REVISTA MATEMÁTICA IBEROAMERICANA
VOL. 10, N.° 3, 1994

# Reconstructing a neural net from its output

**Charles Fefferman**

## Introduction.

Neural nets were originally introduced as highly simplified models of the nervous system. Today they are widely used in technology and studied theoretically by scientists from several disciplines. (See *e.g.* [N]). However, they remain little understood.

Mathematically, a (feed-forward) neural net consists of

(1) A finite sequence of positive integers $(D_0, D_1, \ldots, D_L)$,

(2) A family of real numbers $(\omega^{\ell}_{jk})$ defined for $1 \le \ell \le L$, $1 \le j \le D_\ell$, $1 \le k \le D_{\ell-1}$, and

(3) A family of real numbers $(\theta^{\ell}_j)$ defined for $1 \le \ell \le L$, $1 \le j \le D_\ell$.

The sequence $(D_0, D_1, \ldots, D_L)$ is called the *architecture* of the neural net, while the $\omega^{\ell}_{jk}$ are called *weights* and the $\theta^{\ell}_j$ *thresholds*.

Neural nets are used to compute non-linear maps from $\mathbb{R}^N$ to $\mathbb{R}^M$ by the following construction. We begin by fixing a nonlinear function $\sigma(x)$ of one variable. Analogy with the nervous system suggests that we take $\sigma(t)$ asymptotic to constants as $t$ tends to $\pm\infty$; a standard choice, which we adopt throughout this paper, is $\sigma(x) = \tanh(x/2)$. Given an "input" $(t_1, \ldots, t_{D_0}) \in \mathbb{R}^{D_0}$, we define real numbers $x^{\ell}_j$ for $0 \le \ell \le L$, $1 \le j \le D_\ell$ by the following induction on $\ell$.

(4) If $\ell = 0$ then $x^{\ell}_j = t_j$.

(5) If the $x_k^{\ell-1}$ are known, $1 \le \ell \le L$, then we set

$$x_j^\ell = \sigma\Big( \sum_{1 \le k \le D_{\ell-1}} \omega_{jk}^\ell x_k^{\ell-1} + \theta_j^\ell \Big), \qquad \text{for } 1 \le j \le D_\ell .$$

Here $x_1^\ell, \dots, x_{D_\ell}^\ell$ are interpreted as the outputs of $D_\ell$ "neurons" in the $\ell^{\text{th}}$ "layer" of the net. The *output map* of the net is defined as the map

$$\Phi: (t_1, \dots, t_{D_0}) \longmapsto (x_1^L, \dots, x_{D_L}^L) . \tag{6}$$

In practical applications, one tries to pick the neural net

$$[(D_0, D_1, \dots, D_L), \ (\omega_{jk}^\ell), \ (\theta_j^\ell)]$$

so that the output map $\Phi$ approximates a given map about which we have only imperfect information. The main result of this paper is that under generic conditions, perfect knowledge of the output map $\Phi$ uniquely specifies the architecture, the weights and the thresholds of a neural net, up to obvious symmetries. More precisely, the obvious symmetries are as follows. Let $(\gamma_0, \gamma_1, \dots, \gamma_L)$ be permutations, with

$$\gamma_\ell: \{1, \dots, D_\ell\} \to \{1, \dots, D_\ell\} ;$$

and let $\{\varepsilon_j^\ell : 0 \le \ell \le L, \ 1 \le j \le D_\ell\}$ be a collection of $\pm 1$'s. Assume that $\gamma_\ell$ is the identity and $\varepsilon_j^\ell = +1$ whenever $\ell = 0$ or $\ell = L$. Then one checks easily that the neural nets

(7) $[(D_0, D_1, \dots, D_L), (\omega_{jk}^\ell), (\theta_j^\ell)]$ and

(8) $[(D_0, D_1, \dots, D_L), (\widetilde{\omega}_{jk}^\ell), (\widetilde{\theta}_j^\ell)]$

have the same output map if we set

$$\widetilde{\omega}_{jk}^\ell = \varepsilon_j^\ell \, \omega_{[\gamma_\ell j][\gamma_{\ell-1} k]}^\ell \, \varepsilon_k^{\ell-1} \quad \text{and} \quad \widetilde{\theta}_j^\ell = \varepsilon_j^\ell \, \theta_{[\gamma_\ell j]} . \tag{9}$$

This reflects the facts that the neurons in layer $\ell$ are interchangeable, $1 \le \ell \le L-1$, and that the function $\sigma(x)$ is odd. The nets (7) and (8) will be called *isomorphic* if they are related by (9). Note in particular that isomorphic neural nets have the same architecture. Our main theorem asserts that, under generic conditions, any two neural nets with the same output map are isomorphic.

We discuss the generic conditions which we impose on neural nets. We have to avoid obvious counterexamples such as

(10) Suppose all the weights $\omega_{jk}^{\ell}$ are zero. Then the output map $\Phi$ is constant. The architecture and thresholds of the neural net are clearly not uniquely determined by $\Phi$.

(11) Fix $\ell_0, j_1, j_2$ with $1 \leq \ell_0 \leq L-1$ and $1 \leq j_1 < j_2 \leq D_{\ell_0}$. Suppose we have $\theta_{j_1}^{\ell_0} = \theta_{j_2}^{\ell_0}$ and $\omega_{j_1 k}^{\ell_0} = \omega_{j_2 k}^{\ell_0}$ for all $k$. Then (5) gives $x_{j_1}^{\ell_0} = x_{j_2}^{\ell_0}$. Therefore, the output depends on $\omega_{jj_1}^{\ell_0+1}$ and $\omega_{jj_2}^{\ell_0+1}$ only through the sum $\omega_{jj_1}^{\ell_0+1} + \omega_{jj_2}^{\ell_0+1}$. So the output map does not uniquely determine the weights.

Our hypotheses are more than adequate to exclude these counterexamples. Specifically, we assume that

(12) $\theta_j^{\ell} \neq 0$, and $|\theta_j^{\ell}| \neq |\theta_{j'}^{\ell}|$ for $j \neq j'$.

(13) $\omega_{jk}^{\ell} \neq 0$; and for $j \neq j'$, the ratio $\omega_{jk}^{\ell}/\omega_{j'k}^{\ell}$ is not equal to any fraction of the form $p/q$ with $p$, $q$ integers and $1 \leq q \leq 100\, D_{\ell}^2$.

Evidently, these conditions hold for generic neural nets. The precise statement of our main theorem is as follows. *If two neural nets satisfy* (12), (13) *and have the same output, then the nets are isomorphic.* In Section I we give a slightly different but clearly equivalent statement of our main result. It would be interesting to replace (12), (13) by minimal hypotheses, and to study functions $\sigma(x)$ other than $\tanh(x/2)$.

We now sketch the proof of our main result, sacrificing accuracy for simplicity. After a trivial reduction, we may assume $D_0 = D_L = 1$. Thus, the outputs of the nodes $x_j^{\ell}(t)$ are functions of one variable, and the output map of the neural net is $t \mapsto x_1^L(t)$. The key idea is to continue the $x_j^{\ell}(t)$ analytically to complex values of $t$, and to read off the structure of the net from the set of singularities of the $x_j^{\ell}$. Note that $\sigma(x) = \tanh(x/2)$ is meromorphic, with poles at the points of an arithmetic progression $\{(2m+1)\pi i\colon\, m \in \mathbb{Z}\}$. This leads to two crucial observations.

(14) When $\ell = 1$, the poles of $x_j^{\ell}(t)$ form an arithmetic progression $\Pi_j^1$, and

(15) When $\ell > 1$, every pole of any $x_k^{\ell-1}(t)$ is an accumulation point of poles of any $x_j^{\ell}(t)$.

In fact, (14) is immediate from the formula $x_j^1(t) = \sigma(\omega_{j1}^1 t + \theta_j^1)$, which is merely the special case $D_0 = 1$ of (5). We obtain

$$\Pi_j^1 = \left\{ \frac{(2m+1)\pi i - \theta_j^1}{\omega_{j1}^1} : \; m \in \mathbb{Z} \right\}. \tag{16}$$

To see (15), fix $\ell$, $j$, $\bar{k}$, and assume for simplicity that $x_{\bar{k}}^{\ell-1}(t)$ has a simple pole at $t_0$, while $x_k^{\ell-1}(t)$, $k \neq \bar{k}$, is analytic in a neighborhood of $t_0$. Then

$$x_{\bar{k}}^{\ell-1}(t) = \frac{\lambda}{t - t_0} + f(t), \tag{17}$$

with $f$ analytic in a neighborhood of $t_0$.

From (17) and (5), we obtain

$$x_j^\ell(t) = \sigma(\omega_{j\bar{k}}^\ell \lambda (t - t_0)^{-1} + g(t)), \tag{18}$$

with

$$g(t) = \omega_{j\bar{k}}^\ell f(t) + \sum_{k \neq \bar{k}} \omega_{jk}^\ell \, x_k^{\ell-1}(t) + \theta_j^\ell \tag{19}$$

analytic in a neighborhood of $t_0$.

Thus, in a neighborhood of $t_0$, the poles of $x_j^\ell(t)$ are the solutions $\tilde{t}_m$ of the equation

$$\frac{\omega_{j\bar{k}}^\ell \lambda}{\tilde{t}_m - t_0} + g(\tilde{t}_m) = (2m+1)\pi i, \qquad m \in \mathbb{Z}. \tag{20}$$

There are infinitely many solutions of (20), accumulating at $t_0$. Hence, $t_0$ is an accumulation point of poles of $x_j^\ell(t)$, which completes the proof of (15).

