Representing Piecewise-Linear Functions by Functions with Minimal Arity

Christoph Koutschan, Anton Ponomarchuk, Josef Schicho

Abstract

Any continuous piecewise-linear function $F\colon \mathbb{R}^n \to \mathbb{R}$ can be represented as a linear combination of maxima of at most $n+1$ affine-linear functions. It is well-known that this upper bound is sharp, that is, for every n there exists a function that is no linear combination of maxima of at most n arguments. In this paper, we give a method to determine, for each F, the minimal number k such that F can be written as a linear combination of maxima of at most k arguments.

Index Terms

Continuous piecewise linear functions, neural networks, machine learning, ReLU activation function.

I. Introduction

A continuous piecewise-linear (CPWL) function $F\colon \mathbb{R}^n \to \mathbb{R}$ can be represented in various ways: as a difference between two convex CPWL functions [1], as a lattice polynomial [2]–[4], or in some cases by a canonical piecewise-linear representation [5]. These representations are interchangeable. The function F represented as a lattice polynomial can be transformed into a difference of two convex CPWL functions, see [6]. Let $q \in \mathbb{N}$ be the number of polyhedral regions in the tesselation defined by F. By [7], the function F can be represented as a ReLU-based neural network with at most $\lceil \log_2(n+1) \rceil + 1$ layers. Another bound is given in [8], F can be written as a ReLU-based neural network with at most $\lceil 2\log_2 q \rceil + 1$ layers, where $q \in \mathbb{N}$ is the number of polyhedral regions in the tesselation defined by F. In [9], it was shown that the function F can be represented as a linear combination of $\max$ functions with at most $n+1$ affine-linear functions as arguments, i.e., for a finite set of indices I we have:

$$F = \sum_{i \in I} e_i \max(F_{i,1}, \ldots, F_{i,s}), \tag{1}$$

where $s \le n+1$ and $F_{i,j}\colon \mathbb{R}^n \to \mathbb{R}$ are affine-linear functions, and $e_i = \pm 1$ Moreover, a set of operations was introduced [9] that transform a lattice polynomial into a decomposition of the form (1). In the rest of this paper, an affine-linear function inside a $\max$ function will be called an argument. The upper bound of arguments for $\max$ functions in Eq. (1) is tight, in the sense that there exist functions which do not allow a representation as a linear combination of maxima with fewer arguments. This result was obtained independently by two different research groups [10], [11]. Both give the witness function $\max(0, x_1, \ldots, x_n)$, which cannot be represented as a linear combination of maxima with less than $n+1$ arguments. However, Example I.1, the function $\max(0, x_1, x_2) + \max(0, -x_1, -x_2)$ can be represented as a linear combination of maxima with two arguments – see Example I.1. The representation is not unique. This leads to the question: For a given CPWL function F, what is the minimal number of arguments of maxima that is necessary to represent the function F as in Eq. (1)? This paper gives an answer to the question above.

This paper is structured as follows. In Section II, we provide basic information from polyhedral geometry, see [12]. Also, we introduce and expand the notion of piecewise-constant functions necessary to formulate and prove Theorem III.3. In Section III, we develop the set of conditions in which a CPWL function F can be represented as a linear combination of $\max$ functions with at most $k \le n+1$ arguments. Also, we prove that k is minimal for the given decomposition of the function F.

Example I.1. We consider the two functions $G_1(x_1, x_2) \coloneqq \max(0, x_1, x_2)$ and $G_2(x_1, x_2) \coloneqq \max(0, -x_1, -x_2)$, that are defined on $\mathbb{R}^2$, see Fig. 1. For the functions G_1 and G_2 the following equality holds:

$$\max(0, x_1, x_2) + \max(0, -x_1, -x_2) = \max(0, x_1, x_2, -x_1, -x_2, x_1 - x_2, x_2 - x_1). \tag{2}$$

C. Koutschan is with Johann Radon Institute for Computational and Applied Mathematics, Linz, Austria
A. Ponomarchuk is with Johann Radon Institute for Computational and Applied Mathematics, Linz, Austria
J. Schicho is with Research Institute for Symbolic Computation, Johannes Kepler University, Hagenberg im Mühlkreis, Austria.
The research reported in this paper has been partly funded by BMK, BMDW, and the Province of Upper Austria in the frame of the COMET Programme managed by FFG in the COMET Module S3AI.

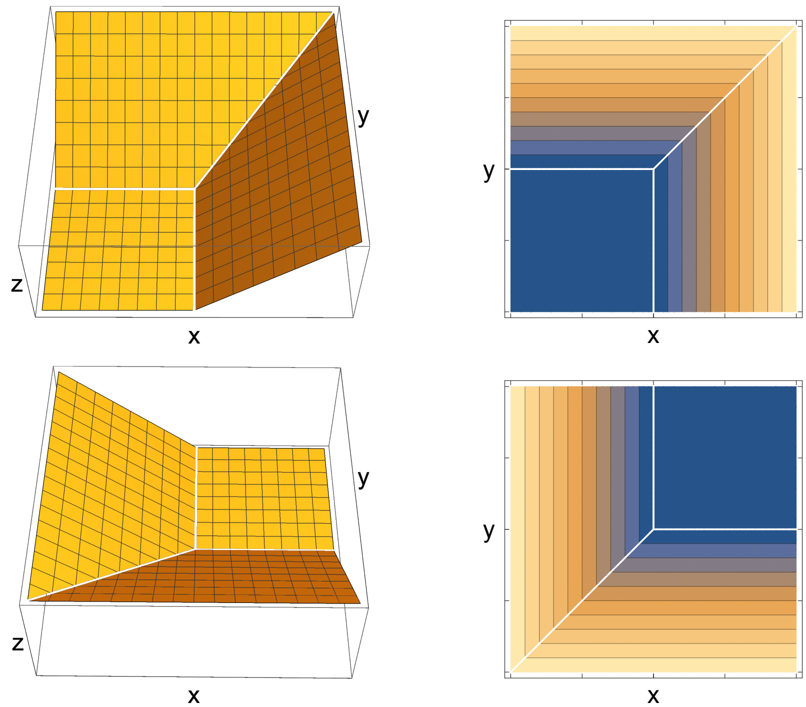

Fig. 1. Two CPWL functions, $G_1(x, y) = \max(0, x, y)$ (first row), and $G_2(x, y) = \max(0, -x, -y)$ (second row). Each row shows the function $\mathbb{R}^2 \to \mathbb{R}$ as a three-dimensional plot (first column) and the function's tessellation as a contour plot (second column). For a hyperplane $H_1 = \{(x, y) \in \mathbb{R}^2 \mid x = y\}$, the delta functions $\Delta_{\nabla G_1}((H_1))$ and $\Delta_{\nabla G_2}((H_1))$ are non-constant.

By [10], [11], the functions G_1 and G_2 cannot be represented as a linear combination of $\max$ functions with at most two arguments. However, the function $G_1 + G_2$, see Fig. 2, can be represented as a linear combination of $\max$ of at most two arguments. The function $G_1 + G_2$ is decomposed as follows:

$$(G_1 + G_2)(x_1, x_2) = \max(x_1, x_2) + \max(-x_2, x_1 - x_2) + \max(-x_1, x_2 - x_1). \tag{3}$$

It is sufficient to show that the right side in Eq. (3) is equivalent to the function $\max(0, x_1, x_2, -x_1, -x_2, x_1 - x_2, x_2 - x_1)$. By applying basic algebraic transformations, we get:

$$\begin{aligned}
&\max(x_1, x_2) + \max(-x_2, x_1 - x_2) + \max(-x_1, x_2 - x_1) = \\
&\max(-x_1, x_2 - x_1) + \max(0, x_1, 2x_1 - x_2, x_1 - x_2) = \\
&\max(0, x_1, x_2, -x_1, -x_2, x_1 - x_2, x_2 - x_1).
\end{aligned} \tag{4}$$

Thus, Eq (2) and Eq. (4) imply that Eq. (3) holds. Theorem III.3 shows why the function $G_1 + G_2$ can be represented as a linear combination of $\max$ with at most two arguments, and the functions G_1 and G_2 cannot.

II. Piecewise-constant functions

Before formulating the main theorem of the paper, we first provide supplementary notions from polyhedral geometry and piecewise-constant functions.

Definition II.1. Let $X \subset \mathbb{R}^n$ be a subset of $\mathbb{R}^n$. The relative interior $\operatorname{relint}(X)$ is the interior of X with respect to the embedding of X into its affine hull. A polyhedron is a connected subset $P \subseteq \mathbb{R}^n$ which is relatively open, i.e., it is equal to its relative interior, and such that its closure $\overline{P}$ is a finite union of convex polyhedra (a convex polyhedron is a the convex hull of a finite set of points). The set $\overline{P} \setminus \operatorname{relint}(P)$ is called the relative boundary $\operatorname{relbd}(P)$ of P.

