利用格多胞体给出积分 ReLU 神经网络深度的下界

Christian Haase 德国柏林自由大学 haase@math.fu-berlin.de

Christoph Hertrich 英国伦敦政治经济学院 c.hertrich@lse.ac.uk

Georg Loho U荷兰特温特大学 g.loho@utwente.nl

arXiv:2302.12553v1 [cs.LG] 24 Feb 2023

摘要

我们证明，允许宽度任意增加时，整数权重 ReLU 神经网络可表示的函数集合会随着网络深度的增加而严格扩大。更确切地说，我们证明，计算 n 个数的最大值确实需要 $\lceil \log_2(n) \rceil$ 个隐藏层，这与已知的上界相匹配。我们的结果基于热带几何中神经网络与 Newton 多胞体之间已知的对偶性。整性假设意味着这些 Newton 多胞体都是格多胞体。随后，我们通过对这类多胞体各面的归一化体积进行奇偶性论证，得到深度下界。

1 引言

神经网络表达能力研究领域的经典结果是所谓的通用逼近定理（Cybenko，1989；Hornik，1991）。这些定理指出，浅层神经网络已经能够近似表示有界定义域上的任意连续函数。然而，要全面理解现代神经网络的运作机制，我们还希望回答以下问题：对于特定深度的神经网络，我们能够精确计算的函数集合究竟是什么？例如，关于精确表示能力的研究，近来推动了我们对训练神经网络计算复杂性的理解，涉及算法（Arora 等人，2018；Khalife 和 Basu，2022）以及困难性结果（Goel 等人，2021；Froese 等人，2022；Bertschinger 等人，2022）。

如今最为突出的激活函数或许是修正线性单元（ReLU）（Glorot 等人，2011；Goodfellow 等人，2016）。尽管它的流行主要源于直觉和实证上的成功，但从数学视角来看，用 ReLU 取代以往使用的平滑激活函数（如 S 形函数）带来了一些有趣的影响：离散几何中研究分段线性函数和多胞体的方法，突然在理解神经网络方面发挥了关键作用（Arora 等人，2018；Zhang 等，2018；Hertrich 等人，2021），补充了传统上占主导地位的解析视角。

这一方向上的一项基本结果由 Arora 等人（2018）提出，他们证明了一个函数能够由 ReLU 神经网络表示，当且仅当它是连续分段线性（CPWL）的。此外，他们的证明表明，$\lceil \log_2(n+1) \rceil$ 个隐藏层足以表示每个输入维数为 n 的 CPWL 函数。一个自然的后续问题是：这种对数级层数究竟是否确有必要，还是更浅的神经网络就已能表示所有 CPWL 函数?

Hertrich 等人（2021）猜想前一种情况成立。更准确地说，若 $\mathrm{ReLU}_n(k)$ 表示定义在 $\mathbb{R}^n$ 上且可由 k 个隐藏层计算的 CPWL 函数集合，则该猜想可表述如下：

猜想 1（Hertrich 等人（2021））。对所有 $k \leq \lceil \log_2(n+1) \rceil$，$\mathrm{ReLU}_n(k-1) \subsetneq \mathrm{ReLU}_n(k)$.

请注意，根据 Arora 等人（2018）的结果，$\mathrm{ReLU}_n(\lceil \log_2(n+1) \rceil)$ 是定义在 $\mathbb{R}^n$ 上的全体连续分段线性函数。

尽管 Hertrich 等人（2021）为他们的猜想提供了一些证据，但对于每个输入维度 $n \geq 4$，该问题仍未解决。更进一步，目前尚无任何一个已知的连续分段线性函数能够证明两个隐藏层不足以表示它。即便对于 $\max\{0, x_1, x_2, x_3, x_4\}$ 这样简单的函数，也无法确定两个隐藏层是否足够。

事实上，$\max\{0, x_1, x_2, x_3, x_4\}$ 并非只是一个随意举出的例子。基于 Wang 与 Sun（2005）的一个结果，Hertrich 等人（2021）证明，他们的猜想等价于以下命题。

猜想 2（Hertrich 等人（2021））。对于 $n = 2^k$，函数 $\max\{0, x_1, \ldots, x_n\}$ 不包含在 $\mathrm{ReLU}_n(k)$ 中。

这种重新表述带来了关于实际神经网络架构中常用的两个元素——最大池化和 Maxout——的有趣解释。最大池化单元位于（ReLU 或其他）层之间，仅输出若干输入的最大值（即将它们“池化”在一起）。它们本身不包含可训练参数。相比之下，Maxout 网络是 ReLU 网络的一种替代方案（事实上也是一种推广）。Maxout 网络中的每个神经元都会输出前一层输出的若干个（可训练的）仿射组合的最大值；而在 ReLU 情况下，则是将单个仿射组合与零进行比较。

因此，该猜想将意味着，实际上需要对数级数量的 ReLU 层来替代一个最大池化单元或一个 Maxout 层，这为这些元素确实比纯 ReLU 网络更强大提供了理论上的依据。

1.1 我们的结果

本文证明，在额外假设神经网络中的所有权重均为整数的情况下，Hertrich 等人（2021）的猜想对所有 $n \in \mathbb{N}$ 都成立。换言之，如果 $\mathrm{ReLU}_n^{\mathbb{Z}}(k)$ 表示由定义在 $\mathbb{R}^n$ 上、具有 k 个隐藏层且仅含整数权重的神经网络所可表示的函数集合，我们证明如下结果。

定理 3. 对于 $n = 2^k$，函数 $\max\{0, x_1, \ldots, x_n\}$ 不属于 $\mathrm{ReLU}_n^{\mathbb{Z}}(k)$。

证明定理 3 是我们的主要贡献。我们在第 1.2 节概述了整体思路，并在第 4 节整合所有论据，给出形式化证明。

Hertrich 等人（2021）中的论证可以加以改编，从而证明两个猜想之间的等价关系在整数情形下同样成立。因此，我们得出结论：增加整数神经网络的层数，确实能够扩大其可表示函数集合，直至层数达到对数级。形式化证明见第 4 节。

Corollary 4. $\mathrm{ReLU}_n^{\mathbb{Z}}(k-1) \subsetneq \mathrm{ReLU}_n^{\mathbb{Z}}(k)$ *for all* $k \leq \lceil \log_2(n+1) \rceil$.

