Combinatorial Capacity of modReLU Complex Networks: VC-Dimension Bounds and Lower Limits

Anonymous authors
Paper under double-blind review

Abstract

Complex-valued neural networks (CVNNs) are increasingly used in settings where both magnitude and phase of the signal carry information; see, e.g., (10; 11; 13; 15). In particular, deep networks with the modReLU activation (18) have become standard in applications such as MRI reconstruction, radar, and complex-valued time-series modeling. While approximation properties of such networks have recently been analyzed in detail (6), their statistical capacity in the sense of VC-dimension has not, to the best of our knowledge, been studied.

In this paper we formalize a natural class of fully connected deep complex-valued networks with modReLU activation and real sign output. Via the standard identification $\mathbb{C}^d \cong \mathbb{R}^{2d}$, we view these models as binary classifiers on $\mathbb{R}^{2d}$. Let W denote the total number of real-valued trainable parameters, including the real and imaginary parts of all weights and biases as well as the real modReLU bias parameters. Using tools from real algebraic geometry and a VC-dimension bound for semi-algebraic concept classes due to Goldberg and Jerrum (8), together with quantitative bounds for quantifier elimination (4), we prove that for any architecture with W parameters and depth L, the VC-dimension of the corresponding hypothesis class is at most on the order of $W^2 \log W$, with a universal constant independent of the particular architecture.

On the other hand, by restricting to real inputs and parameters and exploiting results of Harvey, Liaw, and Mehrabian (9) and Bartlett et al. (3) on deep networks with piecewise-linear activations, we obtain lower bounds of order $WL \log(W/L)$ for suitable depth-L architectures within the modReLU class. Thus the VC-dimension of these networks grows at least linearly in both W and L, and at most quadratically in W up to a logarithmic factor. Closing this gap is an interesting open problem.

1 Introduction

In many signal-processing problems the natural data space is complex-valued. Examples include MRI, radar, wireless communication channels, and certain Fourier or wavelet representations of time series. For such problems, complex-valued neural networks (CVNNs) have been proposed as a more natural alternative to real networks; see, for instance, (10; 11; 12; 18; 13; 15) and the references therein. In these architectures all weights, biases, and activations are complex, and the network computes a map from $\mathbb{C}^d$ to $\mathbb{C}^k$.

Recently, there has been considerable progress on the approximation theory of CVNNs. In particular, Geuchen and Voigtlaender (6) established explicit quantitative error bounds for approximating smooth functions on compact subsets of $\mathbb{C}^d$ by complex-valued networks with the modReLU activation function (18), and showed that the rates are optimal up to logarithmic factors. Follow-up work has treated phenomena such as depth separation and universality for narrow complex networks (7; 12). A recent survey by Lee, Hasegawa, and Gao (13) gives a broad overview of complex activations, Wirtinger calculus, and optimization aspects.

In contrast, the statistical learning theory of CVNNs is still largely undeveloped. For real networks, a rich body of work has characterized the VC-dimension and pseudo-dimension of various architectures since the classical paper of Baum and Haussler (5) and the monograph of Anthony and Bartlett (1). For example, deep

ReLU networks with W real parameters and depth L are now known to have VC-dimension on the order of $WL \log W$, and this dependence is essentially tight (9; 3). Related results characterize the pseudo-dimension of real networks in regression settings (16; 20).

A natural question is whether these nearly tight results for real networks can be transferred to complex-valued architectures by using the standard identification $\mathbb{C}^d \cong \mathbb{R}^{2d}$ (sending $x+iy$ to (x, y) coordinate-wise). For networks with activations that remain piecewise-polynomial after this identification, one can indeed hope to import the classical region-counting and sign-pattern arguments. For modReLU networks, however, such a transfer is *not* available as a black box. The sharp $WL \log W$ analyses for deep real networks in (9; 3) proceed by exploiting that the network computes a *piecewise-polynomial* (in particular, piecewise-linear) map of the input, with a partition induced by activation gates; this structure enables counting the number of distinct affine regions and hence the number of labelings the network can realize. In contrast, when modReLU is written in real coordinates $z = x + iy$, its active branch contains the normalization factor $z/|z|$ with $|z| = \sqrt{x^2 + y^2}$, i.e., expressions involving division by $\sqrt{x^2 + y^2}$. Consequently, the real-coordinate representation of modReLU is not piecewise-polynomial in (x, y) in the sense required by these arguments, and the existing $\Theta(WL \log W)$ upper-bound proofs for piecewise-linear networks cannot be applied directly to modReLU architectures.

Our contribution is to show that, despite this obstruction, the decision regions of modReLU networks can still be controlled using *semi-algebraic* methods after a suitable lifting. By introducing auxiliary radius variables r satisfying $r^2 = x^2 + y^2$, the modReLU gate and the full forward pass admit an existential description by polynomial equalities and inequalities of bounded degree. Projecting out the auxiliary variables yields a semi-algebraic description of the decision set via Tarski–Seidenberg (quantifier elimination) (17; 4). This reduction allows us to invoke the general VC-dimension bound for semi-algebraic concept classes due to Goldberg and Jerrum (8), together with quantitative bounds for quantifier elimination (4), to obtain parameter-count-dependent VC-dimension upper bounds for modReLU complex networks. The price of using a generic projection/quantifier-elimination step is an additional factor in the resulting description complexity, which is reflected in the $O(W^2 \log W)$ bound proved in this work.

Beyond serving as a combinatorial complexity measure, the VC-dimension controls the growth function (via Sauer–Shelah type bounds) and yields classical sample-complexity and uniform-convergence guarantees for ERM-style learning (19; 1). Thus, bounding the VC-dimension of CVNNs provides a principled way to compare their statistical capacity to real-valued architectures and complements existing norm-based generalization analyses (21).

For complex-valued networks there are, to the best of our knowledge, no analogous capacity results in terms of VC-dimension. The only global complexity bounds we are aware of are recent norm-based generalization bounds that control the generalization gap via a spectral-complexity measure for CVNNs (21), which is a different notion from combinatorial capacity. The main purpose of this work is to begin closing this gap by providing VC-dimension bounds for a widely used complex-valued architecture family.