In general, a CPWL function splits the input space into a finite set of polyhedra, but this tesselation is not necessarily a polyhedral complex. However, it always defines a stratified family, in the sense of the following definition.

Definition II.2. Let $\mathcal{P}$ be a family of polyhedra in $\mathbb{R}^n$. We say that $\mathcal{P}$ is stratified if it satisfies the following set of properties:

(a) $\mathcal{P}$ is finite.
(b) P is relatively open, for all $P \in \mathcal{P}$.
(c) $\overline{P_1} \cap P_2 \neq \emptyset \Rightarrow P_2 \subseteq \overline{P_1}$ for all $P_1, P_2 \in \mathcal{P}$.

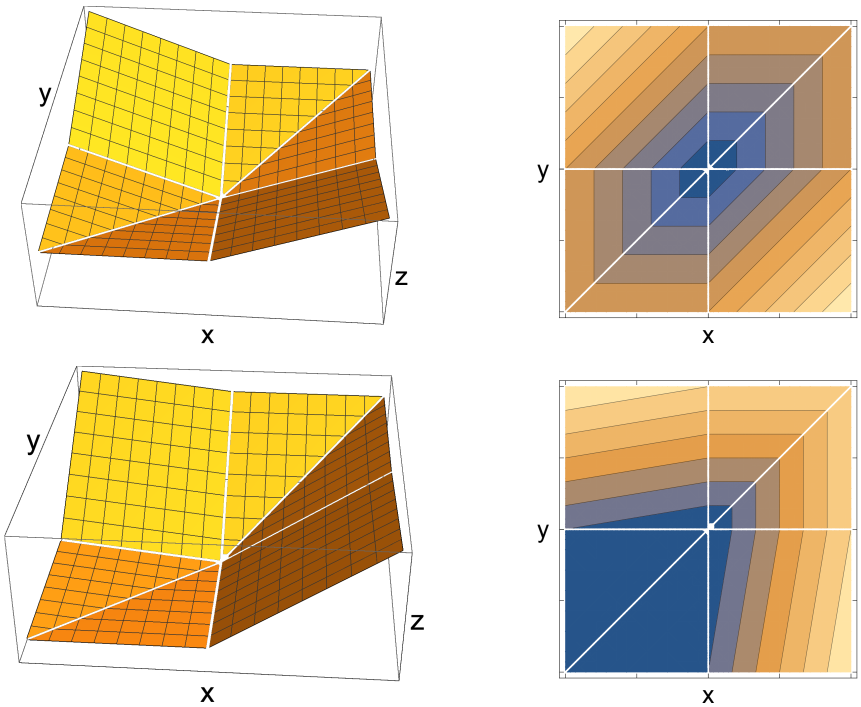

Fig. 2. Two CPWL functions, $G_3(x, y) = \max(0, x, y) + \max(0, -x, -y)$ (first row), and $G_4(x, y) = 6\max(0, x, y) + \max(0, -x, -y)$ (second row). Each row shows the function $\mathbb{R}^2 \to \mathbb{R}$ as a three-dimensional plot (first column) and the function's tessellation as a contour plot (second column). For any flag $\mathcal{H}$ of length 1, the delta function $\Delta_{\nabla G_3}(\mathcal{H})$ is constant. For a hyperplane $H_1 = \{(x, y) \in \mathbb{R}^2 \mid x = y\}$, the delta function $\Delta_{\nabla G_4}((H_1))$ is non-constant.

(d) The polyhedra are disjoint, and their union is an affine linear subspace $H \subseteq \mathbb{R}^n$. This affine subspace is called support and it is denoted by $|\mathcal{P}|$.

Definition II.3. Let $\mathcal{P}$ be a stratified family of polyhedra. We say that a point $x \in |\mathcal{P}|$ is generic with respect to $\mathcal{P}$ if the unique polyhedron containing x has maximal dimension $k := \dim(|\mathcal{P}|)$.

Note that the k-dimensional polyhedra $P \in \mathcal{P}$ cover "almost" the whole affine subspace H because there are only finitely many polyhedra of smaller dimension that are not included in the union of k-dimensional ones.

The concept of piecewise-constant function was applied to analyse convex CPWL functions and their underlying polyhedral complexes [10]. We extend this notion to the general case of CPWL functions.

Definition II.4. Let $\mathcal{P}$ be a finite set of n-dimensional polyhedra that are open, pairwise disjoint, and the union of which is dense in $\mathbb{R}^n$. Let G be an abelian group. A function $f\colon \bigcup \mathcal{P} \to G$ is called a piecewise-constant (PC) if it is constant in every polyhedron in $\mathcal{P}$. We assume that there are no two polyhedra in $\mathcal{P}^*(f)$ with the same value that share a boundary of dimension $n-1$ – in this case, we may simply replace the union of these two polyhedra by the relative interior of the closure of the union (which is the union of the two plus the shared boundary).

The function f does not need to be defined everywhere in $\mathbb{R}^n$. By $\mathcal{P}^0(f)$, we denote the family of n-dimensional polyhedra on which the function f is defined. For any index $i \in \{1, \dots, n\}$, the family of $(n-i)$-dimensional polyhedra $\mathcal{P}^i(f)$ is defined as follows:

$$\mathcal{P}^i(f) := \big\{\operatorname{relint}(\overline{P_1} \cap \overline{P_2}) \bigm| P_1, P_2 \in \mathcal{P}^{i-1}(f),\ \dim(\overline{P_1} \cap \overline{P_2}) = n - i\big\}.$$

Then, any PC function f induces a stratified family of polyhedra $\mathcal{P}^*(f)$ in the following way:

$$\mathcal{P}^*(f) := \big\{P \bigm| P \in \mathcal{P}^i(f),\ \text{for all } i \in \{0, \dots, n\}\big\}.$$

The input space $\mathbb{R}^n$ is the support of the stratified family of polyhedra $\mathcal{P}^*(f)$. The union of all generic points forms with respect to $\mathcal{P}^*(f)$ is equal to the domain of definition $\operatorname{def}(f)$ of f. It has the following form:

$$\operatorname{def}(f) = \bigcup_{P \in \mathcal{P}^0(f)} P.$$

The closure of all generic points $x \in \mathbb{R}^n$ in which the PC function f is non-zero is called the support $\operatorname{supp}(f)$ of f, i.e., it has the following form:

$$\operatorname{supp}(f) = \bigcup_{P \in \mathcal{P}^0(f), f(P) \neq 0} \overline{P}.$$

Later in the paper, we will show that $\operatorname{supp}(f)$ contains information about possible representations of the PC function f. Another helpful notion will be the lineality space of the PC function f.

Definition II.5. Let f be a PC function. The lineality space $\mathrm{L}(f)$ of the function f is the following set of points:

$$\mathrm{L}(f) \coloneqq \{y \in \mathbb{R}^n \mid f(x+y) = f(x) \text{ for all generic } x \in \mathbb{R}^n\}.$$

Lemma II.6. *Let f be a PC function. The lineality space $\mathrm{L}(f) \subseteq \mathbb{R}^n$ of f is a linear subspace.*

Proof. Note that for any non-trivial vectors $v_1, v_2 \in \mathrm{L}(f)$ and for any point $x \in \operatorname{def}(f)$ the following equality holds:

$$f(x + v_1 + v_2) = f(x + v_1) = f(x),$$

i.e., $v_1 + v_2 \in \mathrm{L}(f)$ and $\mathrm{L}(f)$ is closed under addition. Obviously, the vector 0 belongs to $\mathrm{L}(f)$. For non-zero $v \in \mathrm{L}(f)$ also $-v$ is in $\mathrm{L}(f)$. The last thing to show is that if $v \in \mathrm{L}(f)$, then also $\lambda v \in \mathrm{L}(f)$ for any $\lambda \in \mathbb{R}_{\geq 0}$. Without loss of generality, we assume that there exists a non-trivial vector $v \in \mathrm{L}(f)$ and $\lambda' \in (0, 1)$ such that:

$$f(x + \lambda' v) \neq f(x).$$

It follows that there exist disjoint polyhedra $P_1, P_2 \in \mathcal{P}(f)$ such that $x \in P_1$, and $x + \lambda' v \in P_2$. Because $\mathcal{P}(f)$ is finite there exists a polyhedron $P_3 \in \mathcal{P}(f)$ and $\lambda_0 \in \mathbb{N}$ such that for any $\lambda \in [\lambda_0, \infty)$, $x + \lambda v \in P_3$ and the following equality holds:

$$f(x + \lambda v) = f(x).$$

Thus for the scalar $\lambda' + \lambda_0$, the point $x + (\lambda' + \lambda_0)v$ belongs to P_3 and the following equality holds:

$$f(x + (\lambda' + \lambda_0)v) = f(x).$$

By definition of the lineality space $\mathrm{L}(f)$, we get

$$f(x + (\lambda' + \lambda_0)v) = f(x + \lambda' v).$$

As a result, $f(x + \lambda' v) = f(x)$, which contradicts our assumption. So, our assumption is incorrect, and if v is in $\mathrm{L}(f)$, then for any $\lambda \in \mathbb{R}_{\geq 0}$, the vector λv belongs to $\mathrm{L}(f)$, from which it follows that the lineality space $\mathrm{L}(f)$ is a linear subspace. □