据我们所知，我们的结果是首个在宽度不受任何限制的情况下，针对 ReLU 神经网络深度给出的非常数（即对数）下界。在没有整性假设的情况下，已知的最佳下界仍然是两个隐藏层（Mukherjee 和 Basu（2017））；对于简单函数 $\max\{0, x_1, x_2\}$，该下界已经成立。

虽然对于实际的神经网络应用而言，整性假设颇不现实，因为权重通常通过梯度下降进行调整；但从分析理论表达能力的角度来看，这一假设或许是合理的。为此，假设某个 ReLU 网络表示函数 $\max\{0, x_1, \ldots, x_n\}$。那么，每个分数权重都必须相互抵消，或与其他分数权重相加后得到某些整数，因为最终函数中的每个线性片段都只有整数系数。因此，假设一开始就不存在分数权重是有道理的。然而遗憾的是，这种直觉很难转化为证明，因为分数权重的组合可能会产生整数系数，而没有分数权重时则无法得到这些系数。

1.2 论证概要

我们证明的第一个关键步骤，是运用关于神经网络与热带几何之间联系的已有研究，这一研究由 Zhang 等人（2018）开创。每个连续分段线性函数都可以分解为两个凸分段线性连续函数之差。凸分段线性连续函数与某些多胞体之间存在简洁的对偶关系；在热带几何中，这些多胞体称为 Newton多胞体。对于任意神经网络，这种对偶关系使我们能够构造一对 Newton多胞体，它们可以唯一确定该神经网络所表示的连续分段线性函数。这些 Newton多胞体是逐层从点构造而成的，过程中交替取闵可夫斯基和与凸包。交替操作的次数对应于神经网络的深度。因此，分析具有特定深度的神经网络所能表示的函数集合，等价于理解经过特定次数的交替操作能够以这种方式构造出哪些多胞体（定理 8）。

将问题转化到多胞体的世界后，第二个要素是更好地理解构造这些多胞体时涉及的两种运算：闵可夫斯基和与凸包。我们分别证明，对于这两种运算，其结果都可以细分为结构更简单的多胞体：对于两个多胞体的闵可夫斯基和，细分中的每个胞腔都是原多胞体的面的仿射积，即一个与笛卡尔积仿射等价的多胞体（命题 9）。对于两个多胞体的凸包，每个胞腔都是原多胞体的两个面的连接，即两个面的凸包；这两个面的仿射包彼此斜交，也就是说，它们既不相交，也不具有平行的方向（命题 10）。

最后，有了这些细分，我们的第三个要素便是这些多胞体的体积。得益于我们的整性假设，相关多胞体的所有顶点的所有坐标都是整数；换言之，这些多胞体都是格多胞体。对于格多胞体，可以定义所谓的归一化体积（第 2.3 节）。这是欧几里得体积的一个缩放版本，其缩放因子取决于多胞体的仿射包。它的构造方式保证了所有格多胞体的归一化体积都是整数。利用此前得到的细分，我们证明：对于至多具有 k 个隐藏层的神经网络所对应的多胞体，其维数至少为 2^k 的面始终具有偶数的归一化体积（定理 16）。这意味着并非所有格多胞体都能以这种方式构造出来。再次利用受热带几何启发的多胞体与函数之间的对偶性（定理 8），我们将这一结果转回由神经网络计算的连续分段线性函数领域，并得到：k 个隐藏层不足以计算 $\max\{0, x_1, \ldots, x_{2^k}\}$，从而证明定理 3。

1.3 超越整数权重

鉴于整性假设，一个自然的想法是：能否将神经网络的所有权重乘以它们的公分母，从而将我们的结果简单推广到有理权重？遗憾的是，这行不通。我们的证明排除了具有 k 个隐藏层的整数 ReLU 网络计算 $\max\{0, x_1, \ldots, x_{2^k}\}$ 的可能性，但并未排除使用整数权重计算$2 \cdot \max\{0, x_1, \ldots, x_{2^k}\}$ 的可能性。特别是，将输出层的权重除以二，可能会得到一个能够计算 $\max\{0, x_1, \ldots, x_{2^k}\}$ 的半整数权重神经网络。

若要全面处理这一猜想，也就是考虑任意权重的情形，仅通过体积奇偶性进行论证似乎并不足够。奇偶性论证本质上是离散的，并且存在前文所述的问题。尽管如此，我们相信，本文的技术确实为全面解决这一猜想铺平了道路。特别是，第 3 节中构造的细分适用于任意多胞体，而不仅仅是格多胞体。因此，可以设想用另一种不变量取代标准化体积的奇偶性；将该不变量应用于这些细分后，便能得到非整数情形下的一般深度分离。

1.4 与先前工作的进一步联系

与我们的整性假设类似，Hertrich 等人（2021）也采用了一个附加假设，并在所谓的 H 一致神经网络这一特殊情形下证明了他们的猜想。需要指出的是，这两个假设互不蕴含：存在权重为非整数的 H一致网络，也存在不符合 H一致条件的整数神经网络。

我们的结果与所谓的深度和宽度之间的权衡结果有关（但在概念上有所不同）。这类结果表明，深度略微降低，就可能使维持相同的（精确或近似）表达能力所需的宽度呈指数级增加，

Telgarsky（2015；2016）证明了此类结果中的首批结论，而 Eldan & Shamir（2016）甚至证明了两层与三层之间的指数级分离。此后又有许多改进和相关结果相继建立（Arora 等人，2018；Daniely，2017；Hanin，2019；Hanin & Sellke，2017；Liang & Srikant，2017；Nguyen 等，2018；Raghu 等，2017；Safran & Shamir，2017；Safran 等，2019；Vardi 等，2021；Yarotsky，2017）。相比之下，我们关注精确表示，并展示了一种更为显著的权衡：将深度从对数级降至次对数级后，任何有限宽度都不再足够。

受热带几何启发，神经网络计算的连续分段线性函数与 Newton多胞体之间的对偶性，此前已用于多项神经网络相关研究（Maragos 等，2021），例如分析决策边界的形状（Alfarra 等，2020；Zhang 等，2018），或计算并界定线性片段的数量（Charisopoulos & Maragos，2018；Hertrich 等人，2021；Mont´ufar 等，2022）。

2 预备知识

我们将 $[n] \coloneqq \{1, 2, \ldots, n\}$ 记作不超过 n （不含零）的自然数集合。

2.1 ReLU 神经网络

对于任意 $n \in \mathbb{N}$，令 $\sigma\colon \mathbb{R}^n \to \mathbb{R}^n$ 为逐分量的整流函数

$$\sigma(x) = (\max\{0, x_1\}, \max\{0, x_2\}, \ldots, \max\{0, x_n\}).$$

对于任意数量的隐藏层 $k \in \mathbb{N}$，一个具有修正线性单元（ReLU 神经网络）的 $(k+1)$ 层前馈神经网络由 $k+1$ 个仿射变换 $T_{(\ell)}$给出： $\mathbb{R}^{n_{\ell-1}} \to \mathbb{R}^{n_\ell}$,$x \mapsto A^{(\ell)}x + b^{(\ell)}$，其中 $\ell \in [k+1]$。称其计算或表示函数 $f\colon \mathbb{R}^{n_0} \to \mathbb{R}^{n_{k+1}}$，该函数定义为

$$f = T^{(k+1)} \circ \sigma \circ T^{(k)} \circ \sigma \circ \cdots \circ T^{(2)} \circ \sigma \circ T^{(1)}.$$

矩阵 $A^{(\ell)} \in \mathbb{R}^{n_\ell \times n_{\ell-1}}$ 称为权重，向量 $b^{(\ell)} \in \mathbb{R}^{n_\ell}$ 称为第 ℓ 层的偏置。数值 $n_\ell \in \mathbb{N}$ 称为第 ℓ 层的宽度。所有隐藏层的最大宽度 $\max^{\ell \in [k]} n_\ell$ 称为神经网络的宽度。此外，我们称该神经网络的深度为 $k+1$，规模为 $\sum_{\ell=1}^{k} n_\ell$。

通常，神经网络表示为分层的有向无环图，其中每一层（包括输入层 $\ell = 0$ 和输出层 $\ell = k+1$）的每个维度对应一个顶点，权重作为弧的标签，偏置作为节点的标签。该图的顶点称为神经元。每个神经元对其前驱节点的输出进行仿射变换，再应用 ReLU 函数，并输出结果。

