches associated to different most correlated bases obtained by $J^*(x)$. While the facets provide a very intuitive and simple decomposition into sub-frames, the decomposition via the most correlated bases minimizes the correlation directly.

To guarantee that $K = \bigcup_j F_j^K$ and that 0 does not lie on any of the facets (requirement for using Lemma 3.14) we have to assume that 0 lies in the interior of P_Φ (w.r.t. the topology in $\mathbb{R}^n$). The property of Φ that ensures this was introduced in [5, Definition 1], where Φ is called *omnidirectional*. Equivalently, we can say that there is no half-space containing all elements of Φ, which can be easily verified numerically via convex optimization, see the appendix in [5]. Moreover, every non-omnidirectional frame can be made omnidirectional by including a vector constructed as the negative normalized mean of all frame elements, see Appendix B.

Remark. *The construction of the F_j^K using cones and the assumption of omnidirectionality are natural if K is centered around the origin. In other situations, one might want to come up with alternative constructions of F_j^K and a different notion of omnidirectionality that are more suited to the geometry K. In this work, we restrict ourselves to the described setting.*

Assuming omnidirectionality, we can use Lemma 3.14 to identify the minimal analysis coefficient $\langle x, \phi_i \rangle$ that can occur for any x contained in any F_j^K that contains ϕ_i. This guarantees that for any $x \in F_j^K$ the sub-frame $\Phi_{I_{F_j}}$ is active. Following [17], the procedure is called *polytope bias estimation* (PBE). The following theorem generalizes the results in [17] from $\mathbb{B}_r$ and $\mathbb{B}_r^+$ to general bounded K.

Theorem 3.15 (Polytope Bias Estimation)**.** *Let $\Phi \subset \mathbb{S}$ be a normalized omnidirectional frame and $K \subseteq \mathbb{R}^n$ bounded. Then Φ is α_K^Δ-rectifying on $K \subseteq \mathbb{R}^n$ with α_K^Δ given by*

$$\left(\alpha_K^\Delta\right)_i = \inf_{\substack{x \in F_j^K \\ j:\phi_i \in F_j}} \langle x, \phi_i \rangle. \tag{42}$$

Proof. Since Φ is omnidirectional, for any $x \in K$ there is a facet F_j such that $x \in F_j^K$. It follows from (42) that $\langle x, \phi_i \rangle \geq \left(\alpha_K^\Delta\right)_i$ for all $i \in I_{F_j}$. By Lemma 3.14, $\Phi_{I_{F_j}}$ is a frame, hence, Φ is α_K^Δ-rectifying on K. □

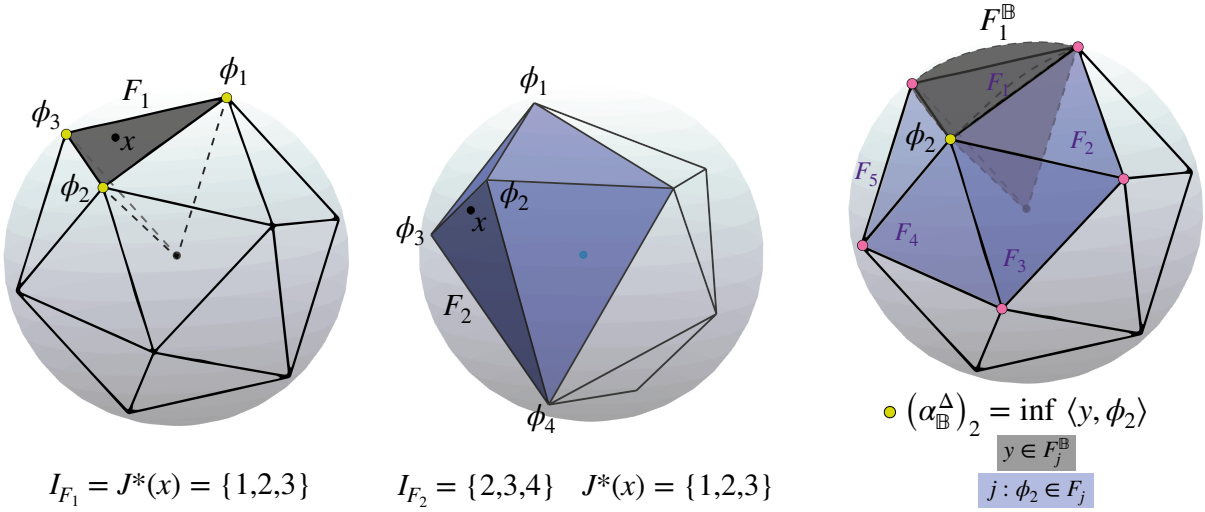

Figure 5: For the Icosahedron frame we have that $x \in F_j \Leftrightarrow I_{F_j} = J^*(x)$ (left). For less regular frames, this does not hold anymore (mid). The right picture illustrates the PBE on $\mathbb{B}$ for the Icosahedron frame. To get $\left(\alpha_{\mathbb{B}}^{\Delta}\right)_2$ the infima are taken over the points in the conical parts (dark gray area) for all adjacent facets of ϕ_2 (blue).

This procedure naturally takes the geometry of the frame into account and provides an intuitive way of estimating a large bias vector such that the frame becomes α-rectifying. Note, however, that in general this only yields a sufficient condition. A special case where it is also necessary is discussed in Lemma 3.17. In the following, we demonstrate how the PBE simplifies for specific concrete situations.

Algorithmic solution for Approach B. The advantage of the PBE is that for simple standard domains K it is easy to compute α_K^{Δ} via linear programs. In the following proposition, we formulate the PBE for five prototypical domains. We provide a detailed discussion and corresponding pseudo-code for implementing the corresponding optimizations in Appendix C.

Proposition 3.16. *Let $\Phi \subset \mathbb{S}$ be a normalized omnidirectional frame. The following holds.*

(i) Φ is α_Φ^{Δ}-rectifying on the boundary of the polytope $\partial P_\Phi = \bigcup_j F_j$ with α_Φ^{Δ} given by

$$\left(\alpha_\Phi^{\Delta}\right)_i = \min_{\substack{\ell \in I_{F_j} \\ j:\phi_i \in F_j}} \langle \phi_\ell, \phi_i \rangle.$$

(ii) Φ is $\alpha_{\mathbb{S}}^{\Delta}$-rectifying on the sphere $\mathbb{S}$ with $\alpha_{\mathbb{S}}^{\Delta}$ given by

$$\left(\alpha_{\mathbb{S}}^{\Delta}\right)_i = \min\{ \min_{\substack{x \in F_j^{\mathbb{S}} \\ j:\phi_i \in F_j}} \langle x, \phi_i \rangle, \left(\alpha_\Phi^{\Delta}\right)_i \}.$$

The inner minima are the solutions of convex linear programs. In particular, they are equal to $\left(\alpha_\Phi^{\Delta}\right)_i$ whenever they are non-negative.

(iii) Φ *is* $(r^{-1}\cdot\alpha_{\mathbb{B}}^{\Delta})$*-rectifying on the donut* $\mathbb{D}_{r,s}$ *for* $0\le s<r$ *with* $\alpha_{\mathbb{B}}^{\Delta}$ *given by*

$$\left(\alpha_{\mathbb{B}}^{\Delta}\right)_i=\min\{s,\left(\alpha_{\mathbb{S}}^{\Delta}\right)_i\}.$$

The case $s=0$ *yields a bias estimation for the closed ball* $\mathbb{B}_r$ *(Figure 5).*

(iv) Let $J^+=\{j\in I: F_j\cap\mathbb{R}^n_+\neq\emptyset\}$ *and* $I^+=\bigcup_{j\in J^+}I_{F_j}$ *then* Φ *is* $(r^{-1}\cdot\alpha_{\mathbb{B}^+}^{\Delta})$*-rectifying on the non-negative part of the ball* $\mathbb{B}_r^+$ *with* $\alpha_{\mathbb{B}^+}^{\Delta}$ *given by*

$$\left(\alpha_{\mathbb{B}^+}^{\Delta}\right)_i=\begin{cases}\left(\alpha_{\mathbb{B}}^{\Delta}\right)_i & \text{for } i\in I^+\\ s_i & \text{else,}\end{cases}\tag{43}$$

where $s_i\in\mathbb{R}$ *is arbitrary.*

(v) If $\alpha_{\Phi}^{\Delta}\ge 0$*, then* $(s\cdot\phi_i)_{i\in I}$ *for* $s\ge 0$ *is* $\left(s\cdot\alpha_{\Phi}^{\Delta}\right)$*-rectifying on* $(\mathring{\mathbb{B}}_s)^{\mathrm{c}}=\mathbb{R}^n\setminus\mathring{\mathbb{B}}_s$.