Contributions. At a high level, the contributions of this paper are as follows.

- We formalize a natural class of fully connected deep complex-valued networks with modReLU activation and real sign output. Via the standard identification $\mathbb{C}^d \cong \mathbb{R}^{2d}$, we view these models as binary classifiers on $\mathbb{R}^{2d}$.
- We show that the corresponding decision regions form a semi-algebraic family in (θ, x) whose description complexity—in terms of the number and degree of defining polynomials—can be bounded in terms of the architecture and the number W of real-valued trainable parameters.
- Using the VC-dimension bound for semi-algebraic concept classes due to Goldberg and Jerrum (8), combined with quantitative bounds on quantifier elimination (4), we derive an upper bound of order $W^2 \log W$ for the VC-dimension of this class.
- By specializing to real inputs and constraining parameters appropriately, we show that our class contains, as a subclass, real-valued networks with a fixed piecewise-linear activation. Results of Harvey,

Liaw, and Mehrabian (9) and Bartlett et al. (3) then imply lower bounds of order $WL\log(W/L)$ for suitable architectures, showing that the VC-dimension must grow at least linearly in both W and L.

- We discuss how the same approach could be extended to other complex activations and to pseudo-dimension, and outline several open problems, including whether the $W^2 \log W$ upper bound can be sharpened towards $WL\log W$ in the complex-valued setting.

2 Preliminaries

We collect notation and background results used in the sequel. Standard references for VC-dimension and statistical learning theory include (19; 1; 20).

Notation. We write $\Re(z)$ and $\Im(z)$ for the real and imaginary parts of $z \in \mathbb{C}$. We identify $\mathbb{C}^d$ with $\mathbb{R}^{2d}$ via the bijection $\Psi : \mathbb{R}^{2d} \to \mathbb{C}^d$ defined by

$$\Psi(a_1, b_1, \ldots, a_d, b_d) := (a_1 + ib_1, \ldots, a_d + ib_d).$$

For binary classification we use $\mathrm{sign} : \mathbb{R} \to \{+1, -1\}$ with the convention $\mathrm{sign}(t) = +1$ if $t \geq 0$ and $\mathrm{sign}(t) = -1$ if $t < 0$. A network of depth L has $L-1$ hidden layers and one output layer; we index layers by $\ell = 0, \ldots, L$ with $n_0 = d$ and $n_L = 1$.

VC-dimension. Let $\mathcal{H} \subseteq \{+1, -1\}^{\mathcal{X}}$ be a class of binary classifiers on a domain $\mathcal{X}$. A finite set $\{x_1, \ldots, x_m\} \subseteq \mathcal{X}$ is *shattered* by $\mathcal{H}$ if for every labeling $(y_1, \ldots, y_m) \in \{+1, -1\}^m$ there exists $h \in \mathcal{H}$ such that $h(x_i) = y_i$ for all i. The *VC-dimension* $\mathrm{VCdim}(\mathcal{H})$ is the largest m such that some set of size m is shattered (or ∞ if no such largest m exists) (19; 1).

Semi-algebraic sets. A set $S \subset \mathbb{R}^n$ is called *semi-algebraic* if it can be expressed as a finite Boolean combination of polynomial equalities and inequalities in the coordinates $x_1, \ldots, x_n$. A family $\{S_\theta\}_{\theta \in \Theta}$ with $\Theta \subseteq \mathbb{R}^W$ is called semi-algebraic if

$$\{(\theta, x) \in \Theta \times \mathbb{R}^n : x \in S_\theta\}$$

is a semi-algebraic subset of $\mathbb{R}^{W+n}$. A standard closure result is the Tarski–Seidenberg theorem: the image (projection) of a semi-algebraic set under a polynomial map is semi-algebraic (17; 4).

We will use the following general bound on the VC-dimension of semi-algebraic families due to Goldberg and Jerrum (8) (see also (1, Theorem 8.3) and (4, Theorem 7.38)).

Theorem 1 (Goldberg–Jerrum (8)) *Let $F = \{S_\theta\}_{\theta \in \Theta}$ be a family of subsets of $\mathbb{R}^n$ parameterized by $\Theta \subset \mathbb{R}^W$. Suppose there exist integers $s \geq 1$ and $d \geq 1$ and polynomials $p_1, \ldots, p_s \in \mathbb{R}[U_1, \ldots, U_W, X_1, \ldots, X_n]$ of degree at most d such that for all $\theta \in \Theta$ and $x \in \mathbb{R}^n$,*

$$x \in S_\theta \iff \Phi\big(\mathrm{sign}(p_1(\theta, x)), \ldots, \mathrm{sign}(p_s(\theta, x))\big) = \mathrm{true},$$

for some fixed Boolean formula Φ. Then there exists a universal constant $C > 0$ such that

$$\mathrm{VCdim}(F) \;\leq\; C\, W\, \log(s\, d).$$

Complex-valued neural networks and modReLU. Let $d \geq 1$ and $L \geq 1$ be integers. A fully connected feedforward complex-valued network (CVNN) of depth L and input dimension d consists of:

- layer widths $n_0 = d,\ n_1,\ \ldots,\ n_{L-1},\ n_L = 1$,
- complex weight matrices $W^{(\ell)} \in \mathbb{C}^{n_\ell \times n_{\ell-1}}$ and complex biases $b^{(\ell)} \in \mathbb{C}^{n_\ell}$ for $\ell = 1, \ldots, L$,
- a real modReLU bias parameter $\beta_j^{(\ell)} \in \mathbb{R}$ for each hidden unit (ℓ, j) with $\ell = 1, \ldots, L-1$.