In view of (14), (15), it is natural to make the following definitions. The *natural domain* of a neural net is the largest open subset of the complex plane to which the output map $t \mapsto x_1^L(t)$ can be analytically continued. For $\ell \geq 0$ we define the $\ell^{\text{th}}$ *singular set* Sing$(\ell)$ by setting

Sing$(0)$ = complement of the natural domain in $\mathbb{C}$, and

Sing$(\ell + 1)$ = the set of all accumulation points of Sing$(\ell)$.

These definitions are made entirely in terms of the output map, without reference to the structure of the given neural net. On the other hand, the sets Sing$(\ell)$ contain nearly complete information on the architecture, weights and thresholds of the net.

This will allow us to read off the structure of a neural net from the analytic continuation of its output map. To see how the sets Sing$(\ell)$ reflect the structure of the net, we reason as follows.

From (14) and (15) we expect that

(21) For $1 \leq \ell \leq L$, Sing$(L-\ell)$ is the union over $j = 1, \ldots, D_\ell$ of the set of poles of $x_j^\ell(t)$, together with their accumulation points (which we ignore here), and

(22) For $\ell \geq L$, Sing$(\ell)$ is empty.

Immediately, then, we can read off the "depth" $L$ of the neural net; it is simply the smallest $\ell$ for which Sing$(\ell)$ is empty.

We need to solve for $D_\ell$, $\omega_{jk}^\ell$, $\theta_j^\ell$. We proceed by induction on $\ell$.

*When* $\ell = 1$, (14) and (21) show that Sing$(L-1)$ is the union of arithmetic progressions $\Pi_j^1$, $j = 1, \ldots, D_1$. Therefore, from Sing$(L-1)$ we can read off $D_1$ and the $\Pi_j^1$. (We will return to this point later in the introduction.) In view of (16), $\Pi_j^1$ determines the weights and thresholds at layer 1, modulo signs. Thus, we have found $D_1$, $\omega_{jk}^1$, $\theta_j^1$.

*When* $\ell > 1$, we may assume that

(23) The $D_{\ell'}$, $\omega_{jk}^{\ell'}$, $\theta_j^{\ell'}$ are already known, for $1 \leq \ell' \leq \ell$.

Our task is to find $D_\ell$, $\omega_{jk}^\ell$, $\theta_j^\ell$. In view of (23), we can find a pole $t_0$ of $x_{\bar{k}}^{\ell-1}(t)$ for our favorite $\bar{k}$. Assume for simplicity that $t_0$ is a simple pole of $x_{\bar{k}}^{\ell-1}(t)$, and that the $x_k^{\ell-1}(t)$, $k \neq \bar{k}$, are analytic in a neighborhood of $t_0$. Then $x_{\bar{k}}^{\ell-1}(t)$ is given by (17) in a neighborhood of $t_0$, with $\lambda$ already known by virtue of (23). Let $U$ be a small neighborhood of $t_0$.

We will look at the image $Y$ of $U \cap$ Sing$(L-\ell)$ under the map $t \mapsto \lambda/(t-t_0)$. Since $\lambda$, $t_0$ and Sing$(L-\ell)$ are already known, so is $Y$. On the other hand, we can relate $Y$ to $D_\ell$, $\omega_{jk}^\ell$, $\theta_j^\ell$ as follows. From (21) we see that $Y$ is the union over $j = 1, \ldots, D_\ell$ of

(24) $Y_j =$ image of $U \cap \{$ Poles of $x_j^\ell(t)\}$ under $t \longmapsto \lambda/(t-t_0)$.

For fixed $j$, the poles of $x_j^\ell(t)$ in a neighborhood of $t_0$ are the $\tilde{t}_m$ given by (20). We write

$$\text{(25)} \qquad \frac{\omega^{\ell}_{j\bar{k}}\lambda}{\tilde{t}_m - t_0} = \left[\frac{\omega^{\ell}_{j\bar{k}}\lambda}{(\tilde{t}_m - t_0)} + g(\tilde{t}_m)\right] + \left[g(t_0) - g(\tilde{t}_m)\right] .$$

Equation (20) shows that the first expression in brackets in (25) is equal to $(2m+1)\pi i$. Also, since $\tilde{t}_m \to t_0$ as $|m| \to +\infty$ and $g$ is analytic in a neighborhood of $t_0$, the second expression in brackets in (25) tends to zero. Hence,

$$\frac{\omega^{\ell}_{j\bar{k}}\lambda}{\tilde{t}_m - t_0} = (2m+1)\pi i - g(t_0) + o(1), \qquad \text{for large } m .$$

Comparing this with the definition (24), we see that $Y_j$ is asymptotic to the arithmetic progression

$$\text{(26)} \qquad \Pi^{\ell}_j = \left\{\frac{(2m+1)\pi i - g(t_0)}{\omega^{\ell}_{j\bar{k}}} : \ m \in \mathbb{Z}\right\}.$$

Thus, the known set $Y$ is the union over $j = 1, \dots, D_\ell$ of sets $Y_j$, with $Y_j$ asymptotic to the arithmetic progression $\Pi^\ell_j$. From $Y$, we can therefore read off $D_\ell$ and the $\Pi^\ell_j$. (We will return to this point in a moment). We see at once from (26) that $\omega^{\ell}_{j\bar{k}}$ is determined up to sign by $\Pi^\ell_j$. Thus, we have found $D_\ell$ and $\omega^{\ell}_{j\bar{k}}$. With more work, we can also find the $\theta^\ell_j$, completing the induction on $\ell$.

The above induction shows that the structure of a neural net may be read off from the analytic continuation of its output map. We believe that the analytic continuation of the output map will lead to further consequences in the study of neural nets.

Let us touch briefly on a few points which we glossed over above. First of all, suppose we are given a set $Y \subset \mathbb{C}$, and we know that $Y$ is the union of sets $Y_1, \dots, Y_D$, with $Y_j$ asymptotic to an arithmetic progression $\Pi_j$. We assumed above that $\Pi_1, \dots, \Pi_D$ are uniquely determined by $Y$. In fact, without some further hypothesis on the $\Pi_j$, this need not be true. For instance, we cannot distinguish $\Pi_1 \cup \Pi_2$ from $\Pi_3$ if $\Pi_1 = \{\text{odd integers}\}$, $\Pi_2 = \{\text{even integers}\}$, $\Pi_3 = \{\text{all integers}\}$. On the other hand, we can clearly recognize $\Pi_1 = \{\text{all integers}\}$ and $\Pi_2 = \{m\sqrt{2} : m \text{ an integer}\}$ from their union $\Pi_1 \cup \Pi_2$. Thus, irrational numbers enter the picture. The rôle of our generic hypothesis (13) is to control the arithmetic progressions that arise in our proof.

Secondly, suppose $x_{\bar{k}}^{\ell}(t)$ has a pole at $t_0$. We assumed for simplicity that $x_k^{\ell}(t)$ is analytic in a neighborhood of $t_0$ for $k \neq \bar{k}$. However, one of the $x_k^{\ell}(t)$, $k \neq \bar{k}$, may also have a pole at $t_0$. In that case, the $x_j^{\ell+1}(t)$ may all be analytic in a neighborhood of $t_0$, because the contributions of the singularities of the $x_k^{\ell}$ in $\sigma\big(\sum_k \omega_{jk}^{\ell+1} x_k^{\ell} + \theta_j^{\ell+1}\big)$ may cancel. Thus, the singularity at $t_0$ may disappear from the output map. While this circumstance is hardly generic, it is not ruled out by our hypotheses (12), (13). Because singularities can disappear, we have to make technical changes in our description of Sing$(\ell)$. For example, in the discussion following (23), $Y$ need not be the union of the sets $Y_j$. Rather, $Y$ is their "approximate union", in a sense to be made precise in (II.A.1) below.

Next, we should point out that the signs of the weights and thresholds require some attention, even though we have some freedom to change signs by applying isomorphisms. (See (9).) In effect, we introduce in Section IV.A an extra induction on the number of neurons in the net, in order to show that the signs come out correctly. The induction comes into play in the substantial Lemma IV.B.16 below.

Finally, in the definition of the natural domain, we have assumed that there is a unique maximal open set to which the output map continues analytically. This need not be true of a general real-analytic function on the line -for instance, take $f(t) = (1 + t^2)^{1/2}$. Fortunately, Lemma III.A.1 below shows that the natural domain is well-defined for any function that continues analytically to the complement of a countable set. The defining formula (5) lets us check easily that the output map continues to the complement of a countable set, so the natural domain makes sense. This concludes our overview of the proof of our main theorem.

Both the uniqueness problem and the use of analytic continuation have already appeared in the neural net literature. In particular, it was R. Hecht-Nielson who pointed out the rôle of isomorhisms and posed the uniqueness problem. His paper with Chen and Lu [CLH] on "equioutput transformations" on the space of all neural nets influenced our work. E. Sontag [So] and H. Sussman [Su] proved sharp uniqueness theorems for one hidden layer. The proof in [Su] uses complex variables.

At this stage, few non-trivial results are known for neural nets with more than one hidden layer, *i.e.* with $L > 1$. However, a recent paper of Macintyre and Sontag [MS] proves finiteness of the VC dimension, a measure of the computing power of a neural net.

I am grateful to R. Crane, S. Markel, J. Pearson, E. Sontag, R. Sverdlove, and N. Winarsky for introducing me to the study of neural nets.