The dimension of the lineality space is denoted by $\operatorname{lnum}(f) \coloneqq \dim(\mathrm{L}(f))$ and is called the lineality number of f. In a special case, when f is a PC function such that $\mathcal{P}(f)$ is a family of cones, the following lemma holds:

Lemma II.7. *Let f be a PC function such that every $P \in \mathcal{P}(f)$ is a cone. Assume, without loss of generality, that zero is the common vertex of all cones. For the lineality space $\mathrm{L}(f) \subseteq \mathbb{R}^n$ the following equality holds:*

$$\mathrm{L}(f) = \bigcap \{\operatorname{relbd}(P) \mid P \in \mathcal{P}^0(f)\}.$$

Proof. Let $X \coloneqq \bigcap \{\operatorname{relbd}(P) \mid P \in \mathcal{P}^0(f)\}$. The goal is to show that for X, the following equality holds:

$$\mathrm{L}(f) = X.$$

Firstly, we show that $X \subseteq \mathrm{L}(f)$. Let $P \in \mathcal{P}^0(f)$ be an n-dimensional cone. For every $x \in X$ and $x_P \in \operatorname{relint}(P)$, the point $x + x_P$ is in $\operatorname{relint}(P)$ as well. The PC function f is defined on $x + x_P$ and its value is equal to $f(x + x_P) = f(x_P)$. This equality holds for every polyhedron in $\mathcal{P}^0(f)$ which implies that $x \in \mathrm{L}(f)$ and $X \subseteq \mathrm{L}(f)$.

Secondly, we show that $\mathrm{L}(f) \subseteq X$. The inclusion is proven by contradiction. Let $P, P' \in \mathcal{P}^0(f)$ be distinct cones such that

$$\operatorname{relint}(P) \cap \mathrm{L}(f) \neq \emptyset.$$

For any two points $x \in \operatorname{relint}(P) \cap \mathrm{L}(f)$ and $x' \in \operatorname{relint}(P')$, there exists $\lambda \in (0, 1)$ such that the point $y = \lambda x + (1 - \lambda)x'$ belongs to $\operatorname{relint}(P)$. Thus, the PC function f is defined on y and $f(y) \neq f(x')$, which contradicts the definition of $\mathrm{L}(f)$. So, our assumption is incorrect and $\mathrm{L}(f) \subseteq \operatorname{relbd}(P)$. Furthermore, by the same set of arguments, it follows that the lineality space $\mathrm{L}(f) \subseteq \operatorname{relbd}(P)$ for every $P \in \mathcal{P}(f)$, and $\mathrm{L}(f) \subseteq X$. As a result, $X \subseteq \mathrm{L}(f)$ and $\mathrm{L}(f) \subseteq X$ that leads to the equality $\mathrm{L}(f) = X$. □

To define PC functions for a subspace $H \subseteq \mathbb{R}^n$, such that $\operatorname{codim}(H) = i$ and $i \in \{1, \dots, n\}$, we introduce the notion of flags.

Definition II.8. A flag $\mathcal{H} \coloneqq (H_1, \dots, H_k)$ of length $k \in \{1, \dots, n\}$ is a sequence of affine subspaces $H_1 \supset H_2 \supset \cdots \supset H_k$, where $\dim(H_i) = n - i$, for all $i \in \{1, \dots, k\}$.

Definition II.9. For a given flag $\mathcal{H}$ of length k and a PC function f, we define the family of polyhedra $\mathcal{P}_{\mathcal{H}}(f)$, as follows:

$$\mathcal{P}_{\mathcal{H}}(f) \coloneqq \big\{ H_k \cap P \;\big|\; P \in \mathcal{P}_{\mathcal{H}'}(f),\ H_k \cap P \neq \emptyset \big\},$$

where $\mathcal{H}' \coloneqq (H_1, \dots, H_{k-1})$ is a flag of length $k-1$. If $\mathcal{H} = (H_1)$, the family of polyhedra $\mathcal{P}_{\mathcal{H}}(f)$ is defined as follows:

$$\mathcal{P}_{\mathcal{H}}(f) \coloneqq \{ H_1 \cap P \mid P \in \mathcal{P}^*(f),\ H_1 \cap P \neq \emptyset \}.$$

Note that all polyhedra P in $\mathcal{P}_{\mathcal{H}}(f)$ are open, pairwise disjoint and at most of dimension $n-k$.

Lemma II.10. *Let f be a PC function, and let $\mathcal{H} = (H_1, \dots, H_k)$ be a flag of length $k \in \{1, \dots, n\}$. The family of polyhedra $\mathcal{P}_{\mathcal{H}}(f)$ is stratified.*

Proof. The proof is done by induction.

Base: Let $k = 1$, i.e., $\mathcal{H} = (H_1)$. Our goal is to show that $\mathcal{P}_{\mathcal{H}}(f)$ is stratified. From Definition II.9, it follows that $\mathcal{P}_{\mathcal{H}}(f)$ is finite, every polyhedron $P \in \mathcal{P}_{\mathcal{H}}(f)$ is open and $H_1 = \bigcup_{P \in \mathcal{P}_{\mathcal{H}}(f)} P$. To finish the base case, one has to show that the property (c) from Definition II.2 is also satisfied for all polyhedra in $\mathcal{P}_{\mathcal{H}}(f)$. Let $Q_1, Q_2 \in \mathcal{P}_{\mathcal{H}}(f)$ be polyhedra such that

$$\overline{Q_1} \cap Q_2 \neq \emptyset. \tag{5}$$

There exist polyhedra $P_1, P_2 \in \mathcal{P}^*(f)$ such that (5) can be rewritten as follows:

$$\overline{P_1 \cap H} \cap P_2 \cap H \neq \emptyset.$$

Because the intersection $P_1 \cap H$ is non-empty and open, it follows that $\overline{P_1 \cap H} = \overline{P_1} \cap H$. From $P_2 \cap H \subseteq \overline{P_1} \cap H$ it follows that:

$$P_2 \cap H \subseteq \overline{P_1 \cap H}.$$

Thus, the family of polyhedra $\mathcal{P}_{\mathcal{H}}(f)$ poses all four properties and is stratified.

Step: Let $k > 1$ and $\mathcal{H} = (H_1, \dots, H_k)$. Without loss of generality, we may assume that for the flag $\mathcal{H}' = (H_1, \dots, H_{k-1})$, the family of polyhedra $\mathcal{P}_{\mathcal{H}'}(f)$ is stratified. Because H_k is a hyperplane for H_{k-1}, with the same set of arguments as in the base case, one can show that the family of polyhedra $\mathcal{P}_{\mathcal{H}}(f)$ is stratified. ☐

By Lemma II.10, all polyhedra from $\mathcal{P}_{\mathcal{H}}(f)$ cover the affine subspace H_k from the flag $\mathcal{H}$. Because $\mathcal{P}_{\mathcal{H}}(f)$ is finite, it follows that there are a finite number of polyhedra P with dimension $\leq n-k-1$. Any hyperplane $H = \{x \in \mathbb{R}^n \mid a^T x = b\}$, where $a \in \mathbb{R}^n$, splits the input space $\mathbb{R}^n$ into two open half-spaces $H^+ \coloneqq \{x \in \mathbb{R}^n \mid a^T x > b\}$ and $H^- \coloneqq \{x \in \mathbb{R}^n \mid a^T x < b\}$.

Lemma II.11. *Let f be a PC function and let $H \subset \mathbb{R}^n$ be a hyperplane. Then for any generic $x \in H$ and two vectors $v_1, v_2 \in \mathbb{R}^n$ such that both $x + v_1$ and $x + v_2$ are in H^+, the following equality holds:*

$$\lim_{\varepsilon \to 0} \big(f(x + \varepsilon v_1) - f(x - \varepsilon v_1) \big) = \lim_{\varepsilon \to 0} \big(f(x + \varepsilon v_2) - f(x - \varepsilon v_2) \big).$$

Proof. By Lemma II.10, the hyperplane H is covered by a finite family $\mathcal{P}_{\mathcal{H}}(f)$ of $(n-1)$-dimensional polyhedra, where $\mathcal{H} = (H)$. For every generic point $x \in H$, there exists a unique polyhedron $Q \in \mathcal{P}_{\mathcal{H}}(f)$, that contains the point x. We distinguish two cases.