Given an input $z^{(0)} = x \in \mathbb{C}^d$, the network computes recursively for $\ell = 1, \ldots, L-1$:

$$u^{(\ell)} \;=\; W^{(\ell)} z^{(\ell-1)} + b^{(\ell)} \;\in\; \mathbb{C}^{n_\ell}, \qquad z_j^{(\ell)} \;=\; \sigma_{\mathrm{modR}}\big(u_j^{(\ell)}; \beta_j^{(\ell)}\big), \quad j = 1, \ldots, n_\ell,$$

and the (complex-valued) output is

$$F_\theta(x) \;=\; W^{(L)} z^{(L-1)} + b^{(L)} \;\in\; \mathbb{C},$$

where θ denotes the collection of all trainable real parameters (real and imaginary parts of all weights and biases, together with all $\beta_j^{(\ell)}$). The total number of trainable real parameters is therefore

$$W \;=\; 2\sum_{\ell=1}^{L} n_\ell n_{\ell-1} \;+\; 2\sum_{\ell=1}^{L} n_\ell \;+\; \sum_{\ell=1}^{L-1} n_\ell.$$

We work with the *modReLU* activation introduced by Trabelsi et al. (18) and further analyzed by Geuchen and Voigtlaender (6). For $\beta \in \mathbb{R}$ and $z \in \mathbb{C}$, define

$$\sigma_{\mathrm{modR}}(z; \beta) \;:=\; \begin{cases} 0, & z = 0, \\ 0, & |z| + \beta \le 0, \\ (|z| + \beta)\,\dfrac{z}{|z|}, & |z| + \beta > 0, \end{cases} \tag{1}$$

i.e., $\sigma_{\mathrm{modR}}(z; \beta) = \mathrm{ReLU}(|z| + \beta) \cdot \dfrac{z}{|z|}$ with the convention $\dfrac{z}{|z|} = 0$ at $z = 0$ (18). Equivalently, writing $z = re^{i\vartheta}$ with $r = |z|$ and $\vartheta = \arg(z)$ (defined for $r > 0$), one has $\sigma_{\mathrm{modR}}(z; \beta) = 0$ for $r = 0$ and $\sigma_{\mathrm{modR}}(z; \beta) = (r + \beta)e^{i\vartheta}$ on the active region $\{r > 0,\ r + \beta > 0\}$.

For binary classification we use the sign of the real part:

$$h_\theta(x) \;=\; \mathrm{sign}\big(\Re(F_\theta(x))\big) \;\in\; \{+1, -1\}. \tag{2}$$

Via the identification $\Psi : \mathbb{R}^{2d} \to \mathbb{C}^d$, we view h_θ as a classifier on $\mathbb{R}^{2d}$ by setting, for $\tilde{x} \in \mathbb{R}^{2d}$,

$$h_\theta(\tilde{x}) \;=\; \mathrm{sign}\big(\Re(F_\theta(\Psi(\tilde{x})))\big). \tag{3}$$

Let $\mathcal{H}_{W,L}^{\mathrm{modR}}$ denote the set of all classifiers of the form (3) obtained by varying $\theta \in \mathbb{R}^W$ for a fixed architecture $(n_0, \ldots, n_L)$.

3 VC-dimension upper bound

Theorem 2 (VC-dimension of modReLU networks) *There exists a universal constant $C > 0$ such that for every choice of layer widths $(n_0, \ldots, n_L)$, letting $W \ge 2$ denote the total number of* trainable real parameters *(including real/imaginary parts of all weights and biases and all modReLU bias parameters), the VC-dimension of $\mathcal{H}_{W,L}^{\mathrm{modR}}$ satisfies*

$$\mathrm{VCdim}\big(\mathcal{H}_{W,L}^{\mathrm{modR}}\big) \;\le\; C\, W^2 \log W.$$

In particular, for any fixed depth L, $\mathrm{VCdim}(\mathcal{H}_{W,L}^{\mathrm{modR}}) = O(W^2 \log W)$ uniformly over architectures with W parameters.

3.1 Semi-algebraic representation of the decision map

Fix an architecture $(n_0, \ldots, n_L)$ and write $\tilde{x} \in \mathbb{R}^{2d}$ for the real input and $\theta \in \mathbb{R}^W$ for the trainable parameters. Recall our sign convention $\mathrm{sign}(t) = +1$ for $t \ge 0$ and -1 for $t < 0$, so

$$h_\theta(\tilde{x}) = +1 \quad \Longleftrightarrow \quad \Re(F_\theta(\Psi(\tilde{x}))) \ge 0.$$

Auxiliary variables and constraints. We introduce auxiliary real variables encoding a forward pass. For each hidden layer $\ell = 1, \ldots, L-1$ and unit $j = 1, \ldots, n_\ell$, we introduce

$$u_{j,R}^{(\ell)},\, u_{j,I}^{(\ell)} \quad \text{(pre-activation)}, \qquad z_{j,R}^{(\ell)},\, z_{j,I}^{(\ell)} \quad \text{(post-activation)}, \qquad r_j^{(\ell)} \quad \text{(radius)}.$$

(We treat $z^{(0)} = \Psi(\tilde{x})$ as fixed by $\tilde{x}$.) Let Y denote the full collection of these auxiliary variables. Then

$$|Y| \;=\; 5 \sum_{\ell=1}^{L-1} n_\ell.$$

Moreover, since $n_{\ell-1} \geq 1$ for all ℓ and the parameter count includes the real and imaginary parts of all complex weights, we have

$$W \;\geq\; 2 \sum_{\ell=1}^{L-1} n_\ell n_{\ell-1} \;\geq\; 2 \sum_{\ell=1}^{L-1} n_\ell,$$

and hence $|Y| \leq \frac{5}{2} W$, i.e., $|Y| = O(W)$ with an absolute constant.

We impose the following constraints in the variables $(\theta, \tilde{x}, Y)$. Each atomic predicate is a polynomial equality or inequality of degree at most 2. The only non-affine ingredient is the modReLU nonlinearity: it is captured by introducing the radius variables $r_j^{(\ell)}$ and enforcing a three-branch gate (zero/inactive/active), including the $u_j^{(\ell)} = 0$ case required by the standard convention $\sigma_{\text{modR}}(0; \beta) = 0$ (18).

- *Complex affine maps.* For each $\ell = 1, \ldots, L-1$ and $j = 1, \ldots, n_\ell$, we enforce the real and imaginary parts of

  $$u_j^{(\ell)} \;=\; \sum_{k=1}^{n_{\ell-1}} W_{jk}^{(\ell)} z_k^{(\ell-1)} + b_j^{(\ell)}$$

  by two polynomial equalities. Written in real coordinates, these constraints are bilinear in the parameter variables (real/imaginary parts of $W_{jk}^{(\ell)}$ and $b_j^{(\ell)}$) and the activation variables (real/imaginary parts of $z_k^{(\ell-1)}$), hence have degree 2.