Proof. The points $(i)-(iv)$ are direct consequences of [17, Theorem 4.4 and Theorem 4.6], where detailed proofs can be found. To show (v), note that $(s\cdot\phi_i)_{i\in I}$ has the same combinatorial facet structure as Φ and therefore it is still omnidirectional. In particular,

$$(\mathring{\mathbb{B}}_s)^{\mathrm{c}}=\{x=sty: y\in\mathbb{S}, t\ge 1\}.$$

The statement follows by $\langle sty,\phi_i\rangle\ge s\cdot\left(\alpha_{\Phi}^{\Delta}\right)_i$ for $t\ge 1$ and $s\ge 0$. □

Other than the Monte-Carlo sampling-based approach, this bias estimation procedure yields a deterministic sufficient condition for the α-rectifying property, which is necessary only in special situations. In general, there are two properties of the inscribing polytope P_Φ that affect the PBE in a way that the estimated biases become smaller.

(1) The more vertices a facet has, the smaller the infima in (42) become. In the case where all facets have the minimal number of n vertices, P_Φ is called *simplicial*. It is known that a polytope with vertices that are i.i.d. uniform samples on $\mathbb{S}$ is simplicial with probability one [9].

(2) The decomposition of K using cones in (41) is natural and convenient for implementation, but leads to a sub-optimal partition if P_Φ is geometrically very irregular, i.e., the sizes of the facets are significantly different. In such a scenario, for $x\in F_j^K$ where F_j is a very large facet, there might be a smaller neighboring facet F_k which provides larger analysis coefficients for x and, hence, a better estimation. For an illustration of such a situation see the center plot in Figure 5.

For frames with inscribing polytopes that are simplicial and regular, we can show that the PBE indeed yields a maximal bias. We set $K=\mathbb{S}$.

Lemma 3.17. *Let* $\Phi\subset\mathbb{S}$ *be a normalized omnidirectional frame such that* P_Φ *is simplicial and for all* $i\in I$ *it holds that* $\langle\phi_i,\phi_k\rangle=\langle\phi_i,\phi_\ell\rangle$ *for all* ϕ_k,ϕ_ℓ *sharing a facet with* ϕ_i*. Then* Φ *is* α*-rectifying on* $\mathbb{S}$ *if and only if* $\alpha\le\alpha_{\mathbb{S}}^{\Delta}$.

Proof. The converse direction directly follows from Theorem 3.15. We show the implication direction by counterposition. At first note that the regularity condition implies that $\alpha_{\mathbb{S}}^{\Delta}$ is constant. Let $\alpha \in \mathbb{R}^m$ be such that $\alpha_i > \left(\alpha_{\mathbb{S}}^{\Delta}\right)_i$ for fixed $i \in I$. There is a facet F with $i \in I_F$ and $x^* \in F^{\mathbb{S}}$ satisfying $\langle x^*, \phi_i \rangle = \left(\alpha_{\mathbb{S}}^{\Delta}\right)_i < \alpha_i$. By regularity, it follows that $I_{x^*}^{\alpha} = I_F \setminus \{i\}$. Since P_Φ is simplicial $\Phi_{I_F \setminus \{i\}}$ is not a frame, hence, Φ is not α-rectifying on $\mathbb{S}$. □

Examples for this are frames whose inscribing polytopes are convex regular polytopes in $\mathbb{R}^n$. In a more general sense, we expect the PBE to be very effective for frames with evenly distributed frame elements.

Remark. *The two bias estimation procedures described in Approach A and B are fundamentally linked. For each facet F of P_Φ the hole radius of F is defined as the Euclidean distance from the boundary to the center of its spherical cap $F^{\mathbb{S}}$* (41)*. The Euclidean covering radius* (35) *of Φ for $\mathbb{S}$ is the largest hole radius among all facets [31].*

3.3 Numerical experiments

For a given sampling sequence X_N and a full-spark assumption on Φ, the numerical implementation of the sampling-based bias estimation in Theorem 3.13 is straightforward. Similarly, there are convex hull algorithms available, such that the implementation of the cases of the PBE from Proposition 3.16 is straightforward, too. Besides the pseudo-code found in the appendix, we provide concrete implementations in Python, together with the code for reproducing all experiments in this section in the accompanying repository https://github.com/danedane-haider/Alpha-rectifying-frames.

Evolution towards injectivity (Approach A)

We demonstrate the basic functionality of the Monte-Carlo sampling-based algorithm. For this, we choose the frame Φ and the sampling sequences X_N to consist of i.i.d. uniform samples on $\mathbb{B}$. For every step in the approximation of $\alpha_{\mathbb{B}}^{\flat}$ we measure the proportion of samples from an unseen test sampling sequence Y_M (also i.i.d. uniform on $\mathbb{B}$), for which Φ is $\alpha^{(k)}$-rectifying. Figure 6 shows the empirical mean and variance over 1000 independent trials of this procedure for two different dimensions, $n = 3$ (left) and $n = 30$ (right), and three different redundancies ($2, 3.3$, and 9), respectively. We observe that in some configurations the associated ReLU layer is injective already after a few hundred iterations. In others, it takes up to 1000 iterations. It is especially fast for ReLU layers in low dimensions with low redundancy. This can be explained by the fact that for draws of Φ which yield very unevenly distributed points on $\mathbb{B}$ (which happens more likely in high dimensions with high redundancy) the iterative scheme struggles to update $\alpha^{(k)}$ efficiently, which results in the procedure taking particularly long. In general, all tested examples became injective reliably.

Effect of redundancy (Approach A)

We use approximations of $\alpha_{\mathbb{B}}^{\flat}$ to verify the injectivity of ReLU layers with random weights and biases systematically for different redundancies. Thereby, we numerically verify the conjecture stated in [24] that the transition from non-injectivity to injectivity happens at

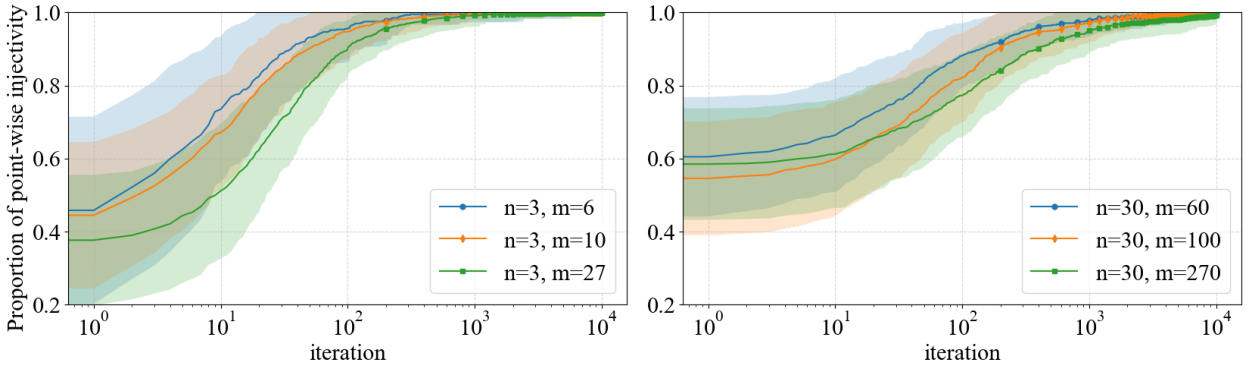

Figure 6: Per iteration k, the plots show the proportion of a test sample sequence Y_M where the ReLU layer with bias $\alpha^{(k)}$ is injective. Left: $n = 3$. Right: $n = 30$; Both for redundancies $2, 3.3$, and 9. The ReLU layers are becoming injective reliably after about 10^4 iterations. The procedure is fastest for low dimensions and low redundancy.

a redundancy $q \in (6.6979, 6.6981)$ in a non-asymptotic setting. We let both, the frame Φ and the sampling sequence X_N contain i.i.d. standard normal points and compute $\alpha^{(N)}$ with $N = 5 \cdot 10^5$ for all redundancy settings where $2 \leq n \leq 30$ and $n \leq m \leq 150$. By Theorem 3.13, we know that for every setting Φ is $\alpha^{(N)}$-rectifying on X_N. Inspired by the asymptotic expression for the covering radius in (38), we subtract a correcting term of $\rho^*(n, N) = 0.05 \cdot \left(\frac{\log(N)}{N}\right)^{\frac{1}{n}}$ to compensate for insufficient amount of sampling in higher dimensions. This yields that for every setting we have that Φ is $\big(\alpha^{(N)} - \rho^*(n, N)\big)$-rectifying on $\mathbb{B}$ with high probability. The factor 0.05 was chosen experimentally.