Case 1: If $Q \subset P$ for some n-dimensional polyhedron $P \in \mathcal{P}(f)$, then there exists $\varepsilon > 0$ such that $\{x + \varepsilon v \mid v \in \mathbb{R}^n\} \subset P$ and the function f is constant on this ε-neighborhood of x. Therefore, for any two vectors $v_1, v_2 \in \mathbb{R}^n$ we have

$$\lim_{\varepsilon \to 0} f(x + \varepsilon v_1) = \lim_{\varepsilon \to 0} f(x + \varepsilon v_2). \tag{6}$$

Case 2: Otherwise, there exist $P_1, P_2 \in \mathcal{P}(f)$ such that $Q = \operatorname{relint}(\overline{P_1} \cap \overline{P_2})$. Without loss of generality, we may assume that $P_1 \cap H^+ = P_1$. Let $v_1, v_2 \in \mathbb{R}^n$ be two vectors such that $x + v_1, x + v_2 \in H^+$. Then there exists $\varepsilon > 0$ for which $x + \varepsilon v_1, x + \varepsilon v_2 \in P_1$. Since f is constant on P_1, it follows that (6) holds in this case as well.

With the same set of arguments, one can show that the following equality holds:

$$\lim_{\varepsilon \to 0} f(x - \varepsilon v_1) = \lim_{\varepsilon \to 0} f(x - \varepsilon v_2). \tag{7}$$

By combining (6) and (7), we receive the desired equality. ☐

For a flag $\mathcal{H} = (H_1, \dots, H_k)$ we say that a collection of vectors $\mathcal{V} \coloneqq (v_1, \dots, v_k)$ is a directional with respect to $\mathcal{H}$, if every vector v_i, where $i \in \{1, \dots, k\}$ satisfies two properties: v_i is non-parallel to H_i and for every point $x \in H_i$, the sum $v_i + x$ belongs to H_{i-1}.

Definition II.12. Let f be a PC function, let $\mathcal{H} = (H_1, \ldots, H_k)$ be a flag of length $k \in \{1, \ldots, n\}$, and let $\mathcal{V} = (v_1, \ldots, v_k)$ be a directional with respect to $\mathcal{H}$. For any generic point $x \in H_k$, we define the delta function $\Delta_f(\mathcal{H}, \mathcal{V})\colon H_k \to G$ recursively as follows:

$$\Delta_f(\mathcal{H}, \mathcal{V})(x) := \lim_{\varepsilon \to 0} \big(\Delta_f(\mathcal{H}', \mathcal{V}')(x + \varepsilon v_k) - \Delta_f(\mathcal{H}', \mathcal{V}')(x - \varepsilon v_k)\big),$$

where $\mathcal{H}' := (H_1, \ldots, H_{k-1})$ and $\mathcal{V}' := (v_1, \ldots, v_{k-1})$.

When $\mathcal{H} = (H_1)$ and $\mathcal{V} = (v_1)$, the delta function $\Delta_f(\mathcal{H}, \mathcal{V})\colon H_1 \to G$ is defined as follows:

$$\Delta_f(\mathcal{H}, \mathcal{V})(x) := \lim_{\varepsilon \to 0} \big(f(x + \varepsilon v_1) - f(x - \varepsilon v_1)\big).$$

The delta function $\Delta_f(\mathcal{H}, \mathcal{V})$ is a PC function with respect to the pair $(\mathcal{H}, \mathcal{V})$. From Lemma II.11, it follows that the delta function is invariant under the choice of the directional $\mathcal{V}$ – it just depends on the orientation of each hyperplane, i.e., a the specification of the two components $H_{i-1} \setminus H_i$ as positive or negative. This means that $\mathcal{V}$ can be omitted in the definition of the delta function, and in the rest of the paper, we, therefore, write $\Delta_f(\mathcal{H})$ instead of $\Delta_f(\mathcal{H}, \mathcal{V})$. By definition, the flag $\mathcal{H}$ should contain at least one affine subspace of codimension 1. For the sake of generality, we say that the flag $\mathcal{H}$ is of length 0, if it contains the input space $\mathbb{R}^n$, i.e., $\mathcal{H} = (\mathbb{R}^n)$. The delta function $\Delta_f(\mathcal{H})$ for the flag $()$ of length zero is the PC function f, i.e., $\Delta_f(()) := f$.

Lemma II.13. *Let f be a PC function. If $\operatorname{supp}(f)$ is non-empty and does not contain any affine space of dimension k, then there exists a hyperplane H for which the support of $\Delta_f(\mathcal{H})$, where $\mathcal{H} = (H)$ is non-empty and does not contain any affine space of dimension k.*

Proof. This holds because $\operatorname{supp}(\Delta_f(\mathcal{H})) \subseteq \operatorname{supp}(f)$. □

Lemma II.14. *Let f be a PC function, and let $\mathcal{H}$ be a flag of length $k \in \{0, \ldots, n-1\}$ such that $\Delta_f(\mathcal{H})$ is non-constant. Then, for any affine subspace $H_{k+1} \subset H_k$ of codimension $k+1$, where $H_k \in \mathcal{H}$, the inequality holds:*

$$\operatorname{lnum}(\Delta_f(\mathcal{H}) \cup (H_{k+1})) \geq \operatorname{lnum}(\Delta_f(\mathcal{H})). \tag{8}$$

Proof. Let $\mathcal{H}$ be a flag of fixed length k, where $k \in \{0, \ldots, n-1\}$ such that $\Delta_f(\mathcal{H})$ is non-constant. Without loss of generality, we may assume that $0 \in H_k$, where $H_k \in \mathcal{H}$. The affine subspace H_k is isomorphic to $\mathbb{R}^{n-k}$. Let $\phi\colon H_k \to \mathbb{R}^{n-k}$ be a bijection between H_k and R^{n-k} and let $H_{k+1} \subset H_k$ be an affine subspace of codimension $k+1$. The mapping ϕ translates H_{k+1} into a set $H \subset \mathbb{R}^{n-k}$ as follows: $H := \{\phi^{-1}(x) \mid x \in H_{k+1}\}$. The set H is a hyperplane in $\mathbb{R}^{n-k}$. Also, the mapping ϕ defines the PC function $g\colon \mathbb{R}^{n-k} \to G$ as follows:

$$g := \Delta_f(\mathcal{H}) \circ \phi^{-1}.$$

The functions g and $\Delta_f(\mathcal{H})$ have the lineality spaces of the same dimension. Moreover, the mapping ϕ establishes a one-to-one correspondence between $\Delta_f(\mathcal{H} \cup (H_{k+1}))$ and $\Delta_g((H))$ as follows:

$$\Delta_f(\mathcal{H} \cup (H_{k+1})) = \Delta_g((H)).$$

As a result, the functions $\Delta_f(\mathcal{H} \cup (H_{k+1}))$ and $\Delta_g((H))$ have the lineality spaces of the same dimension as well. To finish the proof, we have to show that for a hyperplane $H \in \mathbb{R}^{n-k}$, the following inequality holds:

$$\operatorname{lnum}(\Delta_g((H))) \geq \operatorname{lnum}(g).$$

We distinguish two cases.

Case 1: Let $\mathrm{L}(g)$ be parallel to H. There exists non-trivial $y \in \mathrm{L}(g)$ that is non-parallel to H and is picked as the direction vector for $\Delta_g((H))$. Then for every $x \in \operatorname{def}(\Delta_g((H)))$ holds:

$$\Delta_g((H))(x) = \lim_{\varepsilon \to 0} g(x + \varepsilon y) - g(x - \varepsilon y) = \lim_{\varepsilon \to 0} g(x) - g(x) = 0.$$

So, the function $\Delta_g((H)) \equiv 0$ with $\operatorname{lnum}(\Delta_g((H))) = n - k - 1$, and Eq. (8) holds.