- *modReLU gate (including the $u = 0$ case).* Fix a hidden unit (ℓ, j) with bias parameter $\beta_j^{(\ell)} \in \mathbb{R}$ (a component of θ). We first enforce the radius relation

  $$r_j^{(\ell)} \geq 0, \qquad (r_j^{(\ell)})^2 = (u_{j,R}^{(\ell)})^2 + (u_{j,I}^{(\ell)})^2.$$

  We then encode $z_j^{(\ell)} = \sigma_{\text{modR}}(u_j^{(\ell)}; \beta_j^{(\ell)})$ (with the convention $\sigma_{\text{modR}}(0; \beta) = 0$, cf. (1) and (18)) by the disjunction of the following polynomial conditions:

  $$\begin{aligned}
  &\text{(zero)}\; r_j^{(\ell)} = 0, \quad z_{j,R}^{(\ell)} = 0, \quad z_{j,I}^{(\ell)} = 0;\\
  &\text{(inactive)}\; r_j^{(\ell)} > 0, \quad r_j^{(\ell)} + \beta_j^{(\ell)} \leq 0, \quad z_{j,R}^{(\ell)} = 0, \quad z_{j,I}^{(\ell)} = 0;\\
  &\text{(active)}\; r_j^{(\ell)} > 0, \quad r_j^{(\ell)} + \beta_j^{(\ell)} \geq 0,\\
  &\qquad z_{j,R}^{(\ell)}\, r_j^{(\ell)} = (r_j^{(\ell)} + \beta_j^{(\ell)})\, u_{j,R}^{(\ell)}, \quad z_{j,I}^{(\ell)}\, r_j^{(\ell)} = (r_j^{(\ell)} + \beta_j^{(\ell)})\, u_{j,I}^{(\ell)}.
  \end{aligned}$$

  The "zero" branch is essential: when $u_j^{(\ell)} = 0$ (hence $r_j^{(\ell)} = 0$), modReLU outputs 0 even if $\beta_j^{(\ell)} > 0$ by the standard convention in (18). Introducing this branch ensures the existential encoding matches the actual network computation for all $(u_j^{(\ell)}, \beta_j^{(\ell)})$.

- *Output and decision.* Write the output as $F_\theta(x) = W^{(L)} z^{(L-1)} + b^{(L)}$ and enforce the decision event by the single polynomial inequality

  $$\Re(F_\theta(\Psi(\tilde{x}))) \geq 0,$$

  again expressed in real coordinates (degree ≤ 2 in $(\theta, \tilde{x}, Y)$).

The total number of atomic predicates across all layers is $s_0 = O(W)$, and the maximum degree is $\delta = 2$.

From constraints to a quantifier-free description. Define

$$G \;:=\; \{(\theta, \tilde{x}) \in \mathbb{R}^{W+2d} :\; h_\theta(\tilde{x}) = +1\}.$$

Let $\widetilde{G} \subset \mathbb{R}^{W+2d+|Y|}$ be the set of all $(\theta, \tilde{x}, Y)$ satisfying the constraints above. Then $\widetilde{G}$ is semi-algebraic by definition, and

$$(\theta, \tilde{x}) \in G \quad \Longleftrightarrow \quad \exists\, Y \in \mathbb{R}^{|Y|} \;\text{ such that }\; (\theta, \tilde{x}, Y) \in \widetilde{G}.$$

Hence G is the projection of the semi-algebraic set $\widetilde{G}$. By the Tarski–Seidenberg theorem (quantifier elimination for real closed fields), G is itself semi-algebraic (17; 4).

Moreover, quantitative quantifier-elimination bounds apply because the defining formula for $\widetilde{G}$ has a single existential block, uses at most $s_0 = O(W)$ input polynomials of degree $\leq \delta = 2$, and ranges over

$$k_0 := W + 2d \quad \text{free variables} \qquad \text{and} \qquad k_1 := |Y| = O(W) \text{ quantified variables.}$$

In particular, by (4, Theorem 14.16) (see also the discussion of singly-exponential bounds for one-block elimination therein), there exist absolute constants $C_1, C_2 > 0$ and an equivalent quantifier-free formula defining G that involves at most

$$s \;\leq\; (s_0\, \delta)^{C_1(k_0+k_1)} \qquad \text{polynomials of degree at most} \qquad d \;\leq\; \delta^{C_2(k_0+k_1)}.$$

Since $n_L = 1$, any fully connected architecture satisfies $W \geq 2d+2$ (already for $L = 1$ one has $W = 2d+2$), so $k_0 + k_1 = O(W)$ and $\log(sd) = O(W \log W)$.

Lemma 1 (Semi-algebraic decision map) *There exist integers $s \geq 1$, $d \geq 1$ and polynomials $p_1, \ldots, p_s \in \mathbb{R}[U_1, \ldots, U_W, X_1, \ldots, X_{2d}]$ of degree at most d such that for all $\theta \in \mathbb{R}^W$ and $\tilde{x} \in \mathbb{R}^{2d}$,*

$$h_\theta(\tilde{x}) = +1 \iff \Phi\big(\mathrm{sign}(p_1(\theta, \tilde{x})), \ldots, \mathrm{sign}(p_s(\theta, \tilde{x}))\big) = \mathrm{true},$$

for a fixed Boolean formula Φ independent of $\theta, \tilde{x}$. Furthermore, one may choose s, d so that for absolute constants $c_0, c_1 > 0$,

$$s \;\leq\; (c_0 W)^{c_1 W} \qquad \textit{and} \qquad d \;\leq\; 2^{c_1 W}.$$

In particular, $\log(sd) = O(W \log W)$ uniformly over architectures with W parameters.

Proof. By construction, $\widetilde{G} \subset \mathbb{R}^{W+2d+|Y|}$ is semi-algebraic and is defined by a quantifier-free formula that uses $s_0 = O(W)$ polynomial (in)equalities of degree at most $\delta = 2$ (including the explicit "zero" branch for the modReLU gate, ensuring correctness when $u_j^{(\ell)} = 0$ and $\beta_j^{(\ell)} > 0$). The set G is the projection of $\widetilde{G}$ onto the $(\theta, \tilde{x})$-coordinates, hence is semi-algebraic by Tarski–Seidenberg (17; 4).