To test if the ReLU layer associated with one of the realizations of Φ is injective for a given bias α we have to verify that $\alpha \leq \alpha^{(N)} - \rho^*(n, N)$. We compare three settings for biases with i.i.d. normal values with mean zero and variances,

$$(i)\ \sigma^2 = 0 \qquad (ii)\ \sigma^2 = 0.1 \qquad (iii)\ \sigma^2 = 1$$

Figure 7 shows the results for the three settings from left to right. Setting (i) is the one where the conjecture in [24] was formulated. Looking at the solid magenta line in Figure 7 (left) we can observe that our method is capable of numerically reproducing the conjecture. On the injectivity of random ReLU layers with non-zero bias, there are no theoretical results in the literature so far. Hence, our approach yields some novel insights here. For small variance (Setting (ii)) we observe that the clear boundary from the previous setting blurs out. For the standard variance in setting (iii) this behavior further intensifies. Note that an according change of the variances in the distribution of Φ or X_N instead gives the same result. These observations show that the way the bias in a ReLU layer is initialized has a big influence on its injectivity.

Approximation of the maximal bias (Approach A & B)

In the setting of Lemma 3.17 we have shown that the PBE yields a maximal bias. This allows us to study the approximation of the sampling-based approach to the maximal bias

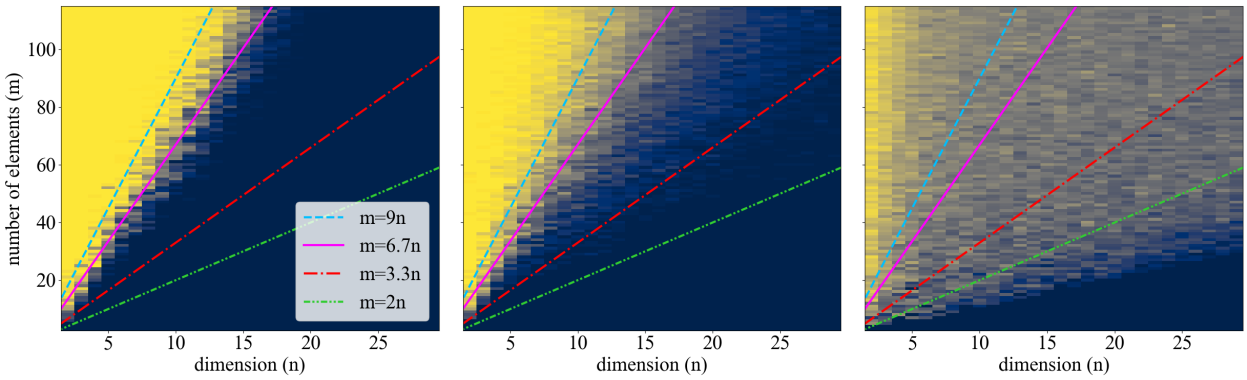

Figure 7: The plots show the injectivity behavior of random ReLU layers for different redundancies ($2 \leq n \leq 30, n \leq m \leq 150$). The brightness encodes the proportion of the values of the given bias α that are smaller than the ones in $(\alpha^{(N)} - \rho^*(n, N))$ for $N = 5 \cdot 10^5$. For values of one (yellow), the ReLU layer is injective on $\mathbb{B}$. In all settings, the samples in X_N are i.i.d. standard normal and the bias i.i.d. normal with different variances. From left to right: $\sigma^2 = 0$, $\sigma^2 = 0.1$, and $\sigma^2 = 1$. For $\sigma^2 = 0$ we observe the clear transition from non-injective to injective at a redundancy of 6.7 (solid magenta line) that aligns well with the conjecture from the literature. For larger variance, the transition blurs out quickly and prevents us from predicting clear statements about injectivity.

$\alpha^{\mathbb{S}}$ by the PBE over the iterations. Figure 8 shows this in the example of the Tetrahedron frame, where the PBE yields that $\alpha^{\mathbb{S}} = \frac{1}{\sqrt{3}}$. We plot $\|\alpha^{(k)} - \alpha^{\mathbb{S}}\|$ as k increases and find that the approximation is very slow, which emphasizes the superiority of the polytope approach in this setting.

Remarks on the limitations

Not surprisingly, both algorithms suffer from high dimensionality. For the sampling-based approach, the asymptotic behavior of the covering radius (38) indicates that it may become infeasible to reach a good approximation of $\alpha_K^\flat$ only by increasing the number of samples. Yet, with some experimenting on the factor, we can use expression (38) to effectively compensate for insufficient sampling in high dimensions. It has particularly high potential when injectivity is only required on specific data points of interest. In such a situation, the sampling set can be constructed in a custom data-driven way that respects the distribution of the data.

For the polytope bias estimation, the numerical computation of the convex hull to obtain the vertex-facet relations becomes infeasible in high dimensions. A possible remedy is to use dimensionality reduction and do the bias estimation in a lower dimensional space. The benefits of the method are that injectivity follows deterministically and that it comes with a lot of intuition an can be studied further using more advanced tools from convex geometry.

With this, we conclude the part of the paper that is concerned with the analysis of the injectivity of a ReLU layer. The last chapter is dedicated to the reconstruction of the input from the output of an injective ReLU layer.

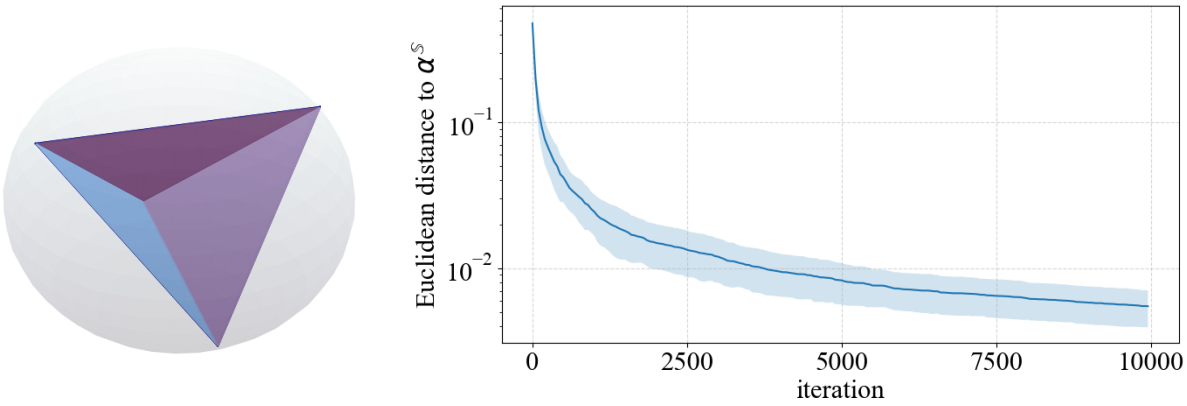

Figure 8: Left: The inscribing polytope for the Tetrahedron frame. Right: The Euclidean distance of $\alpha^{(k)}$ to the maximal bias of the Tetrahedron frame $\alpha^{\mathbb{S}}$ on a log y-scale over 10^5 iterations. In cases where the inscribing polytope is a simplicial and regular polytope, the sampling-based bias estimation is sub-optimal, and the polytope approach already gives the maximal bias.