2.2 多胞体与格

我们简要介绍多胞体和格所需的概念；如需进一步阅读，请参见（Schrijver，1986，第 4、7–9 章）。对于任意两个集合 X、$Y \subseteq \mathbb{R}^n$，可以定义它们的闵可夫斯基和$X + Y = \{x + y \mid x \in X, y \in Y\}$。我们用 Span($X$) 表示集合 X 的通常线性包，也就是包含 X的最小线性子空间。仿射包 Aff(X) 是包含 X 的最小仿射子空间。最后，对于集合 X，我们定义线性空间 Lin(X) $=$Span $\{x - y \mid x, y \in X\}$，它是将 Aff($X$) 平移至原点后得到的。注意，除非 Aff($X$) 包含原点（此时这三个概念完全相同），否则 Lin(X) 通常是 Span(X) 的真子空间。

对于集合 $X \subset \mathbb{R}^n$，其凸包为

$$\operatorname{conv}(X) = \left\{ \sum_{s \in S} \lambda_s s \mid S \subseteq X \text{ finite}, \lambda_s \geq 0 \text{ for all } s \in S, \sum_{s \in S} \lambda_s = 1 \right\} .$$

多胞体是有限集合 $V \subset \mathbb{R}^n$ 的凸包。给定一个多胞体 $P \subset \mathbb{R}^n$， P 的面是对某个 $c \in \mathbb{R}^n$ 而言形如 arg min $\{c^\top x \mid x \in P\}$ 的集合；多胞体的面仍是多胞体。 P 的维数 dim(P) 是 Lin(P) 的维数。按照惯例，我们也称空集为一个

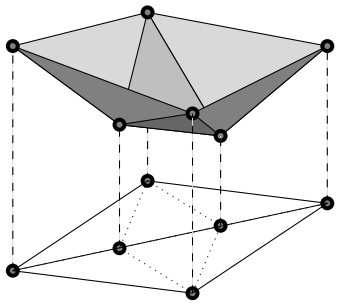


图 1：一个正则细分：$\mathbb{R}^2$ 中的四边形被细分为六个满维单元，这些单元是 $\mathbb{R}^3$中提升点的凸包的下方面的投影。

P 的维数为 -1 的面。维数为 0 的面是顶点，而维数为 $\dim(P) - 1$ 的面是面片。如果所有顶点都属于 $\mathbb{Q}^n$，我们称 P 为有理的；更进一步，如果 P 的所有顶点都是整数的，也就是说它们属于 $\mathbb{Z}^n$，那么 P 就是一个格多胞体。

一个重要的（格）多胞体示例是单纯形 $\Delta_0^n = \text{conv}\{0, e_1, e_2, \ldots, e_n\} \subseteq \mathbb{R}^n$，它由原点和所有 n 标准基向量张成。

最后，我们定义多胞体的（正则多面体）细分。设 $\tilde{P} \subset \mathbb{R}^{n+1}$ 为一个多胞体，并设 $P \subset \mathbb{R}^n$ 为它在投影 $\text{proj}_{[n]}$: $\mathbb{R}^{n+1} \to \mathbb{R}^n$ （即遗忘最后一个坐标）下的像。

$\tilde{P}$的下方面是关于线性目标函数 $(c^\top, 1) \in \mathbb{R}^n \times \mathbb{R}$的极小值点集合。将 P 的下方面投影到 $P \subset \mathbb{R}^n$，得到 P的一个细分~，也就是说，这是一个多胞体集合，其中各多胞体交于公共面，且覆盖 P。图 1 给出了平面中一个多胞体细分的示例。

格是 $\mathbb{R}^n$ 的一个子集，形如

$$L = \{B \cdot z \mid z \in \mathbb{Z}^p\}$$

对于某个矩阵 $B \in \mathbb{R}^{n \times p}$，其列向量线性无关。在这种情况下，我们称 $b^{(1)}$、.....、$b^{(p)}$ 为 B 的格基。L 中的每个元素都能唯一地表示为 $b^{(1)}$、.....、$b^{(p)}$ 的整数系数线性组合。选定一组格基，就将 L 与 $\mathbb{Z}^p$ 对应起来。

格的经典例子就是 $\mathbb{Z}^n$ 本身。在本文中，我们只研究形如以下形式的格：

$$\{x \in \mathbb{Z}^n \mid A \cdot x = 0\} \ ,$$

对于某个 $A \in \mathbb{Q}^{q \times n}$，也就是说，它是 $\mathbb{Z}^n$ 与 $\mathbb{R}^n$ 中某个有理的子空间的交集。

例如，向量 $\binom{2}{1}$、$\binom{1}{2}$ 构成 $\mathbb{R}^2$ 作为向量空间的一组基，但它们并不构成 $\mathbb{Z}^2$ 的格基。相反，它们生成了较小的格 $\{x \in \mathbb{Z}^2 \mid 3 \text{ 整除 } x_1 + x_2\}$。为同一个格选取两组基，会得到一个具有整数系数的基变换矩阵，其逆矩阵也具有整数系数。因此，它的行列式必为 ± 1。

我们主要讨论仿射格，即形如 $K = L + v$ 的集合，其中 $L \subset \mathbb{R}^n$ 是一个格，且 $v \in \mathbb{R}^n$。那么，如果 $b_1 - b_0, \ldots, b_r - b_0$ 是与 L 平行的线性格 $L = K - K$ 的一个（线性）格基，则 $b_0, \ldots, b_r \in K$ 构成一个仿射格基；也就是说，K 中的每个元素都可以唯一表示为 $b_0, \ldots, b_r$ 的仿射组合，且系数为整数。

2.3 归一化体积

我们证明中的主要工具是（牛顿）多面体各面的归一化体积。我们在有理多面体 $P \subset \mathbb{R}^n$ 所在的仿射包 Aff(P) 内测量其体积（该多面体可能不是满维的）。为此，我们选取与 Aff(P) 平行的线性空间 Lin(P) 中 Lin(P) 的一组格基 $\cap \mathbb{Z}^n$。由于 P 是有理多面体，这组格基便确定了一个从 Lin(P) 到 $\mathbb{R}^r$（维数为 $r =$）的线性变换，该变换将格基向量映射为标准单位向量。于是，P 的归一化体积 Vol(P) 等于 $r!$ 乘以其在该变换下的像的欧几里得体积。缩放因子 $r!$ 的选取使得对于每个 $n \geq 1$，单纯形 Δ_0^n 的归一化体积均为 1。