## I. Statement of the Main Results.

### A. Definitions.

A *neural net* consists of the following:

(1) A finite sequence of positive integers $(D_0, D_1, \dots, D_L)$ with $L \geq 1$.

(2) A collection of real numbers $(\omega^\ell_{jk})$, defined for $1 \leq \ell \leq L$, $1 \leq j \leq D_\ell$, $1 \leq k \leq D_{\ell-1}$.

(3) A collection of real numbers $(\theta^\ell_j)$, defined for $1 \leq \ell \leq L$, $1 \leq j \leq D_\ell$.

Here, $L$ is called the *depth* of the net, and $(D_0, D_1, \dots, D_L)$ is called the *architecture* of the net. The $\omega^\ell_{jk}$ are called *weights*, while the $\theta^\ell_j$ are called *thresholds.*

Thus, a neural net has the form $[(D_0, \dots, D_L), (\omega^\ell_{jk}), (\theta^\ell_j)]$. We denote neural nets by $\mathcal{N}$, $\mathcal{N}'$, $\widetilde{\mathcal{N}}$, etc.

For $\mathcal{N} = [(D_0, \dots, D_L), (\omega^\ell_{jk}), (\theta^\ell_j)]$, we define functions $x^\ell_j(t_1, \dots, t_{D_0}, \mathcal{N})$ by the following induction on $\ell$.

$$x^0_j(t_1, \dots, t_{D_0}, \mathcal{N}) = t_j \quad \text{for} \quad 1 \leq j \leq D_0 \,. \tag{4}$$

$$x^\ell_j(t_1, \dots, t_{D_0}, \mathcal{N}) = \sigma\left( \sum_{k=1}^{D_{\ell-1}} \omega^\ell_{jk} x^{\ell-1}_k(t_1, \dots, t_{D_0}, \mathcal{N}) + \theta^\ell_j \right), \tag{5}$$

for $1 \leq j \leq D_\ell$, where

$$\sigma(x) = \tanh\left(\frac{x}{2}\right) \,. \tag{6}$$

We call $(t_0, t_1, \dots, t_{D_0}) \in \mathbb{R}^{D_0}$ the *input* to the neural net; we call $(x^L_j(t_1, \dots, t_{D_0}, \mathcal{N}))_{1 \leq j \leq D_L} \in \mathbb{R}^{D_L}$ the *output*, or the *function computed by the neural net*; and we call $x^\ell_j(t_1, \dots, t_{D_0}, \mathcal{N})$ the *function computed by the* $j^{\text{th}}$ *node of the* $\ell^{\text{th}}$ *layer*.

When it is clear which neural net we are talking about, we may write $x^\ell_j(t_1, \dots, t_{D_0})$ in place of $x^\ell_j(t_1, \dots, t_{D_0}, \mathcal{N})$. Also, when $D_0 = 1$, we may write $x^\ell_j(t)$ or $x^\ell_j(t, \mathcal{N})$ in place of $x^\ell_j(t_1), x^\ell_j(t_1, \mathcal{N})$.

The *size* of a neural net $\mathcal{N} = [(D_0, \ldots, D_L), (\omega^\ell_{jk}), (\theta^\ell_j)]$ is defined simply as $D_0 + D_1 + \cdots + D_L$.

Next we discuss isomorphisms of neural nets. Let

$$\mathcal{N} = [(D_0, \ldots, D_L), (\omega^\ell_{jk}), (\theta^\ell_j)] \tag{7}$$

be a neural net. Then let

$$\gamma_\ell : \{1, \ldots, D_\ell\} \to \{1, \ldots, D_\ell\} \tag{8}$$

be permutations. Finally, let $\varepsilon^\ell_j$ be a collection of signs,

$$\varepsilon^\ell_j = \pm 1, \qquad \text{for } 0 \le \ell \le L,\ 1 \le j \le D_\ell . \tag{9}$$

In terms of $\mathcal{N}$, $(\gamma_\ell)$, $(\varepsilon^\ell_j)$, we define a new neural net

$$\widetilde{\mathcal{N}} = [(D_0, \ldots, D_L), (\widetilde{\omega}^\ell_{jk}), (\widetilde{\theta}^\ell_j)], \tag{10}$$

where

$$\widetilde{\omega}^\ell_{jk} = \varepsilon^\ell_j \, \omega^\ell_{(\gamma_\ell j)(\gamma_{\ell-1} k)} \, \varepsilon^{\ell-1}_k \tag{11}$$

and

$$\widetilde{\theta}^\ell_j = \varepsilon^\ell_j \, \theta^\ell_{(\gamma_\ell j)} . \tag{12}$$

An easy induction on $\ell$ shows that

$$x^\ell_j(\tilde{t}_1, \ldots, \tilde{t}_{D_0}, \widetilde{\mathcal{N}}) = \varepsilon^\ell_j \, x^\ell_{(\gamma_\ell j)}(t_1, \ldots, t_{D_0}, \mathcal{N}), \tag{13}$$

provided $(\tilde{t}_1, \ldots, \tilde{t}_{D_0})$ and $(t_1, \ldots, t_{D_0})$ are related by

$$\tilde{t}_j = \varepsilon^0_j \, t_{(\gamma_0 j)}, \qquad \text{for } 1 \le j \le D_0 . \tag{14}$$

In particular, if we assume

(15) $\varepsilon^\ell_j = 1$ when $\ell = 0$ or $L$ and $\gamma_0$, $\gamma_L$ are the identity permutation,

then (13), (14) show that

$$x^L_j(t_1, \ldots, t_{D_0}, \widetilde{\mathcal{N}}) = x^L_j(t_1, \ldots, t_{D_0}, \mathcal{N}), \qquad \text{for } 1 \le j \le D_L . \tag{16}$$

Thus, the neural nets $\widetilde{\mathcal{N}}$ and $\mathcal{N}$ compute the same function. We say that the nets $\mathcal{N}$, $\widetilde{\mathcal{N}}$ are *isomorphic* if they are related by (7), ..., (12) for some choice of $(\gamma_\ell)$, $(\varepsilon_j^\ell)$ satisfying (15). For fixed

$$[(\gamma_\ell)_{0 \le \ell \le L},\ (\varepsilon_j^\ell)_{0 \le \ell \le L,\ 1 \le j \le D_\ell}]$$

satisfying (15), the map $\mathcal{N} \mapsto \widetilde{\mathcal{N}}$ given by (7), (10), (11), (12) is is called the *isomorphism induced by* $[(\gamma_\ell), (\varepsilon_j^\ell)]$. One checks easily that compositions and inverses of isomorphisms are again isomorphisms. Note that any two isomorphic neural nets $\mathcal{N}$, $\widetilde{\mathcal{N}}$ have the same architecture.

It is useful to pick out a single representative from an isomorphism class of neural nets. Thus, we say that $\mathcal{N} = [(D_0, \ldots, D_L), (\omega_{jk}^\ell), (\theta_j^\ell)]$ is in *standard order* if for each $\ell$, $1 \le \ell < L$, we have

$$0 < \theta_1^\ell < \theta_2^\ell < \cdots < \theta_{D_\ell}^\ell . \tag{17}$$

The proof of the following observation is left to the reader.

(18) **Lemma.** *Every neural net* $\mathcal{N} = [(D_0, \ldots, D_L), (\omega_{jk}^\ell), (\theta_j^\ell)]$ *satisfying the generic condition*

$$\theta_j^\ell \neq 0, \qquad |\theta_j^\ell| \neq |\theta_{j'}^\ell| \quad \text{for } j \neq j', \tag{19}$$

*is isomorphic to one and only one neural net in standard order.*

## B. The Main Theorems.

The main result of this paper is as follows.

(1) **Uniqueness Theorem.** *Let* $\mathcal{N} = [(D_0, \ldots, D_L), (\omega_{jk}^\ell), (\theta_j^\ell)]$ *and* $\widetilde{\mathcal{N}} = [(\widetilde{D}_0, \ldots, \widetilde{D}_{\widetilde{L}}), (\widetilde{\omega}_{jk}^\ell), (\widetilde{\theta}_j^\ell)]$ *be neural nets in standard order, satisfying the following generic conditions.*

$$\omega_{jk}^\ell \neq 0 \qquad \text{and} \qquad \left|\frac{\omega_{jk}^\ell}{\omega_{j'k}^\ell}\right| \neq \frac{p}{q}, \tag{2}$$

*for* $j \neq j'$, $p, q \in \mathbb{Z}$, $1 \le q \le 100\, D_\ell^2$,

$$\widetilde{\omega}_{jk}^\ell \neq 0 \qquad \text{and} \qquad \left|\frac{\widetilde{\omega}_{jk}^\ell}{\widetilde{\omega}_{j'k}^\ell}\right| \neq \frac{p}{q}, \tag{3}$$

*for* $j \neq j'$, $p, q \in \mathbb{Z}$, $1 \le q \le 100\,\widetilde{D}_\ell^2$. *Assume* $D_0 = \widetilde{D}_0$, $D_L = \widetilde{D}_{\widetilde{L}}$, *and*

$$x_j^L(t_1, \ldots, t_{D_0}, \mathcal{N}) = x_j^{\widetilde{L}}(t_1, \ldots, t_{D_0}, \widetilde{\mathcal{N}}), \tag{4}$$

*for all* $(t_1, \ldots, t_{D_0}) \in \mathbb{R}^{D_0}$ *and all* $j$, $1 \le j \le D_L$.