4 Duality and Reconstruction

If a ReLU layer C_α is injective, there is an inverse mapping that can infer any input from the output. This extends to the possibility of synthesizing new data that correspond to arbitrary coefficient vectors in the image of the ReLU layer, or studying the connection of single weights to the input by perturbing the corresponding output coefficient, thereby being able to interpret the output values.

In [2], the authors propose to reconstruct the input from its saturated frame coefficients via a custom-modified version of the frame algorithm [11]. This idea can also be adapted to ReLU layers. Our main focus lies on another approach, where we propose to construct explicit perfect reconstruction formulas in the form of locally linear operators.

4.1 ReLU-synthesis

First, we recall the concept of a *dual* frame, which is closely tied to two operators. The first one is the *synthesis operator* which maps the frame coefficients back to the input space as

$$\begin{aligned} D : \mathbb{R}^m &\to \mathbb{R}^n \\ (c_i)_{i\in I} &\mapsto \sum_{i\in I} c_i \cdot \phi_i. \end{aligned}$$

The application of this operator is realized via the multiplication by the transpose of the analysis matrix from the left, $D = C^\top$. The second operator arises as the concatenation of analysis, followed by synthesis, also known as the *frame operator*,

$$\begin{aligned} S : \mathbb{R}^n &\to \mathbb{R}^n \\ x &\mapsto \sum_{i\in I} \langle x, \phi_i \rangle \cdot \phi_i. \end{aligned}$$

In matrix notation, $S = DC$. The frame operator is positive and self-adjoint, and if Φ is a frame, it is additionally invertible. Hence, one can write any $x \in \mathbb{R}^n$ as

$$x = S^{-1}Sx = \sum_{i \in I} \langle x, \phi_i \rangle \cdot S^{-1}\phi_i. \tag{44}$$

The collection $\tilde{\Phi} = (S^{-1}\phi_i)_{i \in I}$ is called the *canonical dual frame* for Φ. Denoting the synthesis operator associated with $\tilde{\Phi}$ by $\tilde{D}$, then Equation (44) is equivalent to

$$\tilde{D}Cx = x. \tag{45}$$

In other words, $\tilde{D}$ is a left-inverse of C (given by the pseudo-inverse of C) [10]. This can be thought of as reconstructing x from its frame coefficients using the canonical dual frame $\tilde{\Phi}$. If Φ is redundant ($m > n$), there are infinitely many different possibilities of constructing a left-inverse of C. All of them can be interpreted as the synthesis operator of a (non-canonical) dual frame. Using this machinery, we can define a reconstruction operator for C_α analogously to (45). Note, however, that unless $I_x^\alpha \neq I$ for all $x \in K$, there is not *one* reconstruction operator for all $x \in K$.

Definition 4.1. *The ReLU-synthesis operator associated with the collection $\Phi = (\phi_i)_{i \in I} \subset \mathbb{R}^n$, the bias $\alpha \in \mathbb{R}^m$, and the index set $J \subseteq I$ is defined by*

$$\begin{aligned} D_J^\alpha : \mathbb{R}^m &\to \mathbb{R}^n \\ (c_i)_{i \in I} &\mapsto \sum_{i \in J} (c_i + \alpha_i) \cdot \phi_i. \end{aligned} \tag{46}$$

Note that $\mathbb{R}^m$ is fixed as domain, the sum, however, runs over the index set J. When using the ReLU-synthesis operator associated with a dual frame of Φ_J we obtain a reconstruction formula in the spirit of (45).

Theorem 4.2 (ReLU-dual I)**.** *Let Φ be α-rectifying on K for $\alpha \in \mathbb{R}^m$ and choose $x_0 \in K$. Let $\widetilde{\Phi}_{I_{x_0}^\alpha} = (\tilde{\phi}_i)_{i \in I_{x_0}^\alpha}$ be any dual frame for $\Phi_{I_{x_0}^\alpha}$, and $\tilde{D}_{I_{x_0}^\alpha}^\alpha$ the associated ReLU-synthesis operator. Then for all $x \in K$ such that $I_{x_0}^\alpha \subseteq I_x^\alpha$ it holds that*

$$\tilde{D}_{I_{x_0}^\alpha}^\alpha C_\alpha x = x. \tag{47}$$

Proof. For $x \in K$ with $I_{x_0}^\alpha \subseteq I_x^\alpha$, the operator composition in (47) reduces to the usual frame decomposition with $\Phi_{I_{x_0}^\alpha}$,

$$\begin{aligned} \tilde{D}_{I_{x_0}^\alpha}^\alpha C_\alpha x &= \sum_{i \in I_{x_0}^\alpha} (\max(0, \langle x, \phi_i \rangle - \alpha_i) + \alpha_i) \cdot \tilde{\phi}_i \\ &= \sum_{i \in I_{x_0}^\alpha} \langle x, \phi_i \rangle \cdot \tilde{\phi}_i = x. \end{aligned}$$

□

The condition $I_{x_0}^\alpha \subseteq I_x^\alpha$ for a reference vector $x_0 \in K$ means that we may use the same left-inverse $\tilde{D}^\alpha_{I^\alpha_{x_0}}$ for all input elements x that share at least all active elements with $\Phi_{I^\alpha_{x_0}}$. By re-writing the condition on the level of index sets, we can alternatively choose a reference sub-space via an index set J instead of fixing a reference vector x_0.

Corollary 4.3 (ReLU-dual II)**.** *Let Φ be α-rectifying on K for $\alpha \in \mathbb{R}^m$ and choose $J \subseteq I$ such that Φ_J is a frame. Let $\tilde{\Phi}_J$ be a dual frame for Φ_J and $\tilde{D}^\alpha_J$ the associated ReLU-synthesis operator. Then for all $x \in K$ with $J \subseteq I_x^\alpha$ it holds that*

$$\tilde{D}^\alpha_J C_\alpha x = x.$$

The approach in the above corollary is particularly useful in the context of the bias estimation procedures described in Section 3.2: Given a decomposition of Φ into sub-frames, either by all different most correlated bases or via the facets of P_Φ, we can compute all associated left-inverses in advance and use them for reconstruction on demand. More precisely, for every $x \in K$ there is a sub-frame associated with a most correlated bases $J^*(x)$ such that $\tilde{D}^\alpha_{J^*(x)}$ is a left-inverse of C_α. Similarly, if Φ is omnidirectional, then for every $x \in K$ there is a facet F such that $\tilde{D}^\alpha_{I_F}$ is a left-inverse of C_α. In both cases, there are only finitely many such sub-frames. Summarizing, we have the following.

Corollary 4.4. *Let Φ be α-rectifying on K. For any $x \in K$ there is $J \subseteq I$ such that $\tilde{D}^\alpha_J$ is a left-inverse of C_α.*

The numerical implementation of the ReLU-synthesis is straightforward when using canonical duals of the sub-frames Φ_J. A detailed discussion and corresponding pseudo-code can be found in Appendix C.

Excursion: Reconstruction from PReLU layers

There are various modifications of the ReLU activation function, one of them being the parametrized ReLU, or PReLU, given by $\mathrm{PReLU}_\gamma = \max(\gamma s, s)$ with $0 < \gamma \leq 1$ [18]. As this is an injective activation function, the associated PReLU layer with weights given by Φ and any bias α is injective if and only if Φ is a frame. In this case, for any $x \in K$ we obtain a left-inverse of the PReLU layer by

$$\begin{aligned} &\tilde{D}^\alpha_\gamma : \mathbb{R}^m \to \mathbb{R}^n \\ &\quad (c_i)_{i\in I} \mapsto \sum_{i \in I^\alpha_x} (c_i + \alpha_i) \cdot \tilde{\phi}_i + \sum_{i \in I \setminus I^\alpha_x} \gamma^{-1} (c_i + \alpha_i) \cdot \tilde{\phi}_i, \end{aligned} \tag{48}$$

where $\widetilde{\Phi} = (\tilde{\phi}_i)_{i\in I}$ is any dual frame for Φ.