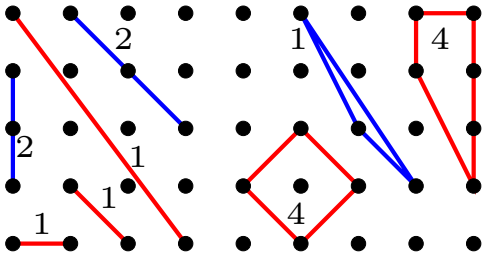


图 2：一些格点线段和格点多边形及其归一化体积。请注意，格多胞体的归一化体积不同于其欧几里得体积，因为它是相对于其仿射包中诱导的格测量的。

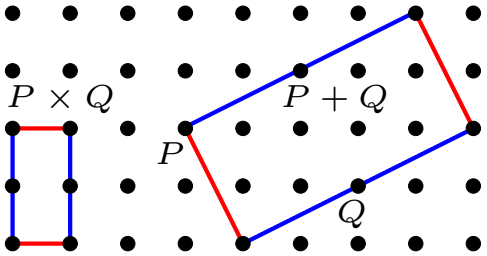


图 3：P 和 Q 的仿射积的归一化体积，可以通过笛卡尔积的体积乘以将 $P \times Q$（左）映射到 $P + Q$（右）的线性映射的行列式绝对值来计算：$\mathrm{Vol}(P+Q) = \binom{1+1}{1} \cdot \left|\det \left(\begin{smallmatrix} -1 & 2 \\ 2 & 1 \end{smallmatrix}\right)\right| \cdot \mathrm{Vol}(P) \cdot \mathrm{Vol}(Q)$。

任意两组格基之间相差一个逆矩阵同样为整矩阵的整矩阵（因此行列式为 ± 1），这保证了上述定义良好。对于满维多胞体，这仅仅是欧几里得体积的一个缩放版本。但对于低维多胞体，我们相对于格进行的归一化不同于欧几里得归一化（参见图 2）。对我们而言，关键性质是每个格多胞体的体积都是整数（例如，参见 Beck & Robins, 2007, §3.5,§5.4）。这使我们能够利用体积的整除性质进行论证。
我们对以下基本结论进行可视化，并将具体证明留在附录中。

引理 5. 设 P、$Q \subset \mathbb{R}^n$ 是两个格多胞体，其中 $i = \dim(P)$且 $j = \dim(Q)$。若有 $\dim(P+Q) = \dim(P) + \dim(Q)$，则

$$\binom{i+j}{i} \cdot \mathrm{Vol}(P) \cdot \mathrm{Vol}(Q) \ \textit{divides} \ \mathrm{Vol}(P+Q)\,.$$

若 P和 Q 满足引理 5 的假设，则称 $P+Q$ 为 P和 Q的仿射积。这与引言中给出的定义等价；为便于比较，我们在图 3 中展示了笛卡尔积$P \times Q$ 。

引理 6. 设 P、$Q \subset \mathbb{R}^n$ 为两个格多胞体，满足 $i = \dim(P)$ 且 $j = \dim(Q)$。若 $\dim(\mathrm{conv}(P \cup Q)) = \dim(P) + \dim(Q) + 1$，则

$$\mathrm{Vol}(P) \cdot \mathrm{Vol}(Q) \ \textit{divides} \ \mathrm{Vol}(\mathrm{conv}(P \cup Q))\,.$$

若 P和 Q 满足引理 6 的假设，则称 $\mathrm{conv}(P \cup Q)$为 P和 Q的连接，参见图 4。此定义与引言中给出的定义等价。

2.4 NEWTON 多胞体与神经网络

证明主要结果的第一个要点，是利用此前发现的、受热带几何启发的分段线性连续函数与多胞体之间的对偶性，将问题转化到多胞体的框架中。

首先，我们注意到可以将讨论限制在不含偏置的神经网络上。为此，若函数 $g\colon \mathbb{R}^n \to \mathbb{R}^m$ 满足对所有 $\lambda \geq 0$，$g(\lambda x) = \lambda g$({v6)，则称其为正齐次函数。

引理 7（Hertrich 等人（2021））。如果一个神经网络表示一个正齐次函数，那么将该神经网络的所有偏置参数 $b^{(\ell)}$ 设为零后，所表示的函数完全相同。

由于 $\max\{0, x_1, \ldots, x_{2^k}\}$ 是正齐次的，这意味着为了证明定理 3，我们可以专注于不含偏置的神经网络。此类神经网络计算得到的函数始终是正齐次的。

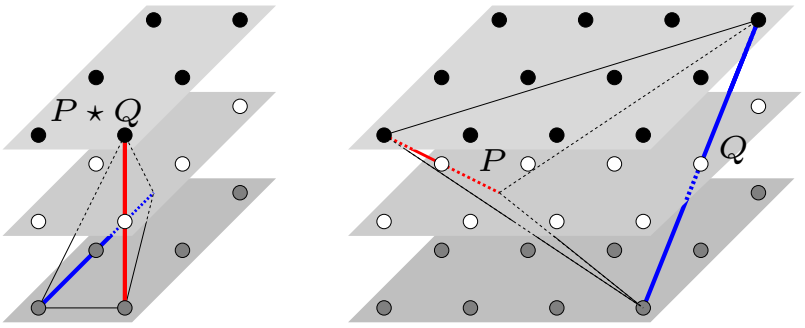


图 4: P 和 Q 连接后的归一化体积，可通过将 P和 Q 的归一化体积相乘，再乘以将 $P \star Q$（左侧）映射到 conv($P \cup Q$)（右侧）的线性映射的行列式绝对值的整数的值来计算：Vol(conv($P \cup Q$)) $= \left|\det \begin{pmatrix} -2 & 0 & 0 \\ 2 & 1 & -1 \\ 0 & 1 & 1 \end{pmatrix}\right| \cdot \text{Vol}(P) \cdot \text{Vol}(Q)$，其中 $P \star Q$ 表示将 P和 Q 嵌入到距离为 1 的正交子空间中。

设 f 是正齐次的凸分段线性函数。凸性意味着我们可以将 f 写成其线性片段的最大值，即 $f(x) = \max\{a_1^\top x, \ldots, a_p^\top x\}$。与 f 对应的 Newton 多胞体 P_f 定义为系数向量的凸包，即 $P_f = \text{conv}\{a_1, \ldots, a_p\}$。事实证明，函数上的两种基本运算 $+$ 和 max（逐点应用）分别对应于相应 Newton 多胞体的闵可夫斯基和与凸包：$P_{f+g} = P_f + P_g$ 和 $P_{\max\{f,g\}} = \text{conv}\{P_f \cup P_g\}$，参见 Zhang 等人（2018）。由于神经网络基本上由仿射变换（即加权和）与最大值运算交替构成，这启发我们定义以下多胞体类 $\mathcal{P}_k$，直观上对应于具有 k 个隐藏层的整数权重神经网络。请注意，类 $\mathcal{P}_k$ 包含维数各异的多胞体。

我们从 $\mathcal{P}_0$（格点集合）开始。对于 $k \geq 0$，我们定义

$$\begin{aligned} \mathcal{P}'_{k+1} &= \{\text{conv}(Q_1 \cup Q_2) \mid Q_1, Q_2 \in \mathcal{P}_k\} \ , \\ \mathcal{P}_{k+1} &= \left\{Q_1 + \cdots + Q_\ell \mid Q_1, \ldots, Q_\ell \in \mathcal{P}'_{k+1}\right\} \ . \end{aligned} \tag{1}$$

将 $\mathcal{P}_k$ 与具有 k 个隐藏层的神经网络联系起来，可以通过以下内容形式化。定理中的差异考虑到了 f 不一定是凸的这一事实；即使它是凸的，神经网络计算出的中间函数也可能是非凸的。

定理 8. 正齐次分段线性连续函数 f 可以表示为一个含 k 个隐藏层的整数神经网络，当且仅当 $f = g - h$ 对于两个凸的正齐次分段线性连续函数 g 和 h，满足 $P_g, P_h \in \mathcal{P}_k$。