Case 2: Let $\mathrm{L}(g)$ be parallel to H. For every $y \in \mathrm{L}(g)$ and $x \in \operatorname{def}(\Delta_g((H)))$, the point $x + y$ belongs to H and $\Delta_g((H))$ for $x + y$ has the following form:

$$\Delta_g((H))(x + y) = \lim_{\varepsilon \to 0} g(x + y + \varepsilon v) - g(x + y - \varepsilon v). \tag{9}$$

Because $y \in \mathrm{L}(g)$, the summands in (9) can be transformed as follows:

$$g(x + y + \varepsilon v) = g(x + \varepsilon v),$$
$$g(x + y - \varepsilon v) = g(x - \varepsilon v).$$

By substituting the summands in (9), one receives the equality:

$$\Delta_g((H))(x + y) = \Delta_g((H))(x).$$

So, $y \in \mathrm{L}(\Delta_g((H)))$ from which follows that $\mathrm{L}(g) \subseteq \mathrm{L}(\Delta_g((H)))$ and (8) holds. □

Lemma II.15. *Let $f\colon \mathbb{R}^n \to G$ be a PC function such that:*

$$f = \sigma_1 f_1 + \cdots + \sigma_p f_p,$$

where f_i is a PC function, for which $\mathrm{L}(f_i)$ contains a k-affine subspace and $\sigma_i \in \{-1, 1\}$ for all $i \in \{1, \dots, p\}$. Then, $\mathrm{supp}(f)$ is either the empty set or contains an affine space of dimension k.

Proof. The proof is done by induction.

Base: Let $n = k$. Then for every $i \in \{1, \dots, p\}$, $L(f_i)$ contains an n-dimensional affine subspace. It is possible only when f_i is a constant function on $\mathrm{def}(f_i)$. So, a linear combination of constant functions is also a constant function, and $\mathrm{L}(f)$ contains the n-affine subspace.

Step: Assume that $n > k$ and $\mathrm{supp}(f) \neq \emptyset$ does not contain an affine k-space. By Lemma II.13, there is a hyperplane H such that the support of $\Delta_f(\mathcal{H})$ is non-empty and does not contain an affine subspace of dimension k. By Lemma II.14, for every $i \in \{1, \dots, p\}$, if f_i is not constant, then the following inequality holds:

$$\mathrm{lnum}(\Delta_{f_i}(\mathcal{H})) \geq k.$$

Note, that for $f = \sigma_1 f_1 + \cdots + \sigma_q f_q + \sigma_{q+1} f_{q+1} + \cdots + \sigma_p f_p$, where $f_1, \dots, f_q$ are constant and $f_{q+1}, \dots, f_p$ are not, the function $\Delta_f(\mathcal{H})$ is represented as follows:

$$\Delta_f(\mathcal{H}) = \sigma_{q+1}\Delta_{f_{q+1}}(\mathcal{H}) + \cdots + \sigma_p \Delta_{f_p}(\mathcal{H}),$$

where the lineality space $\mathrm{L}(\Delta_{f_i}(\mathcal{H}))$ contains a k-linear space, for all $i \in \{q{+}1, \dots, p\}$. By induction hypothesis, $\mathrm{supp}(\Delta_f(\mathcal{H}))$ contains a k-affine subspace. But it contradicts our assumption, so our assumption is incorrect, and $\mathrm{supp}(f)$ contains a k-affine subspace. □

Lemma II.16. *Let $\mathcal{H}'$ be a flag of length k, where $k \in \{0, \dots, n-1\}$, and let $f\colon \mathbb{R}^n \to G$ be a PC function such that for all flags $\mathcal{H} = \mathcal{H}' \cup (H_{k+1})$ of length $k+1$, the delta function $\Delta_f(\mathcal{H})$ is equal to 0, i.e., $\Delta_f(\mathcal{H}) \equiv 0$. Then the delta function $\Delta_f(\mathcal{H}')$ is constant.*

Proof. Let $\mathcal{H}'$ be a fixed flag of length k, where $k \in \{0, \dots, n-1\}$, and let f be a PC function such that for any flag $\mathcal{H} = \mathcal{H}' \cup (H_{k+1})$ of length $k+1$, the delta function $\Delta_f(\mathcal{H}) \equiv 0$. We aim to show that the delta function $\Delta_f(\mathcal{H}')$ is constant.

Without loss of generality, we may assume that $0 \in H_k$. Thus, H_k is a linear subspace of codimension k and is isomorphic to $\mathbb{R}^{n-k}$. Let $\phi\colon H_k \to \mathbb{R}^{n-k}$ be a bijection between H_k and $\mathbb{R}^{n-k}$ and let a function $g\colon \mathbb{R}^{n-k} \to G$ be defined as follows:

$$g \coloneqq \Delta_f(\mathcal{H}') \circ \phi^{-1}.$$

The function g is a PC function on $\mathbb{R}^{n-k}$. A point $x \in \mathbb{R}^{n-k}$ is generic if the point $\phi^{-1}(x)$ is generic in H_k with respect to $\Delta_f(\mathcal{H}')$. Let $H \subset \mathbb{R}^{n-k}$ be a hyperplane. By the mapping ϕ, the hyperplane H translates to a set of points $H_{k+1} = \{\phi^{-1}(x) \mid x \in H\}$. The set H_{k+1} is an affine subspace of H_k of codimension $k+1$. Additionally, for the flag (H), the function g defines the delta function $\Delta_g((H))$ such that:

$$\Delta_g((H)) = \Delta_f(\mathcal{H}) \equiv 0,$$

where $\mathcal{H} = \mathcal{H}' \cup (H_{k+1})$.

Therefore, g is the PC function such that for any flag (H) of length 1, the delta function $\Delta_g((H))$ is equal to 0. It is possible only in a case when g is constant. So, $\Delta_f(\mathcal{H}')$ is constant as well.

□

Lemma II.17. *Let f be a PC function. Then the delta function $\Delta_f(\mathcal{H})$ equals 0 for all but a finite number of flags $\mathcal{H}$.*

Proof. Let $\mathcal{F}$ be the family of all flags such that for any flag $\mathcal{H} \in \mathcal{F}$, the delta function $\Delta_f(\mathcal{H})$ is non-zero. Our goal is to show that $|\mathcal{F}| < \infty$. For any k, let $\mathcal{F}^k$ be the family of flags of length k, such that for every $\mathcal{H} \in \mathcal{F}^k$, $\Delta_f(\mathcal{H})$ is non-zero. The family of all flags $\mathcal{F}$ is a union of families of flags of fixed length, i.e., the following equality holds:

$$\mathcal{F} = \bigcup_{k=1}^{n} \mathcal{F}^k. \tag{10}$$

Any two families in the union in (10) are pairwise disjoint, from which follows that the equality holds between the sizes of the families of flags:

$$|\mathcal{F}| = \sum_{k=1}^{n} |\mathcal{F}^k|.$$

Let $k = 1$ and let $\mathcal{H} = (H_1)$ such that $\mathcal{H} \in \mathcal{F}^1$. The delta function $\Delta_f(\mathcal{H})$ is non-zero, i.e, there exists $x \in \mathbb{R}^n$ such that $\Delta_f(\mathcal{H})(x) \neq 0$. As, $\Delta_f(\mathcal{H})$ is a PC function on H_1 and is non-zero on the point x it follows that there exists $P \in \mathcal{P}^1(f)$ such that $x \in P$, $P \subseteq H_1$ and for every $y \in P$ holds:

$$\Delta_f(\mathcal{H})(y) = \Delta_f(\mathcal{H})(x) \neq 0.$$

The polyhedron P has dimension $n-1$. It implies the existence of a unique hyperplane H for which $P \subseteq H$ and the hyperplane H coincide with H_1. As a result, for any flag $\mathcal{H} \in \mathcal{F}^1$ there exists at least one polyhedron $P \in \mathcal{P}^1(f)$ such that $P \subseteq H_1$, from which the following inequality holds:

$$|\mathcal{F}^1| \leq |\mathcal{P}^1(f)|.$$

Let $k \geq 2$ and let $\mathcal{H}' = (H_1, \ldots, H_{k-1})$ be a flag of length $k-1$. If $\Delta_f(\mathcal{H}')$ is constant, by Lemma II.16 it follows that for every flag $\mathcal{H} = (H_1, \ldots, H_k)$, where $H_k \subset H_{k-1}$, holds $\Delta_f(\mathcal{H}) \equiv 0$. Thus, if for the flag $\mathcal{H}$ of length k the delta function $\Delta_f(\mathcal{H})$ is non-zero, it follows that the delta function $\Delta_f(\mathcal{H}')$ for the flag $\mathcal{H}'$ is non-constant. To finish this proof, one has to show that if $\mathcal{H}' \in \mathcal{F}^{k-1}$, then it has finitely many affine subspaces H_k of codimension k such that the function $\Delta_f((H_1, \ldots, H_k))$ is non-zero.