Applying the quantitative one-block quantifier elimination bound (4, Theorem 14.16) with $k_0 = W + 2d$ free variables, $k_1 = |Y|$ quantified variables, s_0 input polynomials, and degree $\delta = 2$, we obtain an equivalent quantifier-free formula for G involving at most $s \leq (s_0\delta)^{C_1(k_0+k_1)}$ polynomials of degree at most $d \leq \delta^{C_2(k_0+k_1)}$.

Finally, any quantifier-free Boolean combination of atomic comparisons $q(\theta, \tilde{x}) \bowtie 0$ with $\bowtie \in \{<, \leq, =, \geq, >\}$ can be rewritten as a Boolean formula in the values $\mathrm{sign}(q(\theta, \tilde{x}))$ and $\mathrm{sign}(-q(\theta, \tilde{x}))$ under our two-valued convention (e.g., $q > 0$ is equivalent to $q \geq 0$ and not $(q = 0)$, and $q = 0$ is equivalent to $q \geq 0$ and $-q \geq 0$). This transformation increases the number of polynomials by at most a factor of 2 and does not increase degrees, yielding the representation in terms of $\mathrm{sign}(p_i)$ as stated.

Since $W \geq 2d + 2$ and $|Y| = O(W)$, we have $k_0 + k_1 = O(W)$ and $s_0 = O(W)$, hence $s \leq (c_0 W)^{c_1 W}$ and $d \leq 2^{c_1 W}$ for absolute constants $c_0, c_1 > 0$, and therefore $\log(sd) = O(W \log W)$. □

3.2 Proof of the VC-dimension bound

For each $\theta \in \mathbb{R}^W$, let

$$S_\theta := \{\tilde{x} \in \mathbb{R}^{2d} :\; h_\theta(\tilde{x}) = +1\}.$$

By Lemma 1, there exist integers $s \geq 1$, $d \geq 1$ and polynomials $p_1, \ldots, p_s \in \mathbb{R}[U_1, \ldots, U_W, X_1, \ldots, X_{2d}]$ of degree at most d, together with a fixed Boolean formula Φ, such that for all $\theta \in \mathbb{R}^W$ and $\tilde{x} \in \mathbb{R}^{2d}$,

$$\tilde{x} \in S_\theta \quad \Longleftrightarrow \quad \Phi\big(\mathrm{sign}(p_1(\theta, \tilde{x})), \ldots, \mathrm{sign}(p_s(\theta, \tilde{x}))\big) = \mathrm{true}.$$

Thus the family $\{S_\theta\}_{\theta \in \mathbb{R}^W}$ is a semi-algebraic concept class of the form required by the Goldberg–Jerrum theorem (8) (see also (1, Theorem 8.3) and (4, Theorem 7.38)). Applying Theorem 1 with parameter dimension W yields

$$\mathrm{VCdim}(\mathcal{H}^{\mathrm{modR}}_{W,L}) \;=\; \mathrm{VCdim}(\{S_\theta\}_{\theta \in \mathbb{R}^W}) \;\leq\; C\,W\,\log(sd),$$

for a universal constant $C > 0$.

It remains to bound $\log(sd)$ in terms of W. Lemma 1 provides such a bound uniformly over architectures: the quantifier-elimination estimate used there depends on the number of free variables $k_0 = W + 2d$ and quantified variables $k_1 = |Y|$, but these are both $O(W)$ for the present class. Indeed, since $n_L = 1$, even in the shallow case $L = 1$ we have

$$W \;=\; 2n_1 n_0 + 2n_1 \;=\; 2d + 2,$$

and for $L \geq 2$ the parameter count only increases. Hence for all architectures considered here,

$$W \;\geq\; 2d + 2, \tag{4}$$

so in particular $2d \leq W$ and $k_0 = W + 2d \leq 2W$. Together with $|Y| = O(W)$ (as shown in the construction of Lemma 1), we obtain $k_0 + k_1 = O(W)$, and Lemma 1 yields $\log(sd) = O(W \log W)$ with absolute constants (via the quantitative one-block elimination bounds of (4, Theorem 14.16)).

Therefore,

$$\mathrm{VCdim}(\mathcal{H}^{\mathrm{modR}}_{W,L}) \;\leq\; C\,W\,\log(sd) \;\leq\; C'\,W^2 \log W$$

for a universal constant $C' > 0$, completing the proof of Theorem 2. □

Comparison with classical real-network VC-dimension proofs. For deep real networks with piecewise-linear activations (e.g., ReLU), the nearly-tight $\Theta(WL \log W)$ bounds are obtained by counting activation sign patterns/affine regions layer-by-layer and controlling how many distinct dichotomies the network can implement (9; 3). For modReLU networks, the written in real coordinates $z = x + iy$ introduces normalization by $|z| = \sqrt{x^2 + y^2}$, so the resulting map is not piecewise-polynomial in (x, y) in the sense exploited by those region-counting arguments. Our proof instead proceeds by lifting the computation to a polynomial constraint system (Section 3, Lemma 1) and then projecting out auxiliary variables using quantifier elimination (17; 4). The quantitative blow-up incurred by this elimination step is reflected in the bound $\log(sd) = O(W \log W)$ and is precisely what introduces the additional factor of W in the upper bound $O(W^2 \log W)$.

3.3 Lower bounds and comparison

The upper bound of Theorem 2 is obtained via a semi-algebraic encoding and is not expected to be tight. In the opposite direction, we obtain a lower bound by observing that $\mathcal{H}^{\mathrm{modR}}_{W,L}$ contains, as a parameter-restricted subclass, standard real-valued depth-L networks with a fixed piecewise-linear activation. The result below is therefore an *inherited corollary* of the nearly-tight VC-dimension lower bounds for real networks proved in (9; 3).