4.2 The frame algorithm for ReLU layers

The frame algorithm is an iterative scheme that constructs a sequence of vectors in $\mathbb{R}^n$ from given frame coefficients $(\langle x, \phi_i\rangle)_{i\in I}$ that converges to the input x exponentially fast [11]. This sequence $(y_k)_{k=0}^\infty$ is defined as $y_0 = \mathbf{0}$ and

$$y_{k+1} = y_k + \lambda \sum_{i \in I} \big(\langle x, \phi_i\rangle - \langle y_k, \phi_i\rangle\big)\phi_i \tag{49}$$

for $k \geq 0$. Letting A, B be the optimal frame bounds for Φ then for $0 < \lambda < \frac{B}{2}$ we have for all $k \geq 0$ that $\|x - y_{k+1}\| \leq \kappa_\lambda \|x - y_k\|$, where $\kappa_\lambda = \max\{|1 - \lambda A|, |1 - \lambda B|\}$. Note that the optimal value for the parameter λ is $\frac{2}{A+B}$. In practice, the frame algorithm is a great tool to do reconstruction in situations, where computing the exact solution with the canonical dual (or any dual) frame becomes too expensive.

Clearly, the procedure can be directly applied for the reconstruction of x from the output of a ReLU layer by reducing the sum in (49) to run only over I_x^α (the active frame elements). To see that this is really what we want, note that for all $i \in I_x^\alpha$ we have $\langle x, \phi_i \rangle = \mathrm{ReLU}(\langle x, \phi_i \rangle - \alpha_i) + \alpha_i$. Therefore, the differences in the sum in (49) over I_x^α are indeed taken between the values of the unbiased output of the ReLU layer and $\langle y_k, \phi_i \rangle$. Hence, if our frame Φ is α-rectifying on K then for any $x \in K$ we obtain a ReLU-reconstruction sequence $(y_k)_{k=0}^\infty$ that satisfies $\|x - y_{k+1}\| \leq \kappa_{x,\lambda} \|x - y_k\|$ for all $k \geq 0$, where $\kappa_{x,\lambda} = \max\{|1 - \lambda A_x|, |1 - \lambda B_x|\}$, and A_x, B_x are the optimal frame bounds for $\Phi_{I_x^\alpha}$.

However, we can do better than this. Following the idea in [2], we can extend the frame algorithm by using the bias values α_i for all inactive frame elements as a proxy for the lost frame coefficients. This gives an algorithm that always outperforms the naive approach in the setting where we use the optimal parameter for the full frame.

Proposition 4.5 (ReLU frame algorithm)**.** *Let Φ be α-rectifying for $\alpha \in \mathbb{R}^m$ on $K \subseteq \mathbb{R}^n$. For any $x \in K$, let the ReLU-reconstruction sequence $(y_k)_{k=0}^\infty$ be given as $y_0 = \mathbf{0}$ and*

$$y_{k+1} = y_k + \lambda \sum_{i \in I_x^\alpha} \big(\langle x, \phi_i \rangle - \langle y_k, \phi_i \rangle\big) \phi_i + \lambda_0 \sum_{i \in I_{y_k}^\alpha \setminus I_x^\alpha} \big(\alpha_i - \langle y_k, \phi_i \rangle\big) \phi_i \tag{50}$$

for all $k \geq 0$. Let $\lambda = \frac{2}{A+B}$. If $\lambda_0 = 0$ then $\|x - y_{k+1}\| \leq \kappa_\lambda \|x - y_k\|$ for all $k \geq 0$, where $\kappa_\lambda = 1 - A_x \frac{2}{A+B}$. If $\lambda_0 = \frac{2}{A+B}$ and for every $k \geq 0$ the optimal lower frame bound for $\Phi_{I_x^\alpha}$ is strictly less than the optimal frame bound for $\Phi_{I_x^\alpha \cup I_{y_k}^\alpha}$ then there is $0 < \varepsilon_{x,y_k} < 1$ such that

$$\|x - y_{k+1}\| \leq (1 - \varepsilon_{x,y_k}) \kappa_\lambda \|x - y_k\|. \tag{51}$$

Since the proof is analogous to the one of Theorem 5.2 in [2], we omit it here and refer to the appendix.

We note that using $\lambda = \lambda_0 = \frac{2}{A+B}$ as parameters in (50) is very natural as we may not always want to compute the optimal frame bounds for each activated sub-frame, but instead use a reasonably universal parameter that we only have to compute once. However, the design of the algorithm leaves it open to also use other parameters. A comprehensive analysis of which parameters work well is left as an open problem. Note further that the assumption on the frame bounds for $\Phi_{I_x^\alpha}$ and $\Phi_{I_x^\alpha \cup I_{y_k}^\alpha}$ is very mild. In fact, it is always fulfilled as long as y_k does not lie in the span of one of the eigenvectors of the associated frame operator. In the worst case where this happens for all $k \geq 0$ then $\varepsilon_{x,y_k} = 0$ and the extended frame algorithm is as fast as the naive one.

4.3 Stability of the reconstruction

In this section, we revisit a result by [26] on the lower Lipschitz bound of a ReLU layer and translate it into the language of frame theory as done in [4] and [2]. As a small extension,

we present a result on the local lower Lipschitz stability. A general revision of the Lipschitz stability analysis of ReLU layers is left for future work.

Definition 4.6. *A frame Φ allows κ-stable α-rectification on K if there is $\kappa > 0$ s.t.*

$$\|y - z\|^2 \leq \kappa \cdot \|C_\alpha y - C_\alpha z\|^2 \tag{52}$$

holds for all $y, z \in K$.

First of all, we note that if Φ is α-rectifying on K then a frame-type inequality as (4) holds for C_α. That is, there are constants $0 < A_\alpha \leq B_\alpha < \infty$ such that

$$A_\alpha \cdot \|x\|^2 \leq \|C_\alpha x\|^2 \leq B_\alpha \cdot \|x\|^2 \tag{53}$$

for all $x \in K$. It is easy to see that the largest possibility of choosing the lower bound A_α in (53) is the smallest lower frame bound among all possible active sub-frames $\Phi_{I_x^\alpha}$ with $x \in K$. Analogously, the smallest possibility for the upper bound B_α coincides with the largest of all upper frame bounds. In [26] it was shown that any α-rectifying frame allows $(2mA_\alpha^{-1})$-stable $\mathbf{0}$-rectification on $\mathbb{R}^n$ but not (A_α^{-1})-stable $\mathbf{0}$-rectification. It remains an open problem whether this statement extends to the case of non-zero biases and if the factor m can be replaced by a constant.

In general, active sub-frames are naturally prone to have a bad lower frame bound, such that A_α^{-1} may become very large, and the problem becomes globally ill-conditioned. To get a better understanding of how stable the reconstruction process is for smaller portions of the data, we shall investigate the lower Lipschitz property locally.

Definition 4.7. *For $x_0 \in K$, a frame Φ allows κ_{x_0}-stable α-rectification near x_0 if there is $\varepsilon > 0$ and $\kappa_{x_0} = \kappa(x_0) > 0$ such that*

$$\|y - z\|^2 \leq \kappa_{x_0} \cdot \|C_\alpha y - C_\alpha z\|^2 \tag{54}$$

holds for all $y, z \in \mathring{B}_\varepsilon(x_0)$.

Since locally, a ReLU layer is a linear map we might hope to use A_α^{-1} as a lower bound. In general, however, we can only guarantee that an α-rectifying frame allows A_α^{-1}-stable β-rectification for $\beta < \alpha$.

Proposition 4.8. *Let Φ be α-rectifying on K. For $x_0 \in K$ let $J = J(x_0) \subseteq I_{x_0}^\alpha$ be such that Φ_J is a frame with lower frame bound A_J. Then Φ allows A_J^{-1}-stable β-rectification near x_0 for $\beta < \alpha$.*

Proof. Let $x_0 \in K$. There is $J = J(x_0) \subseteq I_{x_0}^\alpha$ such that Φ_J is a frame with frame bounds $0 < A_J \leq B_J < \infty$. Analog to the global case, the bi-Lipschitz condition

$$A_J \cdot \|y - z\|^2 \leq \|C_\alpha y - C_\alpha z\|^2 \leq B_J \cdot \|y - z\|^2 \tag{55}$$

holds for all $y, z \in K$ with $I_y^\alpha, I_z^\alpha \supseteq J$. Let $\beta \in \mathbb{R}^m$ such that $\langle x_0, \phi_i \rangle \geq \alpha_i > \beta_i$ for all $i \in J$. Hence, there is $\varepsilon > 0$ sufficiently small such that for $y, z \in \mathring{B}_\varepsilon(x_0)$ we have $I_y^\beta, I_z^\beta \supseteq J$. Since (55) still holds for C_β, we get that Φ allows A_J^{-1}-stable β-rectification near x_0. □

Corollary 4.9. *Let Φ be α-rectifying on K, and $x_0 \in K$ such that $\langle x_0, \phi_i \rangle > \alpha_i$ for all $i \in J$. Then Φ allows A_J^{-1}-stable α-rectification near x_0.*

4.4 Image of ReLU layers

In the context of our work, it is specifically interesting to know what the image of a ReLU layers looks like. The impact is two-fold.