Let $\mathcal{H}' \in \mathcal{F}^{k-1}$ be a flag for which the function $\Delta_f(\mathcal{H}')$ is non-constant. Without loss of generality, we may assume that the affine subspace H_{k-1} contains 0. The affine subspace H_{k-1} is isomorphic to a linear space $\mathbb{R}^{n-k+1}$. Let $\phi\colon H_{k-1} \to \mathbb{R}^{n-k+1}$ be a bijection between H_{k-1} and $\mathbb{R}^{n-k+1}$. The composition of ϕ and $\Delta_f(\mathcal{H}')$ defines a function $g\colon \mathbb{R}^{n-k+1} \to G$ as follows:

$$g \coloneqq \Delta_f(\mathcal{H}') \circ \phi^{-1}.$$

The function g is PC and splits the linear space $\mathbb{R}^{n-k+1}$ into a finite family of polyhedra $\mathcal{P}^{n-k+1}(g)$ of codimension $k-1$. The function ϕ transforms any affine subspace $H_k \subset H_{k-1}$ of codimension k to a set of points $\phi(H_k) \subset \mathbb{R}^{n-k+1}$ as follows:

$$\phi(H_k) \coloneqq \{\phi(x) \mid x \in H_k\}.$$

The set of points $\phi(H_k)$ is a hyperplane in $\mathbb{R}^{n-k+1}$. Moreover, the function $\Delta_f((H_1, \ldots, H_k))$ is non-zero if and only if $\Delta_g((\phi(H_k)))$ is non-zero as well. By the same motivation as for the case $k = 1$, there are finitely many hyperplanes $H \subset \mathbb{R}^{n-k+1}$ for which $\Delta_g((H))$ is non-zero. Thus, there are finitely many affine subspaces H_k of codimension k for which $\Delta_f((H_1, \ldots, H_k))$ is non-zero as well. This holds for any flag in $\mathcal{F}^{k-1}$, and it follows that $|\mathcal{F}^k| < \infty$. Therefore, for any $k \in \{1, \ldots, n\}$, $|\mathcal{F}^k|$ is finite which implies that $|\mathcal{F}|$ is finite as well.

□

III. Minimal decomposition

We are interested in the case when $G = \mathbb{R}^n$, for $n \in \mathbb{N}$. Let $F\colon \mathbb{R}^n \to \mathbb{R}$ be a CPWL function. The PC function f for a given CPWL function F is defined as follows:

$$f \coloneqq \nabla F.$$

We are also interested in the connection between the possible decomposition of the function F as a linear combination of $\max$ functions and the properties of the corresponding PC functions.

Lemma III.1. *Let $F \coloneqq \max(g_1, \ldots, g_k)$, where $k \leq n+1$, be a CPWL function such that $g_i\colon \mathbb{R}^n \to \mathbb{R}$ be an affine-linear function for $i \in \{1, \ldots, k\}$. Then, the lineality space for the function $f \coloneqq \nabla F$ has at most codimension $k-1$, i.e.,* $\operatorname{codim}(\mathrm{L}(f)) \leq k-1$.

Proof. Since the following equality holds:

$$F = \max(g_1, \ldots, g_k) = \max(0, g_2 - g_1, \ldots, g_k - g_1) + g_1,$$

we also have the correspondent equality for the derivatives:

$$f = \nabla \max(0, \overline{g}_1, \ldots, \overline{g}_{k-1}) + c,$$

where $c \coloneqq \nabla(g_1)$, $\overline{g_i} \coloneqq g_{i+1} - g_1$, for all $i \in \{2, \ldots, k-1\}$. Thus, the equality holds for the correspondent lineality spaces:

$$\mathrm{lnum}(f) = \mathrm{lnum}(\nabla \max(0, \overline{g}_1, \ldots, \overline{g}_{k-1})). \tag{11}$$

Let $A \in \mathbb{R}^{(k-1)\times n}$ be a matrix, whose rows are $\nabla\overline{g}_1, \ldots, \nabla\overline{g}_{k-1}$. The kernel $\ker(A)$ of the matrix A is a subset of $\mathrm{L}(\nabla \max(0, \overline{g}_1, \ldots, \overline{g}_{k-1}))$. For any $x \in \mathbb{R}^n$ and $y \in \ker(A)$, the following equality holds:

$$\max(0, \overline{g}_1, \ldots \overline{g}_{k-1})(x + y) = \max(0, \overline{g}_1, \ldots, \overline{g}_{k-1})(x).$$

Since, the matrix A has $(k-1)$-rows, it follows that $\dim(\ker(A)) \geq n-k+1$. By the eqaulity $\operatorname{codim}(\ker(A)) = n - \dim(\ker(A))$ and equality in (11), it follows that $\operatorname{codim}(\mathrm{L}(f)) \leq k-1$. □

Example III.2. Let us to consider two functions $G_3(x, y) := \max(0, x, y) + \max(0, -x, -y)$ and $G_4 := 6\max(0, x, y) + \max(0, -x, -y)$. These functions are defined on $\mathbb{R}^2$, see Fig. 2. The functions G_3 and G_4 split the input space into the same family of polyhedra. In Example I.1, we have shown that G_3 can be represented as a linear combination of $\max$ with at most two arguments. One may think that since G_3 and G_4 tessellate the input space identically, it implies the function G_4 can be represented in the form of Eq. (1) with at most two arguments as well. However, the function G_4 cannot be represented as a linear combination of $\max$ with at most two arguments. To show this, let us assume the opposite, that for a finite set of indices I, the following equality holds:

$$G_4 = \sum_{i \in I} \max(F_{i,1}, F_{i,2}).$$

Because the function $G_4 = G_3 + 5\max(0, x, y)$, we get that for $\max(0, x, y)$ holds:

$$\max(0, x, y) = \frac{1}{5}(G_4 - G_3),$$

where $G_4 - G_3$ is a linear combination of $\max$ with at most two arguments. This leads to the contradiction because the function $\max(0, x, y)$ cannot be represented in this form. So, the function G_4 cannot be represented as a linear combination of $\max$ with at most two arguments. Although the functions G_3 and G_4 split the input space into the same family of polyhedra, their tessellations are still different with respect to their delta functions. To show this, we define two PC functions $g_3 := \nabla G_3$ and $g_4 := \nabla G_4$. For any flag $\mathcal{H}$ of length 1, the delta function $\Delta_{g_3}(\mathcal{H})$ is constant. At the same time, for the hyperplane $H_1 := \{(x, y) \in \mathbb{R}^2 \mid x = y\}$, the delta function $\Delta_{g_4}((H_1))$ is non-constant.

Eventually, the fact that in Example III.2 for any flag $\mathcal{H}$ of length 1, the delta function $\Delta_{g_3}(\mathcal{H})$ is constant has a direct implication on the existence of the decomposition of the function G_3 as a linear combination of $\max$ with at most two arguments. The observations made in Example III.2 lead to the following theorem:

Theorem III.3. *Let $F\colon \mathbb{R}^n \to \mathbb{R}$ be a CPWL, let $f := \nabla F$. The function F can be represented as a linear combination of $\max$ with at most $k \in \mathbb{N}$ arguments if and only if for any flag $\mathcal{H} = (H_1, \dots, H_{k-1})$ of length $k-1$ holds that the delta function $\Delta_f(\mathcal{H})$ is constant.*

Proof. By [9], it was shown that the function F can be represented as a linear combination of $\max$ with at most $n+1$ arguments. Thus, without loss of generality, we may assume that $k \le n+1$.

Firstly, we show ($\Rightarrow$) implication. Let F be a CPWL function that can be represented as a linear combination of $\max$ of at most k arguments, i.e., there exists a finite set of indices $I \subset \mathbb{N}$ for which the following equality holds:

$$F = \sum_{i \in I} \sigma_i F_i, \tag{12}$$

where $\sigma_i \in \{-1, 1\}$, $F_i := \max(g_{i,1}, \dots, g_{i,p_i})$, such that $p_i \le k$ and $g_{i,j}\colon \mathbb{R}^n \to \mathbb{R}$ is an affine-linear function $j \in \{1, \dots, p_i\}$, for all $i \in I$. For every $i \in I$, we denote $f_i := \nabla F_i$. The function f is a PC function and by (12), it is decomposed as follows:

$$f = \sum_{i \in I} \sigma_i f_i.$$

Let $\mathcal{F}_f$ be the family of all flags for which the function $\Delta_f(\mathcal{H})$ is not equal to 0. By Lemma II.17, the family of flags $\mathcal{F}_f$ is finite. Our goal is to show that for every flag $\mathcal{A} \in \mathcal{F}_f$, the function $\Delta_f(\mathcal{A})$ is constant.