关于非整数情形的同一定理的证明可见 Hertrich (2022, 定理 3.35)。仔细审查其中的证明可以发现，该证明同样适用于整数情形。为完整起见，我们在附录中给出证明。

3 细分Minkowski和与凸包

本节旨在介绍我们证明中使用的主要几何工具：将闵可夫斯基和与并集分别细分为仿射积与连接。更确切地说，我们将证明以下两个结论。这些结论适用于一般多胞体，而不仅仅适用于格多胞体。

命题 9. 对于 $\mathbb{R}^n$ 中的两个多胞体 P和 Q，存在 $P + Q$ 的一个细分，使得每个满维胞腔都是 P 的一个面与 Q 的一个面的仿射积。

命题 10. 对于 $\mathbb{R}^n$ 中的两个多胞体 P 和 Q，存在 conv$\{P \cup Q\}$ 的一个细分，使得每个满维胞腔都是 P 的一个面与 Q 的一个面的连接。

证明这些结论的策略是：以一般方式将多胞体 P和 Q 提升一个维数至 $\mathbb{R}^{n+1}$，在该空间中执行相应的运算（闵可夫斯基和或凸包），然后将所得多胞体在 $\mathbb{R}^{n+1}$ 中的下方面投影回 $\mathbb{R}^n$，从而得到细分。

更确切地说，给定一个 $\alpha \in \mathbb{R}^n$ 和 $\beta \in \mathbb{R}$，我们定义

$$P^{\mathbf{0}} \coloneqq \{(p,0) \mid p \in P\} \quad \text{and} \quad Q^{\alpha,\beta} \coloneqq \left\{(q, \alpha^\top q + \beta) \mid q \in Q\right\} \ .$$

注意，将 $\pi_{[n]}(P^{\mathbf{0}})$ 和 $Q^{\alpha,\,\beta}$ 投影到前 n 个坐标上，分别得到 P 和 Q。此外，将 $P^{\mathbf{0}} + Q^{\alpha,\,\beta}$ 和 $\{P^{\mathbf{0}} \cup Q^{\alpha,\beta}\}$ 的下方面分别向下投影到前 n 个坐标上，可得到 $P{+}Q$ 和 $\{P{\cup}Q\}$ 的细分。接下来需要证明，可以选择参数，使这些细分具有所需的性质。

为此，我们引入以下记号来表示相应多胞体的面。对于每个 $c \in \mathbb{R}^n \setminus \{0\}$，令

$$F_c^{\mathbf{0}} = \arg\min\left\{(c^\top, 1)z \mid z \in P^{\mathbf{0}}\right\} \quad \text{and} \quad G_c^{\alpha,\beta} = \arg\min\left\{(c^\top, 1)z \mid z \in Q^{\alpha,\beta}\right\} \ ,$$

以及

$$F_c = \arg\min\{c^\top x \mid x \in P\} \quad \text{and} \quad G_c = \arg\min\{(c+\alpha)^\top x \mid x \in Q\} \ .$$

注意，$F_c = \pi_{[n]}(F_c^{\mathbf{0}})$ 和 $G_c = \pi_{[n]}(G^{\alpha,\,{}_c\,\beta})$。

据此，我们终于可以明确"泛化地选择 α 和 β"的含义：这意味着 F_c 和 G_c 的线性包与仿射包尽可能少地相交。

引理 11。可以选择 α和 β，使得 $\text{Lin } F_c \cap \text{Lin } G_c = \{\mathbf{0}\}$ 且 $\text{Aff } F_c^{\mathbf{0}} \cap \text{Aff } G^{\alpha,\,{}_c\,\beta} = \emptyset$对每个 $c \in \mathbb{R}^n \setminus \{0\}$都成立。

利用关于 α 和 β的泛型选择的引理，我们得到了证明所需细分存在性的工具。具体证明见附录。

4 利用归一化体积证明深度下界

本节中，我们通过对上一节构造的细分中各胞腔的归一化体积进行奇偶性论证，完成证明。

设 $\mathcal{Q}_k$ 为满足以下性质的格多胞体 P 的集合：对于 P 的每个面 F，若 $\dim(F) \geq 2^k$，则 $\text{Vol}(F)$为偶数。注意，类 $\mathcal{Q}_k$ 包含不同维数的多胞体，尤其包含所有维数小于 2^k的格多胞体。

我们证明定理 3 的计划如下。首先证明类 $\mathcal{Q}_k$ 在闵可夫斯基加法下封闭，并且对 $\mathcal{Q}_k$ 中多胞体的凸包取并集总能得到一个多胞体 $\mathcal{Q}_{k+1}$。利用这一点，通过归纳可得 $\mathcal{P}_k \subseteq \mathcal{Q}_k$。然后证明，将单纯形 $\Delta_0^{2^k}$ 加到 $\mathcal{Q}_k$ 中的多胞体上，会得到一个绝不包含于 $\mathcal{Q}_k$中的多胞体。结合这一分离结果、 $\mathcal{P}_k \subseteq \mathcal{Q}_k$ 和定理 8，即可证明定理 3。本节大多数证明背后的一般思路是使用上一节构造的细分，并利用第 2.3 节的引理，分别论证细分中每个胞腔的体积。

4.1 $\mathcal{Q}$的闭性

命题 12. 对于 P、$Q \in \mathcal{Q}_k$，有 $P + Q \in \mathcal{Q}_k$。

命题 12 的证明基于以下引理。

引理 13. 如果 $d \coloneqq \dim(P+Q) \geq 2^k$，那么 $\text{Vol}(P+Q)$ 为偶数。

引理 13 的证明。由命题 9 可知，可以对 $P+Q$ 进行细分，使得细分中的每个满维胞腔 C 都是 P和 Q的两个面 F和 G 的仿射积。令 $i \coloneqq \dim(F)$，且 $j \coloneqq \dim(G)$，则 $d = i + j$。由引理 5 可知，对于某个 $z \in \mathbb{Z}$，$\text{Vol}(C) = z \cdot \binom{d}{i} \cdot \text{Vol}(F) \cdot \text{Vol}(G)$。现在我们证明这个量总是偶数。如果 i 或 j 至少为 2^k，那么 $\text{Vol}(F)$ 或 $\text{Vol}(G)$（分别）为偶数，因为 P和 Q都包含在 $\mathcal{Q}_k$中。否则， $d = i + j \geq 2^k$，但 $i, j < 2^k$。在这种情况下，$\binom{d}{i}$ 总是偶数，这是卢卡斯定理（卢卡斯，1878年）的直接推论。无论哪种情况，对于细分中的每个胞腔 C，$\text{Vol}(C)$都是偶数。因此， $P+Q$ 的总体积也为偶数。 □

命题 12 的证明。为了证明该命题，我们需要确保不仅 $P+Q$，而且 $P+Q$ 中维数至少为 2^k 的每个面都具有偶数归一化体积。若 F 是这样的一个面，那么根据闵可夫斯基和的基本性质，它是 P 和 Q 的两个面 P' 和 Q' 的闵可夫斯基和。由于 $\mathcal{Q}_k$ 在取面运算下封闭，因此 $P', Q' \in \mathcal{Q}_k$。于是，将引理 13 应用于 P' 和 Q'，可知 F 的体积为偶数。对所有维数至少为 2^k 的面 F 执行此操作，即完成证明。 □