*Then the nets* $\mathcal{N}$ *and* $\widetilde{\mathcal{N}}$ *are identical:*

(5) $L = \widetilde{L}$,

(6) $D_\ell = \widetilde{D}_\ell$ *for* $0 \le \ell \le L$,

(7) $\omega_{jk}^\ell = \widetilde{\omega}_{jk}^\ell$ *for* $1 \le \ell \le L$, $1 \le j \le D_\ell$, $1 \le k \le D_{\ell-1}$,

(8) $\theta_j^\ell = \widetilde{\theta}_j^\ell$ *for* $1 \le \ell \le L$, $1 \le j \le D_\ell$.

The Uniqueness Theorem 1 reduces immediately to the special case. $D_0 = 1$, $D_L = 1$. To see this, we fix $j_0$, $1 \le j_0 \le D_L$, and $k_0$, $1 \le k_0 \le D_0$. Then we restrict attention to the $j_0^{\text{th}}$ outputs $x_{j_0}^L(\,\cdot\,, \widetilde{\mathcal{N}})$ for inputs of the form $(0, \ldots, 0, t, 0, \ldots, 0)$, where the $t$ occurs in the $k_0^{\text{th}}$ coordinate. Thus we obtain functions $x^L(t, \mathcal{N})$, $x^{\widetilde{L}}(t, \widetilde{\mathcal{N}})$ of a single variable $t$. These functions are computed by neural nets $\mathcal{N}_{\text{reduced}}$, $\widetilde{\mathcal{N}}_{\text{reduced}}$ obtained from $\mathcal{N}$, $\widetilde{\mathcal{N}}$ by deleting irrelevant input and output nodes. The special case $D_0 = D_L = 1$ of Theorem (1), applied to $\mathcal{N}_{\text{reduced}}$ and $\widetilde{\mathcal{N}}_{\text{reduced}}$, shows that $\mathcal{N}_{\text{reduced}}$ and $\widetilde{\mathcal{N}}_{\text{reduced}}$ are identical. Since $j_0$ and $k_0$ were arbitary, it follows that $\mathcal{N}$ and $\widetilde{\mathcal{N}}$ are identical. Thus, Theorem (1) is reduced to the special case $D_0 = D_L = 1$.

From now on, *we change the definition of neural nets* to include the requirement $D_0 = D_L = 1$. Thus, a neural net computes a single function of one variable. In view of the elementary Lemma A.18, our uniqueness theorem is reduced to the following statement.

(9) **Uniqueness Theorem.** *Let* $\mathcal{N} = [(D_0, \ldots, D_L), (\omega_{jk}^\ell), (\theta_j^\ell)]$ *and* $\widetilde{\mathcal{N}} = [(\widetilde{D}_0, \ldots, \widetilde{D}_{\widetilde{L}}), (\widetilde{\omega}_{jk}^\ell), (\widetilde{\theta}_j^\ell)]$ *be neural nets satisfying the generic conditions*

$$\omega_{jk}^\ell \neq 0, \qquad \textit{and} \qquad \left|\frac{\widetilde{\omega}_{jk}^\ell}{\widetilde{\omega}_{j'k}^\ell}\right| \neq \frac{p}{q}, \tag{10}$$

*for* $j \neq j'$, $p, q \in \mathbb{Z}$, $1 \le q \le 100\, D_\ell^2$, *and*

$$\widetilde{\omega}_{jk}^\ell \neq 0, \qquad \textit{and} \qquad \left|\frac{\widetilde{\omega}_{jk}^\ell}{\widetilde{\omega}_{j'k}^\ell}\right| \neq \frac{p}{q}, \tag{11}$$

*for* $j \neq j'$, $p, q \in \mathbb{Z}$, $1 \leq q \leq 100\, \widetilde{D}_\ell^2$.

*If* $x_1^L(t, \mathcal{N}) = x_1^{\widetilde{L}}(t, \widetilde{\mathcal{N}})$ *for all real* $t$*, then* $\mathcal{N}$ *and* $\widetilde{\mathcal{N}}$ *are isomorphic.*

The rest of this paper is devoted to the proof of Theorem 9.

## C. A Small Technical Lemma.

The following observation on isomorphic neural nets will be used much later, in the proof of Theorem B.9.

(1) **Lemma.** *Let* $\mathcal{N} = [(D_0, \ldots, D_L), (\omega_{jk}^\ell), (\theta_j^\ell)]$ *and* $\widetilde{\mathcal{N}} = [(D_0, \ldots, D_L), (\widetilde{\omega}_{jk}^\ell), (\widetilde{\theta}_j^\ell)]$ *be isomorphic neural nets. Assume that*

(2) $\omega_{jk}^\ell \neq 0$ *for all* $\ell$, $j$, $k$, $1 \leq \ell \leq L$, $1 \leq j \leq D_\ell$, $1 \leq k \leq D_{\ell-1}$,

(3) $|\omega_{jk}^\ell| \neq |\omega_{j'k}^\ell|$ *for all* $\ell$, $j \neq j'$, $k$, $1 \leq \ell \leq L$, $1 \leq j, j' \leq D_\ell$, $1 \leq k \leq D_{\ell-1}$,

(4) $\omega_{jk}^\ell = \widetilde{\omega}_{jk}^\ell$ *for* $1 \leq \ell \leq L-1$, $1 \leq j \leq D_\ell$, $1 \leq k \leq D_{\ell-1}$.

(*Note: We do not assume* (4) *for* $\ell = L$).

*Then* $\mathcal{N}$ *and* $\widetilde{\mathcal{N}}$ *are identical.*

PROOF. Since $\mathcal{N}$, $\widetilde{\mathcal{N}}$ are isomorphic, there are permutations $\gamma_\ell$ and signs $\varepsilon_j^\ell$ such that

(5) $$\omega_{jk}^\ell = \varepsilon_j^\ell\, \widetilde{\omega}_{(\gamma_\ell j)(\gamma_{\ell-1} k)}^\ell\, \varepsilon_k^{\ell-1},$$

(6) $$\theta_j^\ell = \varepsilon_j^\ell\, \widetilde{\theta}_{(\gamma_\ell j)}^\ell,$$

(7) $$\gamma_0 = \text{identity},\ \gamma_L = \text{identity},\ \varepsilon_1^0 = 1,\ \varepsilon_1^L = 1.$$

Since $\widetilde{\omega}_{jk}^\ell = \omega_{jk}^\ell$ for $\ell \leq L-1$, (5) implies

(8) $$\omega_{jk}^\ell = \varepsilon_j^\ell\, \omega_{(\gamma_\ell j)(\gamma_{\ell-1} k)}^\ell\, \varepsilon_k^{\ell-1}, \qquad \text{for } 1 \leq \ell \leq L-1,$$

so that

(9) $$|\omega_{jk}^\ell| = |\omega_{(\gamma_\ell j)(\gamma_{\ell-1} k)}^\ell|, \qquad \text{for } 1 \leq \ell \leq L-1.$$

From (7) we have $\gamma_0 =$ identity. By (3) and (9), $\gamma_{\ell-1} =$ identity implies $\gamma_\ell =$ identity for $1 \le \ell \le L-1$. Hence $\gamma_\ell =$ identity for all $\ell \le L-1$. Since $\gamma_L =$ identity by (7), we know that all the $\gamma_\ell =$ identity. Thus, (8) becomes

$$\omega^\ell_{jk} = \varepsilon^\ell_j \, \omega^\ell_{jk} \, \varepsilon^{\ell-1}_k \,, \qquad \text{for } 1 \le \ell \le L-1 \,. \tag{10}$$

From (7) we have $\varepsilon^0_k = 1$, since $D_0 = 1$. By (2) and (10), $\varepsilon^{\ell-1}_k = 1$ (all $k$) implies $\varepsilon^\ell_j = 1$ (all $j$) for $1 \le \ell \le L-1$. Hence $\varepsilon^\ell_j = 1$ whenever $\ell \le L-1$. Since also (7) gives $\varepsilon^L_j = 1$ because $D_L = 1$, we know that $\varepsilon^\ell_j = 1$ for all $\ell$, $j$, $0 \le \ell \le L$, $1 \le j \le D_\ell$. Since $\varepsilon^\ell_j = 1$ and $\gamma_\ell =$ identity, (5) and (6) show that the nets $\mathcal{N}$, $\widetilde{\mathcal{N}}$ are identical.

## II. Approximate Arithmetic Progressions.

### A. Preliminaries.

(1) **Definition.** *Let $E$, $E_1, \dots, E_n \subset \mathbb{C}$ be given. We say that $E$ is the approximate union of $E_1, \dots, E_n$ if the following conditions hold.*

(2) $E \subset E_1 \cup \cdots \cup E_n$, *and*

(3) *Any point belonging to exactly one of the $E_1, \dots, E_n$ belongs to $E$.*

(4) **Definition.** *For $\omega, \beta \in \mathbb{C}$ with $\omega \ne 0$, define $\Pi(\omega, \beta) = \{\omega k + \beta : k \in \mathbb{Z}\}$. We say that $E \subset \mathbb{C}$ approximates $\Pi(\omega, \beta)$ if for every $\varepsilon > 0$ the following conditions hold.*

(5) *All but finitely many points of $E$ lie within distance $\varepsilon$ of some point in $\Pi(\omega, \beta)$, and*

(6) *All but finitely many points of $\Pi(\omega, \beta)$ lie within distance $\varepsilon$ of some point in $E$.*

Note that

(7) $\Pi(\omega, \beta) = \Pi(\omega', \beta')$ if and only if $\omega' = \pm\,\omega$ and $\beta' = \beta + \omega\, m$ for some $m \in \mathbb{Z}$.