Proposition 1 (Lower bound (inherited from real networks)) *There exists a universal constant $c > 0$ such that the following holds. For all integers $L \geq 2$ and all W sufficiently large compared to L, there exists an architecture of depth L with at most W trainable real parameters (counting real and imaginary parts of weights and biases, as well as all modReLU bias parameters) for which*

$$\mathrm{VCdim}\big(\mathcal{H}^{\mathrm{modR}}_{W,L}\big) \;\geq\; c\,W\,L \log\Big(\frac{W}{L}\Big).$$

In particular, for suitable depth-L architectures, the VC-dimension of modReLU complex networks grows at least on the order of $W\,L \log(W/L)$.

Proof (via a real-valued subclass). Fix the non-degenerate constant $\beta_0 := -1$ and consider the restriction of modReLU to the real line:

$$\phi_{\beta_0}(x) \;:=\; \Re\big(\sigma_{\mathrm{modR}}(x + i0; \beta_0)\big), \qquad x \in \mathbb{R}.$$

By the definition (1) and the convention $\sigma_{\mathrm{modR}}(0; \beta_0) = 0$ (as in (18)), ϕ_{β_0} is piecewise-linear on $\mathbb{R}$ with a constant number of linear pieces (indeed, at most three). Hence the real-network lower bounds for fixed piecewise-linear activations apply: by (9; 3), there exists a universal constant $c_0 > 0$ such that for all L and all P sufficiently large compared to L, there exists a depth-L *real* network class $\mathcal{H}^{\mathrm{real}}_{P,L}(\phi_{\beta_0})$ with at most P trainable real parameters and activation ϕ_{β_0} satisfying

$$\mathrm{VCdim}\big(\mathcal{H}^{\mathrm{real}}_{P,L}(\phi_{\beta_0})\big) \;\geq\; c_0\, PL \log\Big(\frac{P}{L}\Big). \tag{5}$$

We now embed such a real network as a subclass of our complex modReLU networks. Consider the complex-valued architecture with the same layer widths and impose the restrictions:

- inputs are restricted to $\mathbb{R}^d \subset \mathbb{C}^d$ (identified with $x + i0$),
- all complex weights and biases are constrained to be real (imaginary parts fixed to 0),
- all modReLU bias parameters are fixed to the constant value β_0 (i.e., $\beta_j^{(\ell)} = \beta_0$ for all hidden units).

Under these restrictions, the complex network computes exactly the same real-valued function on $\mathbb{R}^d$ as the corresponding real network with activation ϕ_{β_0}, and the induced classifier $\tilde{x} \mapsto \mathrm{sign}(\Re(F_\theta(\Psi(\tilde{x}))))$ coincides with the real-network classifier $\mathrm{sign}(f(x))$ on inputs $x \in \mathbb{R}^d$.

Let P denote the number of *free* real parameters (weights and biases) in the embedded real network. Consider the corresponding complex architecture (same widths) used to realize this subclass. In the ambient count W for complex modReLU networks we include: (i) both real and imaginary parts of all complex weights and biases, and (ii) all modReLU bias parameters. Under the restriction "imaginary parts = 0" and "$\beta_j^{(\ell)} = \beta_0$", the number of free parameters is precisely the number of real weights and biases, namely P. Since each such parameter corresponds to a complex weight/bias coordinate that contributes two real degrees of freedom in the ambient parameterization, we have

$$P \;\leq\; \frac{W}{2}. \tag{6}$$

(Indeed, for this fixed architecture the ambient parameter count can be written as $W = 2P + P_\beta$, where $P_\beta = \sum_{\ell=1}^{L-1} n_\ell$ counts the modReLU bias parameters; in particular $W \geq 2P$.)

Consequently, $\mathcal{H}^{\mathrm{modR}}_{W,L}$ contains a subclass with VC-dimension at least that of $\mathcal{H}^{\mathrm{real}}_{P,L}(\phi_{\beta_0})$, and by (5) we obtain

$$\mathrm{VCdim}\big(\mathcal{H}^{\mathrm{modR}}_{W,L}\big) \;\geq\; \mathrm{VCdim}\big(\mathcal{H}^{\mathrm{real}}_{P,L}(\phi_{\beta_0})\big) \;\geq\; c_0\, PL \log\Big(\frac{P}{L}\Big).$$

Finally, we choose an architecture whose total parameter count is at most W and for which the embedded real subclass has P on the order of W up to constants. Concretely, take any depth-L fully connected real architecture achieving (5) with $P = \lfloor W/2 \rfloor$ trainable parameters. For the corresponding complex architecture (same widths), the ambient parameter count satisfies $W_{\mathrm{arch}} = 2P + P_\beta \geq 2P$ and hence $W_{\mathrm{arch}} \leq W$ after choosing the widths so that $W_{\mathrm{arch}} \leq W$ (equivalently, by selecting a realization that fits within the budget W). Substituting $P = \lfloor W/2 \rfloor$ gives

$$\mathrm{VCdim}\big(\mathcal{H}^{\mathrm{modR}}_{W,L}\big) \;\geq\; c_0 \Big\lfloor \frac{W}{2} \Big\rfloor L \log\Big(\frac{\lfloor W/2 \rfloor}{L}\Big) \;\geq\; c\, WL \log\Big(\frac{W}{L}\Big),$$

for a (possibly smaller) universal constant $c > 0$, provided W is sufficiently large compared to L. This proves the proposition. □

Combining Proposition 1 with Theorem 2, we obtain for suitable depth-L architectures and large W:

$$c\,W\,L\log\Big(\frac{W}{L}\Big) \;\lesssim\; \mathrm{VCdim}\big(\mathcal{H}^{\mathrm{modR}}_{W,L}\big) \;\lesssim\; W^2\log W.$$

Thus the VC-dimension grows at least linearly in both W and L, and at most quadratically in W up to a logarithmic factor. Narrowing this gap—for instance by sharpening the upper bound towards the $WL\log W$ behavior known for real piecewise-linear networks (9; 3)—is left as an open problem.