命题 14. 对于 P、$Q \in \mathcal{Q}_k$，我们有 $\text{conv}(P \cup Q) \in \mathcal{Q}_{k+1}$。

同样，证明的关键在于以下引理。
引理 15. 如果 $d := \dim(\text{conv}(P \cup Q)) \geq 2^{k+1}$，那么 $\text{Vol}(\text{conv}(P \cup Q))$ 是偶数。

我们将引理 15 和命题 14 的证明推迟到附录，因为它们类似于命题 12 和引理 13 的证明；后两者基于凸包的相应结论（引理 6 和命题 10），而非闵可夫斯基和。

定理16. 对所有 $k \in \mathbb{N}$，$\mathcal{P}_k \subseteq \mathcal{Q}_k$ 都成立。

证明。我们通过对 k 进行归纳来证明这一陈述。类 $\mathcal{P}_0$ 只包含点，因此 $\mathcal{P}_0$ 中没有多胞体具有维数至少为 $2^0 = 1$ 的面。因此，显然可得 $\mathcal{P}_0 \subseteq \mathcal{Q}_0$，归纳的起始步骤至此完成。归纳步骤则通过将命题 12 和命题 14 应用于类 $\mathcal{P}_k$ 的定义 (1) 得出。 □

4.2 特立独行者

定理 3 的最后一个关键步骤是说明 Δ_0^n 如何打破 $\mathcal{Q}_k$ 的结构。
命题 17. 如果 $P \in \mathcal{Q}_k$ 是 $\mathbb{R}^n$ 中一个满足 $n = 2^k$ 的多胞体，那么 $P + \Delta_0^n \notin \mathcal{Q}_k$。

证明。将命题 9 应用于 P 和 $Q = \Delta_0^n$，可得 $P + \Delta_0^n$ 能够被细分，使得每个满维胞腔 C 都是 P 的一个面 F 与 Δ_0^n 的一个面 G 的仿射积。
与引理 13 的证明一样，可以推出所有这些单元的体积都是偶数，但有一个例外：如果 $\dim F = 0$ 且 $\dim G = 2^k$。重新审视命题 9 的证明可知，细分中恰好存在一个这样的胞腔 C。该胞腔是 Δ_0^n 的平移版本，因此其体积为 $\text{Vol}(C) = 1$。

由于细分中的所有单元体积都是偶数，除了一个体积为奇数的单元，我们得到 $\text{Vol}(P + \Delta_0^n)$ 为奇数。因此，$P + \Delta_0^n$ 不能包含于 $\mathcal{Q}_k$ 中。 □

定理 3. 对于 $n = 2^k$，函数 $\max\{0, x_1, \ldots, x_n\}$ 不包含于 $\text{ReLU}_n^{\mathbb{Z}}(k)$ 中。

证明。为导出矛盾，假设存在一个具有整数权重和 k 个隐藏层的神经网络，计算 $f(x) = \max\{0, x_1, \ldots, x_n\}$。由于 f 的 Newton多胞体为$P_f = \Delta_0^n$，定理 8 推出存在两个多胞体 P、$Q \in \mathcal{P}_k$，它们位于 $\mathbb{R}^n$ 中，并且满足 $P + \Delta_0^n = Q$。根据定理 16，我们得到 P、$Q \in \mathcal{Q}_k$。这与命题 17 矛盾。 □

由此我们得出结论：增加更多层时，整数 ReLU 神经网络所能表示的函数集合会严格扩大。

推论 4. 对所有 $k \leq \lceil \log_2(n+1) \rceil$，$\text{ReLU}_n^{\mathbb{Z}}(k-1) \subsetneq \text{ReLU}_n^{\mathbb{Z}}(k)$。

证明。注意，$k \leq \lceil \log_2(n+1) \rceil$ 意味着 $2^{k-1} \leq n$。令 $f\colon \mathbb{R}^n \to \mathbb{R}$ 为函数 $f(x) = \max\{0, x_1, \ldots, x_{2^{k-1}}\}$。根据定理 3，我们得到 $f \notin \text{ReLU}_n^{\mathbb{Z}}(k-1)$。相比之下，不难看出，$k-1$ 个权重为整数的隐藏层足以计算 $\max\{x_1, \ldots, x_{2^{k-1}}\}$，例如可参见 Zhang 等人（2018）；Hertrich（2022）。在该网络的输出端附加一个 ReLU 神经元，意味着 $f \in \text{ReLU}_n^{\mathbb{Z}}(k)$。 □

致谢

Christian Haase 得到了德国科学基金会（DFG）HA 4383/8-1 项目资助。

Christoph Hertrich 得到欧洲研究委员会（ERC）在欧盟地平线2020研究与创新计划下的资助（资助协议 ScaleOpt–757481）。

参考文献

Motasem Alfarra、Adel Bibi、Hasan Hammoud、Mohamed Gaafar 和 Bernard Ghanem。《深度神经网络的决策边界：一种热带几何视角》。arXiv:2002.08838，2020 年。

Raman Arora、Amitabh Basu、Poorya Mianjy 和 Anirbit Mukherjee。《利用整流线性单元理解深度神经网络》。载于《国际学习表征会议》（ICLR），2018 年。

Matthias Beck 和 Sinai Robins。《离散地计算连续问题》。数学本科生教材。Springer，纽约，2007。ISBN 978-0-387-29139-0；0-387-29139-3。多面体中的整数点枚举。

Daniel Bertschinger、Christoph Hertrich、Paul Jungeblut、Tillmann Miltzow 和 Simon Weber。训练全连接神经网络是 ∃ℝ-完全问题。arXiv:2204.01368，2022。

Vasileios Charisopoulos 和 Petros Maragos。用热带方法研究具有分段线性激活函数的神经网络。arXiv:1805.08749，2018。

George Cybenko。通过S形函数的叠加进行逼近。《控制、信号与系统数学》，2(4):303–314，1989。

Amit Daniely。神经网络的深度分离。载于《学习理论会议》，第690–696页。PMLR，2017。

Ronen Eldan 和 Ohad Shamir。深度对前馈神经网络的作用。在会议学习理论（COLT），2016 年。

Vincent Froese、Christoph Hertrich 和 Rolf Niedermeier。以数据维数为参数的 ReLU 网络训练的计算复杂性。《人工智能研究杂志》，74：1775–1790，2022 年。

Xavier Glorot、Antoine Bordes 和 Yoshua Bengio。深度稀疏整流神经网络。在国际人工智能与统计会议（AISTATS），2011 年。

Surbhi Goel、Adam R. Klivans、Pasin Manurangsi 和 Daniel Reichman。训练深度为 2 的 ReLU 网络的紧致困难性结果。在第 12 届理论计算机科学创新会议（ ITCS ），2021 年。

Ian Goodfellow、YoshuaBengio 和 Aaron Courville。《深度学习》。麻省理工学院出版社，2016 年。http://www.deeplearningbook.org。

Boris Hanin。利用宽度受限且采用 ReLU 激活函数的深度神经网络实现通用函数逼近。《Mathematics》，7(10):992，2019 年。

Boris Hanin 和 Mark Sellke。用最小宽度的 ReLU 网络逼近连续函数。arXiv:1710.11278，2017 年。

Christoph Hertrich。神经网络复杂度的面片。博士论文，柏林工业大学，柏林，2022 年。URL：http://dx.doi.org/10.14279/depositonce-15271。