(8) **Definition.** *Let $H$ be a set of integers. We define the upper and lower densities $\Delta^*(H)$, $\Delta_*(H)$ by setting*

$$\Delta^*(H) = \limsup_{\substack{N\to\infty \\ M\to-\infty}} \frac{\text{Number of integers in } [M,N]\cap H}{N-M} \tag{9}$$

*and*

$$\Delta_*(H) = \liminf_{\substack{N\to\infty \\ M\to-\infty}} \frac{\text{Number of integers in } [M,N]\cap H}{N-M}. \tag{10}$$

*If $\Delta^*(H) = \Delta_*(H)$, then we write $\Delta(H)$ for their common value, and we say that $H$ has density $\Delta(H)$.*

We will need the following special case of H. Weyl's Theorem on the equidistribution mod1 of arithmetic progressions (See [W]).

(11) **Theorem.** *Suppose $\theta \in \mathbb{R}$ is irrational and $0 < \varepsilon < 1/2$. Let $\beta \in \mathbb{R}$. Then $\Delta(\{k \in \mathbb{Z}\colon |\theta k + \beta - m| < \varepsilon \text{ for some } m \in \mathbb{Z}\}) = 2\varepsilon$.*

(12) **Corollary.** *Let $\omega$, $\omega'$, $\beta$, $\beta'$ be complex numbers, with $\omega, \omega' \neq 0$. Assume that $\omega'/\omega$ is real and irrational. Then we can make the density*

$$\Delta^*(\{k \in \mathbb{Z}\colon \operatorname{dist}\{\omega k + \beta, \Pi(\omega', \beta')\} < \varepsilon\})$$

*arbitrarily small, by taking $\varepsilon > 0$ small enough.*

## B. The Deconstruction Lemma.

Suppose $E \subset \mathbb{C}$ is the approximate union of sets $E_1, \ldots, E_D$; and suppose that each $E_j$ approximates an arithmetic progression $\Pi(\omega_j, \beta_j)$. We want to know that the progressions $\Pi(\omega_j, \beta_j)$ are uniquely determined by $E$. Also, for each $j_0$, we want to pick out infinitely many points $(x_\nu)_{\nu \geq 1}$ that belong to $E_{j_0}$ but not to any $E_j$, $j \neq j_0$. The following result provides this information.

**Deconstruction Lemma.** *Let $E$, $E_1, \ldots, E_D$, $\widetilde{E}_1, \ldots, \widetilde{E}_{\widetilde{D}}$ be subsets of $\mathbb{C}$, and let $\Pi(\omega_1, \beta_1), \ldots, \Pi(\omega_D, \beta_D)$, $\Pi(\widetilde{\omega}_1, \tilde{\beta}_1), \ldots, \Pi(\widetilde{\omega}_{\widetilde{D}}, \tilde{\beta}_{\widetilde{D}})$ be arithmetic progressions. Assume the following conditions.*

(1) *$E$ is the approximate union of $E_1, \ldots, E_D$.*

(2) *Each $E_j$ approximates the arithmetic progression $\Pi(\omega_j, \beta_j)$.*

(3) *For $j \neq j'$ and $p$, $q$ integers with $1 \le q \le 100\, D^2$, we have $|\omega_j / \omega_{j'}| \neq p/q$.*

(4) *$E$ is the approximate union of $\widetilde{E}_1, \ldots, \widetilde{E}_{\widetilde{D}}$.*

(5) *Each $\widetilde{E}_j$ approximates the arithmetic progression $\Pi(\widetilde{\omega}_j, \tilde{\beta}_j)$.*

(6) *For $j \neq j'$, and $p$, $q$ integers with $1 \le q \le 100\, \widetilde{D}^2$, we have $|\widetilde{\omega}_j / \widetilde{\omega}_{j'}| \neq p/q$.*

*Then $D = \widetilde{D}$, and there is a permutation*

$$\gamma: \{1, \ldots, D\} \to \{1, \ldots, D\}$$

*with the following properties:*

(7) *$\Pi(\omega_j, \beta_j) = \Pi(\widetilde{\omega}_{\gamma j}, \tilde{\beta}_{\gamma j})$ for $1 \le j \le D$.*

(8) *Given $j_0$, $1 \le j_0 \le D$, there is a sequence $(x_\nu)_{\nu \ge 1}$ in $\mathbb{C}$ such that*

(9) *$|x_\nu| \to \infty$ as $\nu \to \infty$,*

(10) *Each $x_\nu$ belongs to $E_{j_0}$ but not to $E_j$ for $j \neq j_0$,*

(11) *Each $x_\nu$ belongs to $\widetilde{E}_{\gamma j_0}$, but not to $\widetilde{E}_j$ for $j \neq \gamma j_0$.*

## C. Preparation for the proof of the Deconstruction Lemma.

We begin with a definition. We say that $\Pi(\omega, \beta)$ *fits into* $E \subset \mathbb{C}$ if for any $\varepsilon > 0$ we have

$$(1) \qquad \Delta_*(\{k \in \mathbb{Z}\colon \operatorname{dist}\{\omega k + \beta, E\} < \varepsilon\}) \ge \frac{9}{10}\,.$$

Note that (1) is phrased in terms of $\omega$ and $\beta$, but in fact depends only on $\Pi(\omega, \beta)$. (See (A.7)).

(2) **Lemma.** *If $E$, $E_j$, $\Pi(\omega_j, \beta_j)$ are as in the Deconstruction Lemma, then $\Pi(\omega_j, \beta_j)$ fits into $E$.*

PROOF. Fix $j' \neq j$. For small enough $\varepsilon > 0$ we will estimate

$$(3) \qquad \Delta_\varepsilon(j,j') = \Delta^*(\{k \in \mathbb{Z}\colon \operatorname{dist}\{\omega_j k + \beta_j,\ \Pi(\omega_{j'},\beta_{j'})\} < \varepsilon\}).$$

Let $\ell_j$,$\ell_{j'}$ denote the lines $\omega_j\mathbb{R} + \beta_j$, $\omega_{j'}\mathbb{R} + \beta_{j'}$ in $\mathbb{C}$. We distinguish several cases.

*CASE* 1: $\ell_j \neq \ell_{j'}$. Then $\operatorname{dist}\{x, \ell_{j'}\}$ is bounded below by a positive constant as $x \in \ell_j$ tends to infinity. Hence, $\Delta_\varepsilon(j,j') = 0$ for $\varepsilon > 0$ small enough.

*CASE* 2: $\ell_j = \ell_{j'}$ and $\omega_j/\omega_{j'}$ is irrational. Then by (A.12), we can make $\Delta_\varepsilon(j,j')$ arbitrarily small by taking $\varepsilon > 0$ small enough.

*CASE* 3: $\ell_j = \ell_{j'}$ and $\omega_j/\omega_{j'} = p/q$ in lowest terms, with $p, q \in \mathbb{Z}$ and $q > 0$. In view of (B.3), we have $q > 100D^2$.

Assume we are given distinct integers $k_1, k_2 \in \mathbb{Z}$ with

$$(4) \qquad \operatorname{dist}\{\omega_j k_1 + \beta_j,\ \Pi(\omega_{j'},\beta_{j'})\} < \varepsilon,$$

and

$$(5) \qquad \operatorname{dist}\{\omega_j k_2 + \beta_j,\ \Pi(\omega_{j'},\beta_{j'})\} < \varepsilon.$$

Thus, for integers $m_1$ and $m_2$, we have

$$(6) \qquad |(\omega_j k_1 + \beta_j) - (\omega_{j'} m_1 + \beta_{j'})| < \varepsilon,$$

and

$$(7) \qquad |(\omega_j k_2 + \beta_j) - (\omega_{j'} m_2 + \beta_{j'})| < \varepsilon.$$

Subtracting (6) from (7), and recalling that $\omega_j/\omega_{j'} = p/q$, we get

$$(8) \qquad \left|\frac{p}{q} - \frac{m_2 - m_1}{k_2 - k_1}\right| < \frac{2\varepsilon}{|\omega_{j'}|\,|k_2 - k_1|}.$$

If $p/q \neq (m_2 - m_1)/(k_2 - k_1)$, then

$$\left|\frac{p}{q} - \frac{m_2 - m_1}{k_2 - k_1}\right| \geq \frac{1}{q\,|k_2 - k_1|},$$

which contradicts (8) provided we take $\varepsilon < |\omega_{j'}|/(2q)$. Hence, $p/q = (m_2 - m_1)/(k_2 - k_1)$. Since $p/q$ is in lowest terms, it follows that $k_2 - k_1$

is a multiple of $q$. So we have shown that (4), (5) imply $k_2 \equiv k_1 \bmod q$. It follows at once that $\Delta_\varepsilon(j,j') \le 1/q$, for $\varepsilon > 0$ small enough.

Since $q > 100\,D^2$ in *CASE* 3, our analysis of the above cases gives $\Delta_\varepsilon(j,j') < 1/(100\,D^2)$ for $j' \neq j$, if $\varepsilon > 0$ is small enough.