4 Discussion and related work

We have shown that the VC-dimension of fully connected complex-valued networks with modReLU activation and W trainable real parameters is at most on the order of $W^2\log W$, uniformly over architectures of a given size. The proof has two main ingredients: (i) after introducing auxiliary variables for complex magnitudes, the network computation and the decision event $\Re(F_\theta(\Psi(\tilde{x}))) \geq 0$ admit an existential semi-algebraic description, and (ii) semi-algebraic concept classes with bounded description complexity have VC-dimension bounded by $O(W\log(sd))$ via the Goldberg–Jerrum theorem (8) (see also (1)). Quantitative bounds for quantifier elimination (4; 17) then control $\log(sd)$ in terms of W, yielding the stated $O(W^2\log W)$ upper bound.

Special cases (shallow and linear models). It is instructive to consider the extreme shallow case $L = 1$, where the network reduces to a single complex affine functional with sign output,

$$h_\theta(\tilde{x}) \;=\; \mathrm{sign}\big(\Re(\langle w, \Psi(\tilde{x})\rangle + b)\big), \qquad w \in \mathbb{C}^d,\; b \in \mathbb{C}.$$

Under the identification $\mathbb{C}^d \cong \mathbb{R}^{2d}$, the map $\tilde{x} \mapsto \Re(\langle w, \Psi(\tilde{x})\rangle + b)$ is a real affine functional on $\mathbb{R}^{2d}$, and therefore $\mathcal{H}^{\mathrm{modR}}_{W,1}$ is a class of halfspaces in $\mathbb{R}^{2d}$. In particular, its VC-dimension is $2d+1$, which is linear in the ambient parameter count $W = 2d+2$ (1; 19). More generally, if all activations are linear (so the network is globally affine in the input), then regardless of depth the induced classifier is again a halfspace in $\mathbb{R}^{2d}$, and the VC-dimension is linear in the effective input dimension (and hence linear in W for fully connected architectures).

Relation to nearly-tight bounds for real networks. For real-valued networks, sharp VC-dimension and pseudo-dimension bounds are known for broad classes of piecewise-linear architectures: in particular, the nearly-tight $\Theta(WL\log W)$ behavior for depth-L networks with W parameters is established in (9; 3). Our lower bound is inherited by embedding such real networks as a parameter-restricted subclass of modReLU complex networks (Section 3, Proposition 1). In contrast, our upper bound is *not* a black-box consequence of (9; 3). The upper-bound techniques in (9; 3) rely on the fact that, with piecewise-linear (more generally, piecewise-polynomial) activations having a constant number of pieces, the network computes a piecewise-polynomial function of the input on a partition induced by activation gates; this enables a layer-by-layer counting of regions/sign patterns without invoking general projection or quantifier-elimination machinery. After identifying $\mathbb{C}^d \cong \mathbb{R}^{2d}$, modReLU does not fit into this framework: writing $z = x+iy$ and $r = \sqrt{x^2+y^2}$, the active branch contains the normalization $z/|z| = (x+iy)/r$ and hence factors of $1/\sqrt{x^2+y^2}$, so the resulting real-coordinate map is not piecewise-polynomial in (x, y) in the sense required by these arguments. For this reason, the existing $\Theta(WL\log W)$ proofs for piecewise-linear networks cannot be applied directly to obtain comparable upper bounds for modReLU complex networks. Our approach instead proceeds by lifting the computation to an existential system of polynomial constraints (introducing auxiliary variables r with $r^2 = x^2+y^2$), and then projecting out these auxiliary variables using Tarski–Seidenberg/quantifier elimination (17; 4). Combined with the Goldberg–Jerrum bound for semi-algebraic concept classes (8), this yields our $O(W^2\log W)$ VC-dimension upper bound.

Can complex networks exceed real-valued capacity? A complex-valued network induces a real-valued classifier on $\mathbb{R}^{2d}$ via the identification $\mathbb{C}^d \simeq \mathbb{R}^{2d}$, and each complex affine map $u = Wz+b$ with $W = A+iB$ can be written as a real affine map with the structured block matrix $\left[\begin{smallmatrix} A & -B \\ B & A \end{smallmatrix}\right]$ acting on $(\Re z, \Im z)$. Thus, for a fixed *real* parameter budget W, the modReLU hypothesis class considered here is (after identification) a *subclass* of real networks with W free real parameters but additional algebraic constraints (weight-tying

induced by complex arithmetic). In this sense, when capacity is measured as a function of W, complex networks cannot have larger VC-dimension than an unconstrained real network family of comparable size; any advantage of complex representations is better viewed as *parameter-efficiency* or inductive bias rather than strictly larger worst-case combinatorial capacity. Although the decision uses $\Re(F_\theta)$, the imaginary part still affects hidden-layer activations and can change $\Re(F_\theta)$, so this does not reduce the model to a purely real network. Finally, the same semi-algebraic lifting applies to other real decision statistics of a complex output (e.g., $\mathrm{sign}(\Im(F_\theta))$, thresholding $|F_\theta|$, or phase regions), and a systematic comparison of such variants to classical real architectures remains an open direction.

Why sharpening the upper bound is technically difficult. The gap between our $O(W^2 \log W)$ upper bound and the $\Theta(WL \log W)$ behavior known for real piecewise-linear networks arises from the use of generic projection/quantifier-elimination bounds. For ReLU-type activations, one can analyze the network layer-by-layer and count the number of activation sign patterns (or affine regions) induced by hyperplane arrangements, leading to sharp control of the shatter function and VC-dimension (9; 3). For modReLU, the normalization $z/|z|$ introduces algebraic structure (involving $\sqrt{x^2+y^2}$) that obstructs a direct reduction to counting finitely many affine regions in the (x, y)-plane. Our proof repairs this by lifting to an existential semi-algebraic description using auxiliary radius variables and then projecting them out using quantifier elimination (Tarski–Seidenberg) (17; 4). Quantitative elimination is necessarily worst-case and can introduce a blow-up in the number and degrees of polynomials describing the projected set; in our argument this manifests as the bound $\log(sd) = O(W \log W)$ (Lemma 1), which, when combined with Goldberg–Jerrum (8), yields the additional factor of W in $O(W^2 \log W)$. Improving the upper bound towards $O(WL \log W)$ would therefore require bypassing generic quantifier elimination and exploiting finer compositional structure of modReLU and complex multiplication to control the number of distinct dichotomies more directly.