Christoph Hertrich、Amitabh Basu、Marco Di Summa 和 Martin Skutella。关于 ReLU 神经网络深度下界的研究。载于《神经信息处理系统进展》（NeurIPS），2021。

Kurt Hornik。多层前馈网络的逼近能力。《神经网络》，4（2）：251–257，1991。Sammy Khalife 和 Amitabh Basu。具有线性阈值激活的神经网络：结构与算法。载于国际整数规划与组合优化会议，第 347–360 页。Springer，2022。Shiyu Liang 和 Rayadurgam Srikant。为什么要用深度神经网络进行函数逼近？载于国际学习表征会议（ICLR），2017。Edouard Lucas。Th´eorie des fonctions num´eriques simplement p´eriodiques。《美国数学杂志》，第 289–321 页，1878。Petros Maragos、Vasileios Charisopoulos 和 Emmanouil Theodosis。热带几何与机器学习。《IEEE会刊》，109（5）：728–755，2021。Guido Mont´ufar、Yue Ren 和 Leon Zhang。Maxout 网络区域数及闵可夫斯基和顶点数的精确界。《SIAM应用代数与几何杂志》，6（4）：618–649，2022。Anirbit Mukherjee 和 Amitabh Basu。深度神经网络在布尔输入上的下界（采用 ReLU 门）。arXiv:1711.03073，2017。Quynh Nguyen、Mahesh Chandra Mukkamala 和 Matthias Hein。神经网络应具有足够宽度，才能学习不连通决策区域。载于国际机器学习会议（ICML），2018。Maithra Raghu、Ben Poole、Jon Kleinberg、Surya Ganguli 和 Jascha Sohl Dickstein。深度神经网络的表达能力。载于国际机器学习会议（ICML），2017。Itay Safran 和 Ohad Shamir。用神经网络逼近自然函数时的深度—宽度权衡。载于国际机器学习会议（ICML），2017。Itay Safran、Ronen Eldan 和 Ohad Shamir。神经网络中的深度分离：究竟分离的是什么？载于学习理论会议，第 2664–2666 页。PMLR，2019。Alexander Schrijver。《线性与整数规划理论》。约翰·威立父子出版公司，纽约，1986。Matus Telgarsky。深度前馈网络的表示优势。arXiv:1509.08101，2015。Matus Telgarsky。深度神经网络的深度优势。载于学习理论会议（COLT），2016。Gal Vardi、Daniel Reichman、Toniann Pitassi 和 Ohad Shamir。用神经网络逼近良性函数时的规模与深度分离。载于学习理论会议，第 4195–4223 页。PMLR，2021。Shuning Wang 和 Xusheng Sun。铰接超平面的泛化。《IEEE信息论汇刊》，51（12）：4425–4431，2005。Dmitry Yarotsky。深度 ReLU 网络逼近的误差界。《神经网络》，94：103–114，2017。Liwen Zhang、Gregory Naitzat 和 Lek-Heng Lim。深度神经网络的热带几何。载于国际机器学习会议（ICML），2018。

A 附加证明

引理 5. 设 P、$Q \subset \mathbb{R}^n$ 是两个格多胞体，满足 $i = \dim(P)$ 和 $j = \dim(Q)$。若有 $\dim(P+Q) = \dim(P) + \dim(Q)$，则

$$\binom{i+j}{i} \cdot \mathrm{Vol}(P) \cdot \mathrm{Vol}(Q) \ \textit{divides} \ \mathrm{Vol}(P+Q)\,.$$

证明. 假设 $\dim(P+Q) = \dim(P) + \dim(Q)$ 等价于 $\mathrm{Lin}(P) \cap \mathrm{Lin}(Q) = \{0\}$，因此存在一个仿射双射 $f\colon \mathrm{Aff}(P) \times \mathrm{Aff}(Q) \to \mathrm{Aff}(P+Q)$，将格点映射为格点。因此，$P+Q$ 的体积等于 $P \times Q$ 的体积乘以 f 的（整数）行列式（参见图 3）。

勒贝格测度的乘法性与我们的归一化共同给出

$$\frac{\mathrm{Vol}(P \times Q)}{(i+j)!} = \frac{\mathrm{Vol}(P)}{i!} \cdot \frac{\mathrm{Vol}(Q)}{j!}\,.$$

□

引理 6. 设 P、$Q \subset \mathbb{R}^n$ 是两个格多胞体，满足 $i = \dim(P)$ 和 $j = \dim(Q)$。若 $\dim(\text{凸包}(P \cup Q)) = \dim(P) + \dim(Q) + 1$，则

$$\mathrm{Vol}(P) \cdot \mathrm{Vol}(Q) \ \textit{divides} \ \mathrm{Vol}(\mathrm{conv}(P \cup Q))\,.$$

证明。其本质上是相同的证明，只需将乘积 $P \times Q$ 替换为连接

$$P \star Q := \mathrm{conv}\left(P \times \{0\} \times \{0\} \ \cup \ \{0\} \times Q \times \{1\}\right) \subset \ \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}\,.$$

（参见图 4。）

□

定理 8。正齐次分段线性连续函数 f 可以表示为一个整数的 k层隐藏层神经网络，当且仅当 $f = g - h$ 对于两个凸的正齐次分段线性连续函数 g 和 h 成立，其中 P_g、$P_h \in \mathcal{P}_k$。

证明。我们对 k进行归纳。对于 $k = 0$，该陈述显然成立，因为这两个性质恰好分别适用于具有整数系数的线性函数，从而完成归纳起点。

对于归纳步骤，假设对于某个 $k \geq 1$，等价性对不超过 $k-1$ 个隐藏层均成立。我们证明它对 k 个隐藏层也成立。

首先，考虑正向方向，即假设 f 可以由一个整数的 k层隐藏层神经网络表示。根据 Zhang 等（2018，引理 6.2），f 是两个热带多项式 g 和 h之差（热带商）。此外，根据同一引理，P_g 和 P_h 是与（$k-1$）层隐藏层神经网络相关联的成对 Newton多胞体的凸包的加权闵可夫斯基和。根据归纳假设，这些多胞体包含于 $\mathcal{P}_{k-1}$中。由（1）可知，这意味着 P_g 和 P_h 包含于 $\mathcal{P}_k$中；其中，神经网络的权重为整数这一事实保证了所得多胞体是格多胞体。至此，第一方向得证。

反过来，假设 f 可以写成 $g - h$，其中 P_g、$P_h \in \mathcal{P}_k$。根据 Hertrich（2022，命题 3.34）和 (1)，这意味着 g 和 h 可以写成凸分段线性函数对的最大值之和，其 Newton 多胞体位于 $\mathcal{P}_{k-1}$中。通过归纳，这些函数可以用具有 $(k-1)$ 个隐藏层的神经网络表示。为得到 g 和 h 而进行的额外最大值运算与求和运算，可以通过再增加一个隐藏层来实现（参见 Hertrich（2022，命题 2.2））。$\mathcal{P}_k$ 仅包含格多胞体这一事实，确保所得神经网络仅具有整数权重。由于 g和 h可以用 k 个隐藏层表示，因此 f也同样如此。