(1) To study the injectivity of a ReLU layer that applies to the output of a previous one it is crucial to know how its image looks like.

(2) Reconstructing samples from the image of a ReLU layer yields a valuable method to generate new consistent data and understand the effect of their ReLU layer.

We give a partial answer to this question in the following. For bounded K we can use the upper bound in (53) to find the smallest closed non-negative ball in $\mathbb{R}^m$ that contains $C_\alpha(K)$.

Lemma 4.10. *Let Φ be α-rectifying on K bounded with $M = \sup_{x\in K} \|x\|$. Letting B_α denote the largest optimal upper frame bound among all sub-frames $\Phi_{I_x^\alpha}$ with $x \in K$, then*

$$C_\alpha(K) \subseteq \mathbb{B}^+_{\sqrt{B_\alpha}M}, \tag{56}$$

where $\sqrt{B_\alpha}M$ is the minimal radius.

Proof. By (53), for $x \in K$ we have that $\|C_\alpha x\|^2 \leq B_\alpha \|x\|^2 \leq B_\alpha M^2$. The application of the ReLU function then corresponds to the projection onto the non-negative part of $\mathbb{B}_{\sqrt{B_\alpha}M}$, i.e., $\mathbb{B}^+_{\sqrt{B_\alpha}M}$. Since all estimations are sharp, the claim follows. □

As already mentioned, understanding the images of specific sets under ReLU layers is crucial to understanding how data is processed and passed on to the next layer. The above lemma is just a small step towards revealing this knowledge which can be used for unraveling certain behaviors of neural networks, and further, enhancing their interpretability and transparency.

5 Conclusion

This manuscript studies the injectivity of ReLU layers and the exact recovery of input vectors from their output using frame theory as a tool. Among many basic properties and insights about ReLU layers on bounded domains, the main theoretical contribution is three different characterizations of the injectivity of ReLU layers that together provide a complete picture of its injectivity behavior as a non-linear deterministic operator. A significant portion of the research focuses on the computation of a maximal bias for a given frame Φ and a domain K, such that the associated ReLU layer is injective on K. This characterization is particularly interesting as it allows us to apply the theoretical results in practical applications. We discuss two different methods to approach this, both of which have distinct advantages and disadvantages, and provide algorithmic solutions to compute approximations of a maximal bias in practice. The second part of this work is devoted to the derivation of reconstruction formulas for injective ReLU layers, based on the concept of duality in frame theory. A brief local stability analysis of the reconstruction operator completes the discussion.

In summary, this paper provides a methodology on how to study the channeling of information in ReLU layers with biases and given input data, made possible by using frame theory as a tool. The results are designed in a general, yet, accessible way such that they may stimulate further theoretical research, but are also directly applicable in practice. While we are pleased to contribute to advancing the understanding of these fundamental and ubiquitous building blocks of neural networks, many critical aspects remain to be explored. A central question in this context is how information propagates through ReLU *networks*, so how can we rigorously characterize the injectivity of the composition of ReLU layers.

Acknowledgement

D. Haider is recipient of a DOC Fellowship of the Austrian Academy of Sciences at the Acoustics Research Institute (A 26355). The work of P. Balazs was supported by the FWF projects LoFT (P 34624) and NoMASP (P 34922). The authors would particularly like to thank Daniel Freeman for his valuable input during very enjoyable discussions and Hannah Eckert for her work on the numerical experiments.

References

[1] B. Alexeev, J. Cahill, and D. G. Mixon, *Full spark frames*, Journal of Fourier Analysis and Applications, 18 (2012), p. 1167–1194.

[2] W. Alharbi, D. Freeman, D. Ghoreishi, B. Johnson, and N. L. Randrianarivony, *Declipping and the recovery of vectors from saturated measurements*, ArXiv, abs/2402.03237 (2024).

[3] R. Balan, P. Casazza, and D. Edidin, *On signal reconstruction without phase*, Applied and Computational Harmonic Analysis, 20 (2006), pp. 345–356.

[4] A. S. Bandeira, J. Cahill, D. G. Mixon, and A. A. Nelson, *Saving phase: Injectivity and stability for phase retrieval*, Applied and Computational Harmonic Analysis, 37 (2014), pp. 106–125.

[5] J. Behrmann, S. Dittmer, P. Fernsel, and P. Maass, *Analysis of invariance and robustness via invertibility of ReLU-networks*, arXiv, abs/1806.09730 (2018).

[6] A. Bora, A. Jalal, E. Price, and A. G. Dimakis, *Compressed sensing using generative models*, in International Conference on Machine Learning (ICML), D. Precup and Y. W. Teh, eds., vol. 70 of Proceedings of Machine Learning Research, PMLR, 2017, pp. 537–546.

[7] A. Breger, M. Ehler, and M. Gräf, *Points on manifolds with asymptotically optimal covering radius*, Journal of Complexity, 48 (2018), p. 1–14.

[8] J. Bruna, A. Szlam, and Y. Lecun, *Signal recovery from pooling representations*, in International Conference on Machine Learning (ICML), 2014, pp. 1585–1598.

[9] C. Buchta and J. Müller, *Random polytopes in a ball*, Journal of Applied Probability, 21 (1984), p. 753–762.

[10] P. G. Casazza and G. Kutyniok, *Finite frames: Theory and applications*, Springer, 2012.

[11] O. Christensen, *An Introduction to Frames and Riezs Bases*, Birkhäuser, 2003.

[12] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, *Fast and accurate deep network learning by exponential linear units (ELUs)*, in International Conference on Learning Representations (ICLR), 2016.

[13] S. B. Damelin and V. Maymeskul, *On point energies, separation radius and mesh norm for s-extremal configurations on compact sets in Rn*, Journal of Complexity, 21 (2005), pp. 845–863.

[14] X. Glorot, A. Bordes, and Y. Bengio, *Deep sparse rectifier neural networks*, in International Conference on Artificial Intelligence and Statistics, G. Gordon, D. Dunson, and M. Dudík, eds., vol. 15 of Proceedings of Machine Learning Research, 2011, pp. 315–323.

[15] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.

[16] V. K. Goyal, J. Kovačević, and J. A. Kelner, *Quantized frame expansions with erasures*, Applied and Computational Harmonic Analysis, 10 (2001), pp. 203–233.

[17] D. Haider, P. Balazs, and M. Ehler, *Convex geometry of ReLU-layers, injectivity on the ball and local reconstruction*, in International Conference on Machine Learning (ICML), 2023.

[18] K. He, X. Zhang, S. Ren, and J. Sun, *Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification*, in International Conference on Computer Vision (ICCV), 2015.

[19] L. Huang, J. Qin, Y. Zhou, F. Zhu, L. Liu, and L. Shao, *Normalization techniques in training DNNs: Methodology, analysis and application*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 45 (2023), pp. 10173–10196.

[20] K. Kothari, A. Khorashadizadeh, M. de Hoop, and I. Dokmanić, *Trumpets: Injective flows for inference and inverse problems*, in Conference on Uncertainty in Artificial Intelligence, C. de Campos and M. H. Maathuis, eds., vol. 161 of Proceedings of Machine Learning Research, PMLR, 2021, pp. 1269–1278.

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, *ImageNet classification with deep convolutional neural networks*, in Advances in Neural Information Processing Systems, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, eds., vol. 25, Curran Associates, Inc., 2012.

[22] Y. A. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, *Efficient BackProp*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 9–48.

[23] A. L. Maas, A. Y. Hannun, and A. Y. Ng, *Rectifier nonlinearities improve neural network acoustic models*, in International Conference on Machine Learning (ICML), 2013.