Consequences of bounded VC-dimension. A bound on $\mathrm{VCdim}(\mathcal{H}^{\mathrm{modR}}_{W,L})$ controls classical combinatorial quantities such as the growth function (shatter coefficient), via Sauer–Shelah type inequalities (19; 1). It also yields standard distribution-free sample-complexity guarantees for uniform convergence and ERM-style learning: informally, achieving small generalization error requires a number of samples scaling on the order of $\mathrm{VCdim}(\mathcal{H}^{\mathrm{modR}}_{W,L})$ up to logarithmic factors (19; 1; 20). While such worst-case guarantees are typically loose for modern overparameterized training, they provide a principled capacity baseline for comparing complex-valued architectures to their real-valued counterparts.

Outlook. A natural open problem is to sharpen the $O(W^2 \log W)$ upper bound towards the $O(WL \log W)$ behavior known for real piecewise-linear networks (9; 3), potentially by avoiding generic quantifier-elimination bounds and instead exploiting finer structure of modReLU and complex multiplication. Another direction is to extend the semi-algebraic approach to other complex activations beyond modReLU, and to other architectures such as convolutional, recurrent, or attention-based complex-valued models, where the interaction of weight sharing, complex arithmetic, and activation structure may lead to different capacity scalings. Finally, it would be interesting to compare VC-type bounds to norm-based generalization analyses for CVNNs, such as spectral-complexity bounds (21), and to study pseudo-dimension for complex-valued regression settings (1; 20).

Limitations. Our analysis is restricted to fully connected feedforward architectures with modReLU activation, and we do not track constants or lower-order terms in the VC-dimension bounds. The semi-algebraic encoding relies on quantifier elimination (4; 17), which is computationally heavy; here it is used purely as a theoretical tool. The resulting $O(W^2 \log W)$ upper bound is likely not optimal, as suggested by the nearly-tight $O(WL \log W)$ bounds for real piecewise-linear networks (9; 3). We also do not address data-dependent or norm-based complexity measures for complex-valued networks, nor do we consider convolutional, recurrent, or attention-based architectures.

Reproducibility Statement

This is a theoretical paper. All claims are stated as formal theorems, lemmas, or propositions, with proofs provided in the main text or as proof sketches that rely on standard results from real algebraic geometry (4; 17) and statistical learning theory (19; 1). No datasets, code, or hyperparameters were used.

Broader Impact Statement. This work is purely theoretical and studies the combinatorial capacity of a class of complex-valued neural networks. We do not anticipate direct negative societal impacts beyond those already associated with general advances in machine learning theory. Any downstream impact will arise only through applications that use complex-valued networks; in those settings, standard considerations around fairness, robustness, and potential misuse of models still apply.

References

[1] M. Anthony and P. L. Bartlett. *Neural Network Learning: Theoretical Foundations*. Cambridge University Press, 1999.

[2] P. L. Bartlett, D. J. Foster, and M. Telgarsky. Spectrally-normalized margin bounds for neural networks. In *Advances in Neural Information Processing Systems 30*, pages 6240–6249. Curran Associates, Inc., 2017.

[3] P. L. Bartlett, N. Harvey, C. Liaw, and A. Mehrabian. Nearly-tight VC-dimension and pseudodimension bounds for deep neural networks. *Journal of Machine Learning Research*, 20(63):1–17, 2019.

[4] S. Basu, R. Pollack, and M.-F. Roy. *Algorithms in Real Algebraic Geometry*. Springer, 2nd edition, 2006.

[5] E. B. Baum and D. Haussler. What size net gives valid generalization? *Neural Computation*, 1(1):151–160, 1989.

[6] P. Geuchen and F. Voigtlaender. Optimal approximation using complex-valued neural networks. In *Advances in Neural Information Processing Systems 36*, 2023.

[7] P. Geuchen. Universal approximation with complex-valued deep narrow networks. *Constructive Approximation*, 2025.

[8] P. W. Goldberg and M. R. Jerrum. Bounding the VC-dimension of concept classes parameterized by real numbers. *Machine Learning*, 18(2–3):131–148, 1995.

[9] N. Harvey, C. Liaw, and A. Mehrabian. Nearly-tight VC-dimension bounds for piecewise linear neural networks. In *Proceedings of the Conference on Learning Theory*, volume 65 of *Proceedings of Machine Learning Research*, pages 1064–1068. PMLR, 2017.

[10] A. Hirose. Complex-valued neural networks for signal processing. In S. Haykin and B. Kosko, editors, *Intelligent Signal Processing*, pages 471–509. Wiley–IEEE Press, 2003.

[11] A. Hirose. *Complex-Valued Neural Networks*. Springer, 2nd edition, 2012.

[12] Y. J. Kim and S. Adali. Approximation by fully complex multilayer perceptrons. *Neural Computation*, 15(7):1641–1666, 2003.

[13] C. Y. Lee, H. Hasegawa, and S. Gao. Complex-valued neural networks: A comprehensive survey. *IEEE/CAA Journal of Automatica Sinica*, 9(8):1406–1426, 2022.

[14] B. Neyshabur. Towards learning convolutions from scratch. arXiv preprint arXiv:1809.01062, 2018.

[15] J. W. Smith, P. J. Wolfe, and V. Tarokh. Complex-valued deep learning for signal processing. arXiv preprint arXiv:2309.07948, 2023.

[16] E. D. Sontag. VC dimension of neural networks. In M. A. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*, pages 1188–1192. MIT Press, 2nd edition, 2003.

[17] A. Tarski. *A Decision Method for Elementary Algebra and Geometry.* University of California Press, 1951.

[18] C. Trabelsi, O. Bilaniuk, Y. Zhang, D. Serdyuk, S. Subramanian, J. F. Santos, S. Mehri, N. Rostamzadeh, Y. Bengio, and C. J. Pal. Deep complex networks. In *International Conference on Learning Representations*, 2018.

[19] V. N. Vapnik. *Statistical Learning Theory.* Wiley, 1998.

[20] M. J. Wainwright. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint.* Cambridge University Press, 2019.

[21] H. Chen, F. He, S. Lei, and D. Tao. Spectral complexity-scaled generalisation bound of complex-valued neural networks. *Artificial Intelligence*, 322:103951, 2023.