The proof is done by contradiction. Assume there exists a flag $\mathcal{H} \in \mathcal{F}_f$ of length $k-1$ such that the function $\Delta_f(\mathcal{H})$ is non-constant. The function $\Delta_f(\mathcal{H})$ is decomposed as follows:

$$\Delta_f(\mathcal{H}) = \sum_{i \in I} \sigma_i \Delta_{f_i}(\mathcal{H}), \tag{13}$$

where $\sigma_i \in \{-1, 1\}$, for all $i \in I$. Due to the assumption that $\Delta_f(\mathcal{H})$ is non-constant, in Eq. (13) exists at least one f_i for which $\Delta_{f_i}(\mathcal{H})$ is non-constant. It implies that there exists a flag $\mathcal{H}' \in \mathcal{F}_{f_i}$ of length $k-2$ such that the delta function $\Delta_{f_i}(\mathcal{H}')$ is non-constant as well. By Lemma III.1, the following chain of inequalities holds:

$$n - k + 1 \le \operatorname{lnum}(f_i) \le \operatorname{lnum}(\Delta_{f_i}(\mathcal{H}')) \le \operatorname{lnum}(\Delta_{f_i}(\mathcal{H})).$$

Since, the function $\Delta_{f_i}(\mathcal{H})$ is defined on an affine subspace of dimension $n-k+1$, it implies that the dimension of $\mathrm{L}(\Delta_{f_i}(\mathcal{H}))$ is $n-k+1$. It is possible when the function $\Delta_{f_i}(\mathcal{H})$ is constant, which contradicts our assumption. So, our assumption is incorrect and $\Delta_f(\mathcal{H})$ is constant on every flag $\mathcal{A} \in \mathcal{F}_f$.

Secondly, we show ($\Leftarrow$) implication. Assume that the PC function $f = \nabla F$ has the property that for all flags $\mathcal{H}$ of fixed length k, the delta function $\Delta_f(\mathcal{H})$ is constant. By Lemma II.17 there exists a finite family of flags $\mathcal{F}_f = \{\mathcal{H}_1, \dots, \mathcal{H}_r\}$ for which the function $\Delta_f(\mathcal{H})$ is non-zero. Our goal is to construct a CPWL function $G_1\colon \mathbb{R}^n \to \mathbb{R}$ with $g_1 := \nabla G_1$ for which the following properties hold:

(a) The function G_1 is a linear combination of maxima of at most $k+1$ affine-linear functions.
(b) The function $\Delta_{g_1}(\mathcal{H})$ is zero for all flags $\mathcal{H}$ that do not belong to $\mathcal{F}_f$.
(c) There exists at least one flag $\mathcal{H} \in \mathcal{F}_f$ for which $\Delta_{g_1}(\mathcal{H}) = \Delta_f(\mathcal{H})$ holds.
(d) The lineality space of g_1 has codimension k and contains the vector space parallel to H_k, where H_k is the last element in $\mathcal{H}$.

Before proving the existence of a function G_1 with the declared properties, let us assume that such a function exists. We define a CPWL function $F_1 \colon \mathbb{R}^n \to \mathbb{R}$ as follows:

$$F_1 \coloneqq F - G_1.$$

The PC function $f_1 \coloneqq \nabla F_1$ satisfies the following equality:

$$f_1 = f - g_1.$$

Thus, for any flag $\mathcal{A}$ of length $s \in \{1, \dots, n\}$, the delta function $\Delta_{f_1}(\mathcal{A})$ satisfies the following equality:

$$\Delta_{f_1}(\mathcal{A}) = \Delta_f(\mathcal{A}) - \Delta_{g_1}(\mathcal{A}). \tag{14}$$

Property (b) and Eq. (14) now imply that for any flag $\mathcal{A} \notin \mathcal{F}_f$, the delta function $\Delta_{f_1}(\mathcal{A})$ is zero. Thus, it follows that $\mathcal{F}_f \subseteq \mathcal{F}_{f_1}$. By property (c) of G_1, there exists a flag $\mathcal{H} \in \mathcal{F}_f$ such that:

$$\Delta_{f_1}(\mathcal{H}) = \Delta_f(\mathcal{H}) - \Delta_{g_1}(\mathcal{H}) \equiv 0. \tag{15}$$

By Eq. (15), the flag $\mathcal{H}$ does not belong to $\mathcal{F}_{f_1}$, and therefore $\mathcal{F}_{f_1} \subset \mathcal{F}_f$. For the function F_1, there exists a CPWL function $G_2 \colon \mathbb{R}^n \to \mathbb{R}$ with the properties (a)–(d) as well. By repeating this procedure a finite number of times $p \leq |\mathcal{F}_f|$, one generates the function $F^* \colon \mathbb{R}^n \to \mathbb{R}$ as follows:

$$F^* = F - \sum_{i=1}^{p} G_i, \tag{16}$$

where for all $i \in \{1, \dots, p\}$, the function G_i is a linear combination of $\max$ of at most $k+1$ arguments. For the PC function ∇F^*, the family of flags $\mathcal{F}_{\nabla F^*}$ is empty, which implies that the function ∇F^* is constant. This is only possible when the function F^* is affine-linear. By Eq. (16), the function F is the sum of $\max$ of at most $k+1$ arguments.

To finish the proof, one is left to show that the function G_1 exists. The function F is a CPWL function such that f is a PC function with $\operatorname{codim}(\mathrm{L}(f)) = k$. By Lemma II.14 for any flag $\mathcal{A}$ of length k, the delta function $\Delta_f(\mathcal{A})$ is constant. Let $\mathcal{H} = (H_1, \dots, H_k)$ such that $\mathcal{H} \in \mathcal{F}_f$, and let $x_0 \in H_k$ such that $\Delta_f(\mathcal{H})(x_0) \neq 0$. Let $U \subset \mathbb{R}^n$ be an n-dimensional ball with the center in x_0 with the following property:

$$U \subset \bigcup \{\overline{P} \mid P \in \mathcal{P}^0(F), x_0 \in \overline{P}\}.$$

For the sake of simplicity and without loss of generality, we assume that $x_0 = 0$ and $F(0) = 0$. In two steps, we define the function $G \colon \mathbb{R}^n \to \mathbb{R}$.

Step 1: For every point $x \in U$, we define $G(x) \coloneqq F(x)$. Because $F(0) = 0$, the function F, restricted to the ball U, is positively homogeneous.

Step 2: For any point $x \in \mathbb{R}^n$, there exists $\lambda \geq 0$ and $y \in U$, such that $x = \lambda y$. In this case, we define $G(x) \coloneqq \lambda G(y)$.

The function G is a CPWL function. It defines a PC function $g = \nabla G$. From the definition of G, it follows that the linear regions of G are cones. These cones generate a stratified family of polyhedra $\mathcal{P}(g)$. The rest of the proof shows that the functions G and g satisfy properties (a)–(d). Let $\mathcal{H} = (H_1, \dots, H_k)$ be the flag used to define G.

Property (b): The family of flags $\mathcal{F}_g$ is a subset of the family flags $\mathcal{F}_f$ of the PC function f. Let $i \in \{1, \dots, k\}$ and let $\mathcal{A} = (A_1, \dots, A_i)$ be a flag such that $\mathcal{A} \in \mathcal{F}_g$. For the flag $\mathcal{A}$ the delta function $\Delta_g(\mathcal{A})$ is non-zero, from which follows that A_i contains a face $P \in \mathcal{P}(g)$ of codimension i. The face P is a cone, so the closure of P contains 0 and the intersection with the ball U is non-empty. Because $P \cap U \neq \emptyset$ and g coincides with f in U, it follows that $\Delta_f(\mathcal{A})(x) = \Delta_g(\mathcal{A})(x) \neq 0$ for every point $x \in P \cap U$. As a result, $\mathcal{A} \in \mathcal{F}_f$ and $\mathcal{F}_g \subseteq \mathcal{F}_f$.

Property (c): Because $\mathcal{H} \in \mathcal{F}_f$ it follows that $\Delta_f(\mathcal{H})$ is constant and $0 \in \operatorname{supp}(\Delta_f(\mathcal{H}))$. Since f and g coincide in U, the equality holds:

$$\Delta_g(\mathcal{H})(0) = \Delta_f(\mathcal{H})(0) \neq 0. \tag{17}$$

Eq. (17) implies $\mathcal{H} \in \mathcal{F}_g$. The family of flags $\mathcal{F}_g$ contains flags of length at most k, from which follows that for any flag $\mathcal{H}' = (H_1, \dots, H_k, H_{k+1})$ holds $\Delta_g(\mathcal{H}') = 0$. By Lemma II.16, this is only possible when $\Delta_g(\mathcal{H})$ is constant. Thus, the delta function $\Delta_g(\mathcal{H})$ is constant and by Eq. (17) implies $\Delta_g(\mathcal{H}) = \Delta_f(\mathcal{H})$.