□

引理 11。可以选择 α和 β，使得 $\mathrm{Lin}\ F_c \cap \mathrm{Lin}\ G_c = \{\mathbf{0}\}$ 且 $\mathrm{Aff}\ F_c^{\mathbf{0}} \cap \mathrm{Aff}\ G_{c}^{\alpha,\beta} = \emptyset$对每个 $c \in \mathbb{R}^n \setminus \{0\}$都成立。

证明。我们将证明，不满足这些条件的参数 α 和 β 构成一个测度为零集。

首先来看第一条陈述。假设存在一个方向 $c \in \mathbb{R}^n \setminus \{0\}$，使得 Lin $F_c \cap$Lin $G_c \supsetneq \{\mathbf{0}\}$ 是一个维数至少为一的子空间。根据 F_c 和 G_c 的定义，我们得到 $c \in$(Lin F_c)$^\perp$ 以及 $c + \alpha \in$(Lin G_c)$^\perp$。这意味着 $\alpha \in$(Lin F_c)$^\perp$ +(Lin G_c)$^\perp$ =(Lin $F_c \cap$Lin G_c)$^\perp$，这是一个维数至多为 $n-1$的子空间，因此是一个测度为零集。因此，选取 α，使其对所有满足 Lin $F \cap$ Lin $G \supsetneq \{\mathbf{0}\}$ 的面 F和 G组成的面偶，都不属于 (Lin $F \cap$Lin G)$^\perp$，即可保证第一条条件成立。

对于第二个陈述，我们选择满足第一个陈述的 α，并注意到对于所有$c \in \mathbb{R}^n \setminus \{0\}$，都有 Lin $F_c \cap$ Lin $G_c = \{\mathbf{0}\}$。这意味着 $|$Aff $F_c \cap$Aff $G_c| \leq 1$，因此|Aff $F_c^{\mathbf{0}} \cap$Aff $G^{\alpha,\beta}_c| \leq 1$。假设存在一个方向 $c \in \mathbb{R}^n \setminus \{0\}$，使得 Aff $F_c^{\mathbf{0}} \cap$Aff $G^{\alpha,\beta}_c = \{(x,y)\} \subseteq \mathbb{R}^n \times \mathbb{R}$。根据我们从 $\mathbb{R}^n$到 $\mathbb{R}^{n+1}$的提升函数的选取，可得 $0 = y = \alpha^\top x + \beta$。因此，选择满足对所有下方面对 $F^{\mathbf{0}}$，其中 $F^{\mathbf{0}}$是 $P^{\mathbf{0}}$的下方面，且 $^\beta$是 Q^α的下方面，并且 Aff $F^{\mathbf{0}} \cap$Aff $G^{\alpha,\beta} = \{(x,y)\}$ 的所有配对均有 $0 \neq \alpha x + \beta$ 的 β，即可保证第二个条件成立。 □

命题 9. 对于 $\mathbb{R}^n$ 中的两个多胞体 P 和 Q，存在 $P+Q$ 的一种细分，使得每个满维胞腔都是 P 的一个面与 Q 的一个面的仿射积。

证明. 不失一般性，我们可以假设 $P+Q$ 是满维的。根据引理 11 选取 α 和 β，并考虑由将 $P^{\mathbf{0}} + Q^{\alpha,\beta}$ 的下方面向下投影所诱导的细分。设 C 是该细分中维数为 n 的一个胞腔，它是 $P^{\mathbf{0}} + Q^{\alpha,\beta}$ 的一个下方面 $\tilde{C}$ 的投影，且该下方面使方向 $(c^\top, 1) \in \mathbb{R}^n \times \mathbb{R}$ 最小。由闵可夫斯基和的基本性质可知 $\tilde{C} = F_c^{\mathbf{0}} + G^{\alpha,\beta}_c$，因此 $C = F_c + G_c$。由引理 11 可得$n \geq \dim F_c + \dim G_c \geq \dim C = n$。因此，$C$ 是 F_c 和 G_c 的仿射积。 □

命题 10. 对于 $\mathbb{R}^n$ 中的两个多胞体 P 和 Q，存在 conv$\{P \cup Q\}$ 的一个细分，使得每个满维胞腔都是 P 的一个面与 Q 的一个面的连接。

证明. 不失一般性，我们可以假设 conv$\{P \cup Q\}$ 是满维的。根据引理 11 选取 α和 β，并考虑通过向下投影 conv$\{P^{\mathbf{0}} \cup Q^{\alpha,\beta}\}$的下方面所诱导的细分。设 C 是该细分中一个维数为 n 的胞腔，它是 conv$\{P^{\mathbf{0}} \cup Q^{\alpha,\beta}\}$的一个下方面 $\tilde{C}$ 的投影，且该下方面沿方向 $(c^\top, 1) \in \mathbb{R}^n \times \mathbb{R}$ 取最小值。如果 $\tilde{C}$ 仅包含来自 $P^{\mathbf{0}}$ 或 $Q^{\alpha,\beta}$的顶点，那么 C 是 P 或 Q 的一个面，结论即得。否则，可得 $\tilde{C} = \text{conv}\{F_c^{\mathbf{0}} \cup G_c^{\alpha,\beta}\}$，因而 $C = \text{conv}\{F_c \cup G_c\}$。由引理 11 可得 $n = \dim C = \dim \tilde{C} = \dim F_c^{\mathbf{0}} + \dim G^{\alpha,\beta}_c + 1 = \dim F_c + \dim G_c + 1$。因此，$C$ 是 F_c 与 G_c 的连接。 □

命题 14. 对于 $P, Q \in \mathcal{Q}_k$，我们有 conv$(P \cup Q) \in \mathcal{Q}_{k+1}$。引理 15. 如果 $d := \dim(\text{conv}(P \cup Q)) \geq 2^{k+1}$，那么 Vol(conv$(P \cup Q)$) 是偶数。

引理 15 的证明。由命题 10 可知，conv$\{P \cup Q\}$ 可以细分，使得每个满维胞腔 C 都是 P 的一个面 F 与 Q的一个面 G 的连接。设 $i := \dim(F)$ 且 $j := \dim(G)$，则 $d = i + j + 1$。由引理 6 可得，对于某个 $z \in \mathbb{Z}$，Vol$(C) = z \cdot$Vol$(F) \cdot$Vol(G)。由于 d 至少为 2^{k+1}，因此 i 或 j 至少为 2^k。因此，Vol(F) 或 Vol(G) 必为偶数，因为 P 和 Q 包含于 $\mathcal{Q}_k$ 中。因此 Vol(C) 也是偶数。对细分中的所有胞腔都这样处理，可推出 Vol(conv$(P \cup Q)$) 是偶数。 □

命题 14 的证明。再次，为了证明该命题，我们需要确保不仅凸包 conv$(P \cup Q)$ 的归一化体积为偶数，而且它的每个维数至少为 2^{k+1} 的面也满足这一点。若 F 是这样的一个面，令 H 为包含 F 的超平面。注意，F 是 P 或 Q 中所有位于 H 内的顶点的凸包。特别地，$F = \text{conv}(P' \cup Q')$，其中 P' 和 Q' 分别是由超平面 H 定义的 P 和 Q 的面（可能为空）。由于 $\mathcal{Q}_k$ 对取面运算封闭，因此可得 P'、$Q' \in \mathcal{Q}_k$。于是，对 P' 和 Q' 应用引理 15，可得 F 的体积为偶数。对所有维数至少为 2^k 的面 F 依此处理，即完成证明。 □