Summing over all $j' \neq j$ and recalling (3), we get

$$\Delta^*(\{k \in \mathbb{Z}\colon \operatorname{dist}\{\omega_j k + \beta_j, \Pi(\omega_{j'}, \beta_{j'})\} < \varepsilon \text{ for some } j' \neq j\}) < \frac{1}{100\,D}\,. \tag{9}$$

Now suppose $k \in \mathbb{Z}$ satisfies

$$\operatorname{dist}\{\omega_j k + \beta_j,\, \Pi(\omega_{j'}, \beta_{j'})\} > \varepsilon\,, \qquad \text{for all } j' \neq j\,. \tag{10}$$

Since $E_j$ approximates $\Pi(\omega_j, \beta_j)$, we can find $x_k^j \in E_j$ for all but finitely many $k$ so that

$$|x_k^j - (\omega_j k + \beta_j)| \le \frac{\varepsilon}{10}\,. \tag{11}$$

In particular, $|x_k^j| \to \infty$ as $k \to \infty$.

On the other hand, since $E_{j'}$ approximates $\Pi(\omega_{j'}, \beta_{j'})$, we have $E_{j'} \subset F_{j'} \cup \{z \in \mathbb{C}\colon \operatorname{dist}\{z,\, \Pi(\omega_{j'}, \beta_{j'})\} < \varepsilon/10\}$ with $F_{j'}$ finite. Hence, (10) implies

$$\operatorname{dist}\{\omega_j k + \beta_j, E_{j'}\} \ge \frac{9}{10}\,\varepsilon\,, \qquad \text{for all } j' \neq j\,, \tag{12}$$

for all but finitely many $k$. Comparing (11) and (12), we see that $x_k^j \notin E_{j'}$, $j' \neq j$. Thus, all but finitely many $k$ satisfying (10) have the property $x_k^j \in E_j \setminus \cup_{j' \neq j} E_{j'}$. Since $E$ is the approximate union of $E_1, \ldots, E_D$, it follows that $x_k^j \in E$. Hence, (11) implies

$$\operatorname{dist}\{\omega_j k + \beta_j, E\} < \varepsilon\,. \tag{13}$$

So (13) holds for all but finitely many of the $k$ that satisfy (10). Therefore, by (9), we have $\Delta_*(\{k \in \mathbb{Z}\colon \operatorname{dist}\{\omega_j k + \beta_j, E\} < \varepsilon\}) \ge 1 - 1/(100\,D)$, which shows that $\Pi(\omega_j, \beta_j)$ fits into $E$.

(14) **Lemma.** *Let* $\Pi(\omega,\beta)$ *and* $\Pi(\omega',\beta')$ *be arithmetic progressions, with* $\Pi(\omega,\beta) \not\subset \Pi(\omega',\beta')$. *Fix* $D > 0$, *and define*

$$q(\omega,\omega') = \begin{cases} q\,, & \text{if } \dfrac{\omega}{\omega'} = \dfrac{p}{q} \text{ in lowest terms, with } p, q \in \mathbb{Z},\ q \ge 2\,. \\ 100\,D^2\,, & \text{if } \dfrac{\omega}{\omega'} \text{ is irrational, an integer, or non-real.} \end{cases} \tag{15}$$

*Then*

$$\Delta^*(\{k \in \mathbb{Z}: \operatorname{dist}\{\omega k + \beta, \Pi(\omega', \beta')\} < \varepsilon\}) \leq \frac{1}{q(\omega, \omega')}, \tag{16}$$

*for $\varepsilon > 0$ small enough.*

PROOF. As in the previous lemma, we set $\ell = \omega\mathbb{R} + \beta$, $\ell' = \omega'\mathbb{R} + \beta'$, and we distinguish several cases.

*CASE* 1: $\ell \neq \ell'$. As in the proof of the previous lemma, the left-hand side of (16) is equal to zero if $\varepsilon$ is small enough.

*CASE* 2: $\ell = \ell'$ and $\omega/\omega'$ irrational. As in the proof of the previous lemma, we can make the left-hand side of (16) arbitrarily small by taking $\varepsilon$ small enough.

*CASE* 3: $\ell = \ell'$ and $\omega/\omega' = p/q$ in lowest terms, with $p, q \in \mathbb{Z}$ and $q \geq 2$. As in the proof of the previous lemma,

$$\operatorname{dist}\{\omega k_1 + \beta, \Pi(\omega', \beta')\} < \varepsilon, \quad \operatorname{dist}\{\omega k_2 + \beta, \Pi(\omega', \beta')\} < \varepsilon,$$

imply $k_2 = k_1 \bmod q$, so that (16) is obvious.

*CASE* 4: $\ell = \ell'$ and $\omega/\omega' = p$ for some integer $p$. Then $\beta' - \beta$ is not a multiple of $\omega'$, since $\Pi(\omega, \beta) \not\subset \Pi(\omega', \beta')$.
Take $\varepsilon < \min_{k \in \mathbb{Z}} |\beta' - \beta - k\omega'|$. Then for all $k, m \in \mathbb{Z}$ we have

$$|(\omega k + \beta) - (\omega' m + \beta')| = |\beta - \beta' - (m - pk)\omega'| > \varepsilon,$$

so that $\operatorname{dist}\{\omega k + \beta, \Pi(\omega', \beta')\} > \varepsilon$, and the left-hand side of (16) equals zero.

(17) **Lemma.** *Assume the hypotheses of the Deconstruction Lemma, and suppose*

$$|\omega_1| < |\omega_2| < \cdots < |\omega_D|. \tag{18}$$

*Fix an integer $s$, $1 \leq s \leq D$. Let $\Pi(\omega, \beta)$ be an arithmetic progression with the following properties:*

(19) $\Pi(\omega, \beta)$ *fits into* $E$,

(20) $\omega \neq p\omega_j/q$ *for* $p, q \in \mathbb{Z}$, $1 \leq q \leq 10\,s$, *if* $j < s$.

*Then either* $\Pi(\omega,\beta)=\Pi(\omega_s,\beta_s)$*, or else* $|\omega|>|\omega_s|$.

PROOF. Assume the lemma is false. Thus,

$$\Pi(\omega,\beta)\neq\Pi(\omega_s,\beta_s)\,, \tag{21}$$

$$|\omega|\leq|\omega_s|\,. \tag{22}$$

Suppose for the moment that $\Pi(\omega,\beta)\subset\Pi(\omega_j,\beta_j)$ for some $j$, $1\leq j\leq D$. Then

$$\omega=p\,\omega_j\,,\qquad\text{for an integer } p\neq 0\,. \tag{23}$$

If $j<s$, then (23) contradicts (20). Hence, $j\geq s$ and (18), (23) yield

$$|\omega|=|p|\,|\omega_j|\geq|\omega_j|\geq|\omega_s|\,. \tag{24}$$

Moreover, at least one of the inequalities in (24) will be strict, unless $p=\pm1$ and $j=s$. Therefore, (22) yields $p=\pm1$ and $j=s$. Since $\Pi(\omega,\beta)\subset\Pi(\omega_j,\beta_j)=\Pi(\omega_s,\beta_s)$ and $\omega=p\,\omega_j=\pm\,\omega_s$, it follows that $\Pi(\omega,\beta)=\Pi(\omega_s,\beta_s)$, contradicting (21). This contradiction proves that

$$\Pi(\omega,\beta)\not\subset\Pi(\omega_j,\beta_j)\,,\qquad\text{for } 1\leq j\leq D\,. \tag{25}$$

Next, we apply (19) and (1) to conclude that

$$\Delta_*(\{k\in\mathbb{Z}\colon \operatorname{dist}\{\omega k+\beta,E\}<\varepsilon\})\geq\frac{9}{10}\,,\qquad\text{for any } \varepsilon>0\,. \tag{26}$$

Since $E$ is the approximate union of $E_1,\dots,E_D$, we have $E\subset E_1\cup\cdots\cup E_D$, so that $\operatorname{dist}\{\omega k+\beta,E\}<\varepsilon$ implies $\operatorname{dist}\{\omega k+\beta,E_j\}<\varepsilon$ for some $j$. Hence, (26) yields

$$\sum_{j=1}^{D}\Delta^*(\{k\in\mathbb{Z}\colon\ \operatorname{dist}\{\omega k+\beta,E_j\}<\varepsilon\})\geq\frac{9}{10}\,,\quad\text{for any } \varepsilon>0\,. \tag{27}$$

Moreover, each $E_j$ approximates $\Pi(\omega_j,\beta_j)$. Hence, given $\varepsilon>0$, we have

$$E_j\subset F_j\cup\{z\in\mathbb{C}\colon\ \operatorname{dist}\{z,\Pi(\omega_j,\beta_j)\}<\varepsilon\}$$

with $F_j$ finite. So, for large integers $k$, $\operatorname{dist}\{\omega k+\beta,E_j\}<\varepsilon$ implies $\operatorname{dist}\{\omega k+\beta,\Pi(\omega_j,\beta_j)\}<2\varepsilon$. Thus, (27) implies

$$\sum_{j=1}^{D}\Delta^*(\{k\in\mathbb{Z}\colon\ \operatorname{dist}\{\omega k+\beta,\Pi(\omega_j,\beta_j)\}<2\,\varepsilon\})\geq\frac{9}{10}\,, \tag{28}$$

for any $\varepsilon > 0$. Taking $\varepsilon$ small enough, and using (25) to apply Lemma (14), we obtain

$$\Delta^*(\{k \in \mathbb{Z}\colon \operatorname{dist}\{\omega k + \beta, \Pi(\omega_j, \beta_j)\} < 2\varepsilon\}) \le \frac{1}{q(\omega, \omega_j)}, \tag{29}$$

where

$$q(\omega, \omega_j) = \begin{cases} q, \text{ if } \dfrac{\omega}{\omega_j} = \dfrac{p}{q} \text{ in lowest terms, with } p, q \in \mathbb{Z},\ q \ge 2, \\ 100\, D^2, \text{ otherwise.} \end{cases} \tag{30}$$