[24] A. Maillard, A. S. Bandeira, D. Belius, I. Dokmanić, and S. Nakajima, *Injectivity of ReLU networks: Perspectives from statistical physics*, 2023, arXiv:2302.14112.

[25] V. Nair and G. E. Hinton, *Rectified linear units improve restricted Boltzmann machines*, in International Conference on International Conference on Machine Learning (ICML), 2010, p. 807–814.

[26] M. Puthawala, K. Kothari, M. Lassas, I. Dokmanić, and M. de Hoop, *Globally injective ReLU networks*, Journal of Machine Learning Research, 23 (2022), pp. 1–55.

[27] M. Puthawala, M. Lassas, I. Dokmanic, and M. De Hoop, *Universal joint approximation of manifolds and densities by simple injective flows*, in International Conference on Machine Learning (ICML), 2022.

[28] A. Reznikov and E. B. Saff, *The covering radius of randomly distributed points on a manifold*, International Mathematics Research Notices, (2016), pp. 6065–6094.

[29] R. M. Richardson, L. Wu, and V. H. Vu, *An inscribing model for random polytopes*, Twentieth Anniversary Volume: Discrete & Computational Geometry, (2009), pp. 1–31.

[30] T. Salimans and D. P. Kingma, *Weight normalization: A simple reparameterization to accelerate training of deep neural networks*, in International Conference on Neural Information Processing Systems (NeurIPS), 2016, p. 901–909.

[31] R. Seri, *Asymptotic distributions of covering and separation measures on the hypersphere*, Discrete and Computational Geometry, 69 (2022).

[32] J. C. Ye, Y. Han, and E. Cha, *Deep convolutional framelets: A general deep learning framework for inverse problems*, SIAM Journal on Imaging Sciences, 11 (2018), pp. 991–1048.

[33] G. M. Ziegler, *Lectures on polytopes*, vol. 152, Springer Science & Business Media, 2012.

Appendix

A - Remarks on Admissibility and Directed Spanning Sets

With this short comment, we aim to complete the circle between three perspectives to characterize the injectivity of a ReLU layer on $\mathbb{R}^n$. In the work by Bruna et al. [8] the injectivity of a ReLU layer, or *half-rectification operator* was linked to an admissibility condition of a $J \subseteq I$. There, J is called admissible for Φ and α if

$$\bigcap_{i\in J}\{x \in \mathbb{R}^n : \langle x, \phi_i\rangle > \alpha_i\} \cap \bigcap_{i\notin J}\{x \in \mathbb{R}^n : \langle x, \phi_i\rangle < \alpha_i\} \neq \emptyset.$$

Puthawala et al. in [26] already pointed out that Proposition 2.2 in [8] is not exactly equivalent to the injectivity of a ReLU layer. However, with a slight modification to

$$\bigcap_{i\in J}\{x \in \mathbb{R}^n : \langle x, \phi_i\rangle \geq \alpha_i\} \cap \bigcap_{i\notin J}\{x \in \mathbb{R}^n : \langle x, \phi_i\rangle < \alpha_i\} \neq \emptyset,$$

this stands in direct relation to the index sets I_x^α, as introduced in Definition 2.2 in the present manuscript. Indeed, with this modified definition, J is α-admissible for Φ if and only if there is $x \in \mathbb{R}^n$ such that $J = I_x^\alpha$. As a consequence, we have that the equivalence of (i) and (ii) in Corollary 1 here with $K = \mathbb{R}^n$, Proposition 2.2 in [8] with the modified admissibility condition, and Theorem 2 in [26] are equivalent.

B - Omnidirectionality

We prove the statement about omnidirectionality mentioned in Approach B. of Section 3.2: By adding a single vector, any non-omnidirectional frame can be made omnidirectional. Furthermore, we recall how omnidirectionality can be checked numerically as in [5].

Lemma 5.1. *Let Φ be a non-omnidirectional frame, then*

$$\Phi' = \left(\Phi, -\frac{\sum_{i\in I}\phi_i}{\|\sum_{i\in I}\phi_i\|}\right)$$

is omnidirectional.

Proof. At first, note that if $\sum_{i\in I}\phi_i = 0$, then Φ is already omnidirectional. Let $c_1, ..., c_m = \frac{1}{1+\|\sum_{i\in I}\phi_i\|}$ and $c_{m+1} = \frac{\|\sum_{i\in I}\phi_i\|}{1+\|\sum_{i\in I}\phi_i\|}$ and $\phi_{m+1} = -\frac{\sum_{i\in I}\phi_i}{\|\sum_{i\in I}\phi_i\|}$, then

$$\sum_{i=1}^{m+1} c_i \cdot \phi_i = 0. \tag{57}$$

Note that $c_i > 0$ for all $i = 1, \ldots, m+1$ and $\sum_{i=1}^{m+1} c_i = 1$. Hence, in (57) we wrote 0 as a convex combination of *all* elements of Φ'. This implies that $0 \in \mathring{P}_{\Phi'}$, hence Φ' is omnidirectional. □

Regarding the verification of omnidirectionality, let D be the synthesis matrix associated with Φ, i.e., it consists of the column vectors ϕ_i for $i \in I$. Then, verifying omnidirectionality is equivalent to the existence of a solution for the convex optimization problem $\min \|Dc\|$ subject to $c > 0$.

Figure 9: Left: The frame is omnidirectional. Right: The frame is not omnidirectional.

C - Algorithms

We discuss the implementation of the presented algorithmic approaches and provide detailed pseudo-code. Our Python implementations can be found under https://github.com/danedane-haider/Alpha-rectifying-frames.

C1. Sampling-based bias estimation. Given a frame Φ, a data domain K, and a sequence of samples $X_N \subset K$, Algorithm 1 demonstrates the sampling-based bias estimation presented in Theorem 3.13. The samples $x_k \in X_N$ can be chosen to be random samples, e.g., $x_k \sim \mathcal{U}(K)$ for suitable K. Assuming the frame to be full-spark, then $J^*(x_k)$ consists of the indices for the largest n frame coefficients $(\langle x_k, \phi_i \rangle)_{i \in I}$.

Algorithm 1 Sampling-based approach for approximating $\alpha_K^\flat$

Input: Φ, X_N, K
initialize $(\alpha^{(0)}) = \alpha_\Phi^\Delta$, $k = 0$
for z in X_N **do**
  compute $(\langle z, \phi_i \rangle)_{i \in I}$
  get $J^*(z)$
  update $(\alpha^{(k+1)})_i \leftarrow \min\{\langle z, \phi_i \rangle, (\alpha^{(k)})_i\}$ for all $i \in J^*(z)$
  $k = k + 1$
end for

The for-loop can be replaced with a while-loop conditioned on $\|\alpha^{(k+steps)} - \alpha^{(k)}\| > \varepsilon > 0$, where the parameter *steps* determines how many updates should be done before checking the condition. This is very useful to avoid early stopping.

C2. Polytope bias estimation. Given a frame Φ and a bounded domain K. The vertex-facet relations for P_Φ encoded in I_{F_j} can be computed with convex hull algorithms, e.g., using the property `simplices` from `scipy.spatial.ConvexHull` in Python. In the following, we demonstrate how to compute the biases from Proposition 3.16.