Property (d): By property (b) the flag $\mathcal{H}$ belongs to $\mathcal{F}_g$. This is possible only when the delta function $\Delta_g((H_1, \dots, H_{k-1}))$ is non-constant. Thus, by Lemma II.14, the codimension of the lineality space $\mathrm{L}(g)$ satisfies the following inequality:

$$\operatorname{codim}(\mathrm{L}(g)) \geq k.$$

Let us assume that $\operatorname{codim}(\mathrm{L}(g)) \geq k+1$, which is only possible when there exists a flag $\mathcal{A}'$ of length k such that the function $\Delta_g(\mathcal{A}')$ is non-constant. By property (b) for any flag $\mathcal{A}$ of length $\geq k+1$ holds:

$$\Delta_g(\mathcal{A}) \equiv 0. \tag{18}$$

Eq. (18) and Lemma II.16 imply that the delta function $\Delta_g(\mathcal{A}')$ is constant. This is a contradiction. So, our assumption is incorrect, and the codimension of $\mathrm{L}(g)$ is k.

Property (a): We have to show that the function G can be represented as a linear combination of $\max$ with at most $k+1$ arguments. Let $W \subseteq \mathbb{R}^n$ be a linear subspace of dimension k, that is complement to $\mathrm{L}(g)$. Let us fix a point $x \in \mathbb{R}^n$. For the point x, there exists a unique pair of points $x_\ell \in \mathrm{L}(g)$ and $x_w \in W$ such that $x = x_\ell + x_w$. There exist two linear maps $\mu_\ell \colon \mathbb{R}^n \to \mathrm{L}(g)$ and $\mu_w \colon \mathbb{R}^n \to W$ that map the point x to its x_ℓ and x_w parts, respectively.

Let $P \in \mathcal{P}^0(g)$ be a cone for which $x \in \overline{P}$ and let $G^{(P)} \colon \mathbb{R}^n \to \mathbb{R}$ be the linear function that defines G on the cone P. Then, the value of G in x is computed as follows:

$$G(x) = G^{(P)}(x) = G^{(P)}(x_\ell + x_w) = (G^{(P)} \circ \mu_\ell)(x) + (G^{(P)} \circ \mu_w)(x). \tag{19}$$

Eq. (19) suggests that the function G can be decomposed as a sum of two functions: $G_\ell \colon \mathbb{R}^n \to \mathbb{R}$ and $G_w \colon \mathbb{R}^n \to \mathbb{R}$. We define the function G_ℓ for the point x as follows:

$$G_\ell(x) \coloneqq (G^{(P)} \circ \mu_\ell)(x).$$

By Lemma II.7, the lineality space $\mathrm{L}(g)$ belongs to the boundary of every cone $P' \in \mathcal{P}^0(g)$, which implies that $G_\ell = (G^{(P')} \circ \mu_\ell)$. Therefore, G_ℓ is not only CPWL but also a linear function.

We define the function G_w for a point $x \in \overline{P}$ as follows:

$$G_w(x) \coloneqq (G^{(P)} \circ \mu_w)(x).$$

Because the set of points $\{x + y \mid y \in \mathrm{L}(g)\}$ is a subset of $\overline{P}$ it follows that x_w belongs to $\overline{P}$ as well, which implies the following equalities:

$$G_w(x) = (G^{(P)} \circ \mu_w)(x) = G^{(P)}(x_w) = G(x_w).$$

Therefore, we can restrict the function G_w on the linear subspace W without losing any information. The restricted G_w on W function is denoted as $G_w|_W$. The linear subspace W is isomorphic to the k-dimensional vector space $\mathbb{R}^k$. We denote the isomorphism between W and $\mathbb{R}^k$ by $\phi \colon W \to \mathbb{R}^k$.

By using ϕ, we define a function $T \colon \mathbb{R}^k \to \mathbb{R}$ as follows:

$$T \coloneqq G_w|_W \circ \phi^{-1}.$$

The function T is a CPWL function defined on $\mathbb{R}^k$. By [9], T can be represented as a linear combination of $\max$ with at most $k+1$ arguments i.e., the following equality holds:

$$T = \sum_{i \in I} \sigma_i \max(F_{i,1}, \ldots, F_{i,s}),$$

where $s \leq k+1$. By composing the isomorphism ϕ and the function T the following equality holds:

$$G_w|_W = \sum_{i \in I} \sigma_i \max(F_{i,1}, \ldots, F_{i,s}) \circ \phi.$$

Moreover, by composing $G_w|_W$ and μ_w, the function $G_w|_W$ is extended to $\mathbb{R}^n$ such that:

$$\begin{aligned} G_w &= \sum_{i \in I} \sigma_i \max(F_{i,1} \circ \phi \circ \mu_w, \ldots, F_{i,s} \circ \phi \circ \mu_w) \\ &= \sum_{i \in I} \sigma_i \max(F^*_{i,1}, \ldots, F^*_{i,s}), \end{aligned} \tag{20}$$

where $F^*_{i,j} \coloneqq F_{i,j} \circ \phi \circ \mu_w$, for every $j \leq s$ and $i \in I$. The functions ϕ and μ_w are linear mappings. Their composition generates a new linear function, which implies that $F^*_{i,j}$ is linear for every $i \in I$ and $j \leq s$. Consequently, the function G_w is decomposed as a linear combination of $\max$ with at most $k+1$ arguments. Thus from Eq. (19) and Eq. (20) follow that the function G is represented as a linear combination of $\max$ with at most $k+1$ arguments and the following equality holds:

$$G = G_\ell + \sum_{i \in I} \sigma_i \max(F^*_{i,1}, \ldots, F^*_{i,s}).$$

□

Let $K \subseteq \{0, \ldots, n\}$ be the set of all indices such that, for any flag $\mathcal{H}$ of length $i \in K$, the delta function $\Delta_{\nabla F}(\mathcal{H})$ is constant. By Theorem III.3, the minimal arguments $k^* \in \mathbb{N}$ necessary to represent the function F as in Eq. (1) is calculated as follows:

$$k^* := \min\{k \mid k \in K\} + 1.$$

Theorem III.3 has the implication for different representations as well. For instance, it shows that the knowledge about the number of linear regions in the decomposition of the input space and the number of unique linear affine-linear functions provided by a CPWL function F are not enough to get optimal representations as a neural network or as a linear combination of $\max$ functions as in Eq. (1). By applying Theorem III.3 to the results from [7], we get the following corollary:

Corollary III.4. *A CPWL function* $F\colon \mathbb{R}^n \to \mathbb{R}$ *can be represented as a ReLU-based neural network with depth at most* $\lceil \log_2(k) \rceil + 1$*, where* $k \leq n+1$*, if for any flag* $\mathcal{H}$ *of length* $k-1$*, the delta function* $\Delta_{\nabla F}(\mathcal{H})$ *is constant.*

Proof. The statement holds by applying Theorem III.3 in Theorem 2.1 from [7]. □

REFERENCES

[1] A. Kripfganz and R. Schulze. Piecewise affine functions as a difference of two convex functions. *Optim.*, 18(1):23–29, 1987.
[2] J.M. Tarela and M.V. Martinez. A representation method for pwl functions oriented to parallel processing. *Math. Comput. Model.*, 13(10):75–83, 1990.
[3] J.M. Tarela and M.V. Martinez. Region configurations for realizability of lattice piecewise-linear models. *Math. Comput. Model.*, 30(11):17–27, 1999.
[4] S. Ovchinnikov. Max-min representation of piecewise linear functions. *Beitr. Algebra Geom.*, 43(1):297–302, 2002.
[5] L.O. Chua and A.-C. Deng. Canonical piecewise-linear representation. *IEEE Trans. Circuits Syst.*, 35(1):101–111, 1988.
[6] S. Wang. General constructive representations for continuous piecewise-linear functions. *IEEE Trans. Circuits Syst. I: Regul. Pap.*, 51(9):1889–1896, 2004.
[7] R. Arora, A. Basu, P. Mianjy, and A. Mukherjee. Understanding deep neural networks with rectified linear units. In *ICLR*, 2018.
[8] K.-L. Chen, H. Garudadri, and B. D. Rao. Improved bounds on neural complexity for representing piecewise linear functions. *NeurIPS*, 35:7167–7180, 2022.
[9] S. Wang and X. Sun. Generalization of hinging hyperplanes. *IEEE Trans. Inf. Theory*, 51:4425–4431, 12 2005.
[10] C. Hertrich, A. Basu, M. Di Summa, and M. Skutella. Towards lower bounds on the depth of relu neural networks. *SIAM J. Discrete Math.*, 37(2):997–1029, 2023.
[11] C. Koutschan, B. Moser, A. Ponomarchuk, and J. Schicho. Representing piecewise linear functions by functions with small arity. *AAECC*, pages 1–16, 2023.
[12] A. Brondsted. *An introduction to convex polytopes*, volume 90. Springer Science & Business Media, 2012.