From (28), (29), we obtain

$$\sum_{j=1}^{D} \frac{1}{q(\omega, \omega_j)} \ge \frac{9}{10}. \tag{31}$$

On the other hand, we can prove an upper bound for the left-hand side of (31). Immediately from (20) and (30), we have $q(\omega, \omega_j) \ge 10\, s$ for $j < s$, so that

$$\sum_{1 \le j < s} \frac{1}{q(\omega, \omega_j)} \le \frac{1}{10}. \tag{32}$$

To control $q(\omega, \omega_j)$ for $j \ge s$, we prove

$$q(\omega, \omega_j) \ge 2, \tag{33}$$

and

$$q(\omega, \omega_j) \le 10\, D, \qquad \text{for at most one } j \ge s. \tag{34}$$

In fact, (33) is immediate from (30). If (34) were false, then we would have

$$\frac{\omega}{\omega_j} = \frac{p}{q}, \qquad \frac{\omega}{\omega_{j'}} = \frac{p'}{q'}, \tag{35}$$

with $p, q, p', q' \in \mathbb{Z}$, $j, j' \ge s$, $j \ne j'$, $1 \le q, q' \le 10\, D$. From (18), (22), (35) we have $|\omega| \le |\omega_s| \le |\omega_j|$ so that $|p| \le |q|$; and similarly, $|p'| \le |q'|$. In particular,

$$0 < |p|, |p'|, |q|, |q'| \le 10\, D, \tag{36}$$

by another application of (35). A final application of (35) gives

$$\frac{\omega_j}{\omega_{j'}} = \frac{q\,p'}{p\,q'} \equiv \frac{P}{Q}\,. \tag{37}$$

Since $1 \le |Q| \le 100\,D^2$ by (36), equation (37) contradicts hypothesis (B.3) of the Deconstruction Lemma. This contradiction proves (34). Immediately from (33), (34), we obtain

$$\sum_{s \le j \le D} \frac{1}{q(\omega,\omega_j)} \le \frac{1}{2} + \frac{D-s}{10\,D} \le \frac{1}{2} + \frac{1}{10}\,. \tag{38}$$

Together, (32) and (38) yield

$$\sum_{j=1}^{D} \frac{1}{q(\omega,\omega_j)} \le \frac{7}{10}\,,$$

contradicting (31). Thus, assuming our lemma to be false, we arrived at a contradiction.

(39) **Lemma.** *Assume the hypotheses of the Deconstruction Lemma. Then there is no arithmetic progression $\Pi(\omega,\beta)$ with the following properties:*

(40) $\Pi(\omega,\beta)$ *fits into* $E$,

(41) $\omega \ne \frac{p}{q}\,\omega_j$ *for* $p, q \in \mathbb{Z}$, $1 \le q \le 10\,D$, *whenever* $1 \le j \le D$.

PROOF. Assume $\Pi(\omega,\beta)$ satisfies (40), (41). By (40) and (1), we have $\Delta_*(\{k \in \mathbb{Z}\colon \operatorname{dist}\{\omega k + \beta, E) < \varepsilon\}\} \ge 9/10$, for any $\varepsilon > 0$. As in the proof of the previous lemma, this implies that

$$\sum_{j=1}^{D} \Delta^*(\{k \in \mathbb{Z}\colon \operatorname{dist}\{\omega k + \beta, \Pi(\omega_j,\beta_j)\} < 2\,\varepsilon\}) \ge \frac{9}{10}\,, \tag{42}$$

for any $\varepsilon > 0$. Moreover, (41) shows that $\Pi(\omega,\beta) \not\subset \Pi(\omega_j,\beta_j)$ for any $j$, $1 \le j \le D$, so that Lemma (14) applies. We obtain from (14) and (42) the estimate

$$\sum_{j=1}^{D} \frac{1}{q(\omega,\omega_j)} \ge \frac{9}{10}\,, \tag{43}$$

with

$$(44)\quad q(\omega,\omega_j) = \begin{cases} q\,, & \text{if } \dfrac{\omega}{\omega_j} = \dfrac{p}{q} \text{ in lowest terms, with } p, q \in \mathbb{Z},\ q \geq 2\,, \\ 100\,D^2\,, & \text{otherwise.} \end{cases}$$

On the other hand, (41) and (44) imply $q(\omega,\omega_j) \geq 10\,D$ for all $j$, so that

$$\sum_{j=1}^{D} \frac{1}{q(\omega,\omega_j)} \leq \frac{1}{10}\,,$$

contradicting (43).

## D. Proving the Deconstruction Lemma.

Let $E$, $E_1,\dots,E_D$, $\widetilde{E}_1,\dots,\widetilde{E}_{\widetilde{D}}$, $\Pi(\omega_j,\beta_j)$, $\Pi(\widetilde{\omega}_j,\tilde{\beta}_j)$ be as in the Deconstruction Lemma. Hypothesis (B.3) shows that the $|\omega_j|$ are all distinct. Without loss of generality, we may therefore permute the $E_j$ and $\Pi(\omega_j,\beta_j)$ to reduce matters to the case

$$(1)\qquad |\omega_1| < \cdots < |\omega_D|\,.$$

Similarly, we may assume

$$(2)\qquad |\widetilde{\omega}_1| < \cdots < |\widetilde{\omega}_{\widetilde{D}}|\,.$$

Also, we may assume

$$(3)\qquad D \leq \widetilde{D}\,.$$

For the rest of the proof, we will assume (1), (2), (3). We will prove that

$$(4)\qquad \Pi(\omega_j,\beta_j) = \Pi(\widetilde{\omega}_j,\tilde{\beta}_j)\,, \qquad \text{for } 1 \leq j \leq D\,.$$

To see this, fix $s$, $1 \leq s \leq D$, and suppose

$$(5)\qquad \Pi(\omega_j,\beta_j) = \Pi(\widetilde{\omega}_j,\tilde{\beta}_j)\,, \qquad \text{for } 1 \leq j < s\,.$$

(This assumption is vacuous for $s = 1$). We will see that (5) implies

$$(6)\qquad \Pi(\omega_s,\beta_s) = \Pi(\widetilde{\omega}_s,\tilde{\beta}_s)\,.$$

In fact, (5) and (A.7) show that

(7) $$\omega_j = \pm\widetilde{\omega}_j \quad \text{for} \quad 1 \le j < s\,.$$

The analogue of Lemma C.2 for the $\widetilde{E}_j$ and $\Pi(\widetilde{\omega}_j, \tilde{\beta}_j)$ shows that

(8) $$\Pi(\widetilde{\omega}_s, \tilde{\beta}_s) \quad \text{fits into } E\,.$$

Since $s \le \min\{D, \widetilde{D}\}$, equation (7) and hypothesis (B.6) show that

(9) $$\widetilde{\omega}_s \ne \frac{p}{q}\,\omega_j\,, \qquad \text{for } 1 \le j < s,\ p, q \in \mathbb{Z},\ 1 \le q \le 10\,s\,.$$

Conditions (8), (9) are the hypotheses of Lemma (C.17), which tells us that

(10) $$\text{either} \quad \Pi(\widetilde{\omega}_s, \tilde{\beta}_s) = \Pi(\omega_s, \beta_s)\,, \quad \text{or else} \quad |\widetilde{\omega}_s| > |\omega_s|\,.$$

The same argument works with the rôles of the $\Pi(\omega_j, \beta_j)$ and $\Pi(\widetilde{\omega}_j, \tilde{\beta}_j)$ interchanged, so we have also

(11) $$\text{either} \quad \Pi(\omega_s, \beta_s) = \Pi(\widetilde{\omega}_s, \tilde{\beta}_s)\,, \quad \text{or else} \quad |\omega_s| > |\widetilde{\omega}_s|\,.$$

Since at least one of the inequalities $|\omega_s| > |\widetilde{\omega}_s|$, $|\widetilde{\omega}_s| > |\omega_s|$ must be false, (10) and (11) imply (6). Thus, (5) implies (6), completing the proof of (4).

Next we show that

(12) $$D = \widetilde{D}\,.$$

If (12) were false, then by (3) we would have

(13) $$\widetilde{D} \ge D + 1\,.$$

By the analogue of Lemma C.2 for the $\widetilde{E}_j$ and $\Pi(\widetilde{\omega}_j, \tilde{\beta}_j)$, we know that

(14) $$\Pi(\widetilde{\omega}_{D+1}, \tilde{\beta}_{D+1}) \quad \text{fits into } E\,.$$

Also, by (4) and (A.7) we have

(15) $$\widetilde{\omega}_j = \pm\,\omega_j\,, \qquad \text{for } 1 \le j \le D\,.$$

Hence, hypothesis (B.6) shows that

$$\widetilde{\omega}_{D+1} \neq \frac{p}{q}\omega_j\,, \qquad \text{for } 1 \le j \le D,\ p,q \in \mathbb{Z},\ 1 \le q \le 10\,D\,. \tag{16}$$

Together, conditions (14) and (16) contradict Lemma C.39. This contradiction completes the proof of (12).