(*i*) $K = \partial P_\Phi$: Recall that α_Φ^Δ is given by

$$\left(\alpha_\Phi^\Delta\right)_i = \min_{\substack{\ell \in I_{F_j} \\ j:\phi_i \in F_j}} \langle \phi_\ell, \phi_i \rangle. \tag{58}$$

Algorithm 2 PBE for ∂P_Φ

Input: Φ
compute I_{F_j} for all facets
for $j = 1, \ldots, \#\text{facets}$ **do**
  $\beta_j = \min_{k<\ell \in I_{F_j}} \langle \phi_k, \phi_\ell \rangle$
end for
for $i = 1, \ldots, m$ **do**
  $\left(\alpha_\Phi^\Delta\right)_i = \min_{j:i \in I_{F_j}} \beta_j$
end for

(*ii*) $K = \mathbb{S}$: Recall that $\alpha_\mathbb{S}^\Delta$ is given by

$$\left(\alpha_\mathbb{S}^\Delta\right)_i = \min\{ \min_{\substack{y \in F_j^\mathbb{S} \\ j:\phi_i \in F_j}} \langle y, \phi_i \rangle, \left(\alpha_\Phi^\Delta\right)_i \}. \tag{59}$$

First, we show that for fixed $i \in I_{F_j}$,

$$\min_{\substack{y \in F_j^\mathbb{S} \\ j:\phi_i \in F_j}} \langle y, \phi_i \rangle \tag{60}$$

can be computed using convex linear programs. Letting $C_{I_{F_j}}, D_{I_{F_j}}$ denote the analysis and synthesis operator associated with $\Phi_{I_{F_j}}$, respectively. We can write any $x \in F_j^\mathbb{S}$ as $x = \sum_{\ell \in I_{F_j}} c_\ell \phi_\ell = D_{I_{F_j}} c$ for some vector $c \geq 0$. So for any $i \in I_{F_j}$ the solution of (60) is found by solving the linear program

$$\begin{aligned} \min_{j:\phi_i \in F_j} \ & \left(C_{I_{F_j}} D_{I_{F_j}} c\right)_i \\ & \text{subject to } c \geq 0 \\ & \|D_{I_{F_j}} c\|_2 = 1. \end{aligned} \tag{61}$$

If $\left(\alpha_\Phi^\Delta\right)_i < 0$, then the above minimum is negative since $\left(\alpha_\mathbb{S}^\Delta\right)_i \leq \left(\alpha_\Phi^\Delta\right)_i < 0$. Therefore, we can replace $\|D_{I_{F_j}} c\|_2 = 1$ by $\|D_{I_{F_j}} c\|_2 \leq 1$ making the problem convex.

(*iii*) $K = \mathbb{D}_{r,s}$: Let $0 \leq s < r$ and recall that $\alpha_\mathbb{B}^\Delta$ is given by

$$\left(\alpha_\mathbb{B}^\Delta\right)_i = \min\{s, \left(\alpha_\mathbb{S}^\Delta\right)_i\}. \tag{62}$$

One gets the general case by scaling with r^{-1}. The case $s = 0$ yields a bias estimation for $\mathbb{B}_r$. Since (62) depends on $\alpha_\mathbb{S}^\Delta$ in a trivial way, we omit the algorithm.

Algorithm 3 PBE for $\mathbb{S}$

Input: Φ
compute α_Φ^Δ
for $i = 1, \dots, m$ **do**
  if $\left(\alpha_\Phi^\Delta\right)_i \geq 0$ **then**
    $\left(\alpha_\mathbb{S}^\Delta\right)_i \leftarrow \left(\alpha_\Phi^\Delta\right)_i$
  else
    $\left(\alpha_\mathbb{S}^\Delta\right)_i \leftarrow$ solution of (61)
  end if
end for

(iv) $K = \mathbb{B}_r^+$: Let $e \in \mathbb{R}^m$ be arbitrary and recall that $\alpha^{\mathbb{B}^+}$ is given by

$$\left(\alpha_{\mathbb{B}^+}^\Delta\right)_i = \begin{cases} \left(\alpha_\mathbb{B}^\Delta\right)_i & \text{for } i \in I^+ \\ s & \text{else,} \end{cases} \tag{63}$$

where s is arbitrary. The crux here is to compute the index set $I^+ = \bigcup_{j \in J^+} I_{F_j}$ defined via $J^+ = \{j \in I : F_j \cap \mathbb{R}_+^n \neq \emptyset\}$. One way to verify that $F_j \cap \mathbb{R}_+^n \neq \emptyset$ is to check the feasibility of the convex optimization problem

$$\begin{aligned} \min \ & \|D_{I_{F_j}} c\|_2 \\ \text{subject to } & c \geq 0 \\ & \sum_i c_i = 1. \end{aligned} \tag{64}$$

If (64) has a solution for the facet F_j, then there is $c \in \mathbb{R}_+^n$ that can be written as a convex linear combination of the vertices of F_j, hence, $F_j \cap \mathbb{R}_+^n \neq \emptyset$.

C3. ReLU-duals and reconstruction. For $J \subseteq I$ we denote by C_J the analysis operator for the collection Φ_J. This corresponds to the $|J| \times n$ matrix C_J consisting of the row vectors ϕ_i for $i \in J$. Algorithm 4 describes how to build the synthesis matrices for doing ReLU-synthesis for an index set J from Corollary 4.3.

Algorithm 4 Construction of the matrix for ReLU-synthesis

Input: $\Phi, J = \{i_1, \dots, i_{|J|}\} \subseteq I$ such that $\Psi = \Phi_J$ is a frame
$S_J^{-1} \leftarrow \left((C_J)^\top C_J\right)^{-1}$

$$\tilde{D}_J \leftarrow \begin{pmatrix} | & | & & | \\ S_J^{-1}\psi_{i_1} & S_J^{-1}\psi_{i_2} & \cdots & S_J^{-1}\psi_{i_{|J|}} \\ | & | & & | \end{pmatrix}$$

Note that if we insert columns of zeros in $\tilde{D}_J$ for all coordinates that have not been activated, then the resulting matrix would be the synthesis operator associated with a non-canonical dual frame for Φ in the classical sense. Applying it to an output vector that is restricted to only the coordinates in J, as presented here, computes the corresponding input by avoiding unnecessary multiplications with zeros. Finally, we want to perform the actual reconstruction: Given $z = C_\alpha x_0$ for some $x_0 \in K$, we can read off $I^\alpha_{x_0}$ from z directly (under the assumption that $\langle x_0, \phi_i\rangle \neq \alpha_i$ for all $i \in I$). Assuming that we have a suitable list of index sets of sub-frames (e.g., all most correlated bases of all facet sub-frames), finding those that are contained in $I^\alpha_{x_0}$ is easy. Choose one, say J. The choice $J = I^\alpha_{x_0}$ is valid too. Then the matrix $\tilde{D}_J$ provides reconstruction as follows.

Algorithm 5 Applying the ReLU-synthesis: Reconstruction of x_0 from $z = C_\alpha x_0$

Input: Φ, α, z
find $J \subseteq I^\alpha_{x_0}$ such that Φ_J is a frame
$z \leftarrow z + \alpha$ (unbias)
restrict to J via $\zeta \leftarrow (z_i)_{i\in J}$
reconstruct $\tilde{D}_J\zeta = x_0$ (see Algorithm 4)

C4. ReLU frame algorithm. For the sake of completeness, we give a proof of Proposition 4.5 that follows the lines of the one of Theorem 5.2 in [2].

Proof of Proposition 4.5. At first note that applying the (classical) frame algorithm (49) with $\Phi_{I^\alpha_x}$ and $\lambda = \frac{2}{A+B}$ gives the constant $\kappa_\lambda = 1 - A_x\frac{2}{A+B}$ since $\frac{2}{A+B} < \frac{2}{A_x+B_x}$. Now, for every $i \in I^\alpha_{y_k} \setminus I^\alpha_x$ we define

$$\gamma_i = \frac{\alpha_i - \langle y_k, \phi_i\rangle}{\langle x - y_k, \phi_i\rangle}$$

and note that $0 \leq \gamma_i < 1$. It is easy to see that the extended frame algorithm (50) constructs the same sequence of vectors $(y_k)_{k=0}^\infty$ as applying the (classical) frame algorithm (49) using the frame $\Phi_\gamma = (\phi_i)_{i\in I^\alpha_x} \cup (\gamma^{1/2}\phi_i)_{i\in I^\alpha_{y_k}}$. Let A' denote the optimal lower frame bound for $\Phi_{I^\alpha_x \cup I^\alpha_{y_k}}$ and A'' the one for Φ_γ. Then, by assumption that $A_x < A'$, together with $\gamma_i < 1$, we have $A_x < A' < A'' \leq A$. Applying the convergence result for the (classical) frame algorithm for Φ_γ gives

$$\|x - y_{k+1}\| \leq \left(1 - A''\tfrac{2}{A+B}\right)\|x - y_k\| \leq (1 - \varepsilon_{x,y_k})\underbrace{\left(1 - A_x\tfrac{2}{A+B}\right)}_{\kappa_\lambda}\|x - y_k\|, \tag{65}$$

where $0 < \varepsilon_{x,y_k} < 1$. □