Next, fix $j_0$, $1 \le j_0 \le D$. We will construct a sequence $(x_\nu)_{\nu \ge 1}$ such that

$$|x_\nu| \to \infty\,, \tag{17}$$

$$x_\nu \in E_{j_0} \setminus \bigcup_{j \neq j_0} E_j\,, \tag{18}$$

$$x_\nu \in \widetilde{E}_{j_0} \setminus \bigcup_{j \neq j_0} \widetilde{E}_j\,. \tag{19}$$

To construct $(x_\nu)$ with these properties, we first note that

$$\Pi(\omega_{j_0}, \beta_{j_0}) \not\subset \Pi(\omega_j, \beta_j)\,, \qquad \text{for } j \neq j_0\,, \tag{20}$$

since $\omega_{j_0}$ is not an integer multiple of $\omega_j$. Hence for $\varepsilon > 0$ small enough, Lemma C.14 shows that

$$\Delta^*(\{k \in \mathbb{Z}\colon\ \operatorname{dist}\{\omega_{j_0}k + \beta_{j_0}, \Pi(\omega_j,\beta_j)\} < \varepsilon\}) \le \frac{1}{q(\omega_{j_0},\omega_j)}\,, \tag{21}$$

where

$$q(\omega_{j_0},\omega_j) = \begin{cases} q\,, \text{ if } \dfrac{\omega_{j_0}}{\omega_j} = \dfrac{p}{q} \text{ in lowest terms, with } p,q \in \mathbb{Z},\ q \ge 2\,, \\ 100\,D^2\,, \text{ otherwise}\,. \end{cases}$$

Hypothesis (B.3) shows that $q(\omega_{j_0},\omega_j) \ge 100D^2$ for $j \neq j_0$, so that (21) implies

$$\Delta^*(\{k \in \mathbb{Z}\colon\ \operatorname{dist}\{\omega_{j_0}k + \beta_{j_0}, \Pi(\omega_j,\beta_j)\} < \varepsilon\}) \le \frac{1}{100\,D^2}\,, \quad \text{for } j \neq j_0\,.$$

Summing over $j$, we obtain

$$\begin{aligned} &\Delta^*(\{k \in \mathbb{Z}\colon\ \operatorname{dist}\{\omega_{j_0}k + \beta_{j_0}, \Pi(\omega_j,\beta_j)\} < \varepsilon\,, \text{ for some } j \neq j_0\}) \\ &\qquad \le \frac{1}{100\,D}\,. \end{aligned} \tag{22}$$

Let $K = \{k \in \mathbb{Z}\colon \operatorname{dist}\{\omega_{j_0}k + \beta_{j_0}, \Pi(\omega_j, \beta_j)\} \geq \varepsilon$ for all $j \neq j_0\}$. Recall that $E_j$ approximates $\Pi(\omega_j, \beta_j)$, so that

$$E_j \subset F_j \cup \{z \in \mathbb{C}\colon \operatorname{dist}\{z, \Pi(\omega_j, \beta_j)\} < \varepsilon/3\}$$

with $F_j$ finite. Therefore, for all but finitely many $k \in \mathbb{Z}$ we have

$$\operatorname{dist}\{\omega_{j_0}k + \beta_{j_0}, E_j\} < \frac{2\varepsilon}{3} \quad \text{implies} \quad \operatorname{dist}\{\omega_{j_0}k + \beta_{j_0}, \Pi(\omega_j, \beta_j)\} < \varepsilon\,.$$

It follows that

$$\operatorname{dist}\{\omega_{j_0}k + \beta_{j_0}, E_j\} \geq \frac{2}{3}\varepsilon\,, \qquad \text{for all } j \neq j_0\,, \tag{23}$$

for all but finitely many $k \in K$.

Similarly, $\widetilde{E}_j$ approximates $\Pi(\widetilde{\omega}_j, \tilde{\beta}_j) = \Pi(\omega_j, \beta_j)$ by (4) and (12), so the proof of (23) yields also

$$\operatorname{dist}\{\omega_{j_0}k + \beta_{j_0}, \widetilde{E}_j\} \geq \frac{2}{3}\varepsilon\,, \qquad \text{for all } j \neq j_0\,, \tag{24}$$

for all but finitely many $k \in K$.

On the other hand, $E_{j_0}$ approximates $\Pi(\omega_{j_0}, \beta_{j_0})$. Hence, for all but finitely many $k \in K$ we can find

(25) $\hat{x}_k \in E_{j_0}$ satisfying

(26) $|\hat{x}_k - (\omega_{j_0}k + \beta_{j_0})| < \dfrac{\varepsilon}{3}$.

Comparing (26) with (23), (24), we conclude that all but finitely many $k \in K$ satisfy

(27) $\hat{x}_k \notin E_j$ for $j \neq j_0$, and

(28) $\hat{x}_k \notin \widetilde{E}_j$ for $j \neq j_0$.

Since $E$ is the approximate union of $E_1, \ldots, E_D$, we know from (25) and (27) that $\hat{x}_k \in E$. This in turn gives $\hat{x}_k \in \widetilde{E}_1 \cup \cdots \cup \widetilde{E}_{\widetilde{D}}$, since $E$ is the approximate union of the $\widetilde{E}_j$. In view of (28), we obtain

$$\hat{x}_k \in \widetilde{E}_{j_0}\,, \tag{29}$$

for any $k$ satisfying (25)-(28).

Thus, for all but finitely many $k \in K$, the following hold:

(30) $|\hat{x}_k - (\omega_{j_0} k + \beta_{j_0})| < \varepsilon$,

(31) $\hat{x}_k \in E_{j_0} \setminus \bigcup_{j \neq j_0} E_j$,

(32) $\hat{x}_k \in \widetilde{E}_{j_0} \setminus \bigcup_{j \neq j_0} \widetilde{E}_j$.

Finally, let $(x_\nu)_{\nu \geq 1}$ be an enumeration of the $\hat{x}_k$ for $k \in K$ that satisfy (30), (31), (32). Estimate (22) and the definition of $K$ show that there are indeed infinitely many such $\hat{x}_k$, so we get an infinite sequence. Estimate (30) shows that $|x_\nu| \to \infty$ as $\nu \to \infty$. Hence, (17), (18), (19) follow at once from (30), (31), (32). We have proven the conclusions of the Deconstruction Lemma, with $\gamma =$ identity.

We conclude this section with a simple special case of the Deconstruction Lemma.

(33) **Corollary.** *Suppose $E_j$ approximates $\Pi(\omega_j, \beta_j)$ for $1 \leq j \leq D$. Assume that $|\omega_j/\omega_{j'}| \neq p/q$ for $j \neq j'$, $p, q \in \mathbb{Z}$, $1 \leq q \leq 100\, D^2$. Then for each $j_0$, $1 \leq j_0 \leq D$, we can find a sequence $(x_\nu)_{\nu \geq 1}$ of complex numbers, such that*

(34) *$|x_\nu| \to \infty$ as $\nu \to \infty$, and*

(35) *$x_\nu \in E_{j_0} \setminus \bigcup_{j \neq j_0} E_j$ for each $\nu$.*

PROOF. Set $\widetilde{E}_j = E_j$, $\Pi(\widetilde{\omega}_j, \tilde{\beta}_j) = \Pi(\omega_j, \beta_j)$, $E = E_1 \cup \cdots \cup E_D$. Then the Deconstruction Lemma applies, and it gives a sequence $(x_\nu)_{\nu \geq 1}$ satisfying (34) and (35).

## III. Analytic Continuation of Neural Nets.

### A. Preliminaries.

Let $\mathcal{N} = [(D_0, \ldots, D_L), (\omega_{jk}^\ell), (\theta_j^\ell)]$ be a neural net. We will show that the functions $x_j^\ell(t, \mathcal{N})$, defined initially for $t$ real, continue analytically to an open subset of $\mathbb{C}$ with countable complement. We will analyze the largest domain $\Omega$ to which we can analytically continue the output $x_1^L(t, \mathcal{N})$. The point-set topology of $\Omega$ leads us to define a hierarchy of singular sets $\mathrm{Sing}(\ell, \mathcal{N})$ in the complex plane. The sets $\mathrm{Sing}(\ell, \mathcal{N})$ are defined entirely in terms of the output function $t \mapsto x_1^L(t, \mathcal{N})$ $(t \in \mathbb{R})$, yet they carry a lot of information on the architecture, weights and thresholds of $\mathcal{N}$.

We begin our discussion with a simple, general result on analytic functions defined in the complement of a countable set.

(1) **Lemma.** *Let $f(t)$ be a function on $\mathbb{R}$, and suppose that $f$ continues analytically to an open set $\Omega \subset \mathbb{C}$, with $\mathbb{R} \subset \Omega$ and $\mathbb{C} \setminus \Omega$ countable. Then there is one and only one open set $\Omega_* \subset \mathbb{C}$ with the following properties:*

(2) $\mathbb{R} \subset \Omega_*$,

(3) $\mathbb{C} \setminus \Omega_*$ *is countable,*

(4) *Let $\Omega' \subset \mathbb{C}$ be any connected open set that meets $\mathbb{R}$. Then $f$ continues analytically into $\Omega'$ if and only if $\Omega' \subset \Omega_*$.*

*We call $\Omega_*$ the natural domain of $f$.*

PROOF. We start with the following remark.

(5) Suppose $\Omega_1, \Omega_2 \subset \mathbb{C}$ are open sets, with $\Omega_1$ connected and $\mathbb{C} \setminus \Omega_2$ countable. Let $F_1$, $F_2$ be analytic on $\Omega_1$, $\Omega_2$ respectively, and assume $F_1 = F_2$ to infinite order at some point of $\Omega_1 \cap \Omega_2$. Then $F_1 = F_2$ on all of $\Omega_1 \cap \Omega_2$.

Indeed, (5) is immediate from the fact that $\Omega_1 \cap \Omega_2$ is the complement of a countable set in $\Omega_1$, and thus $\Omega_1 \cap \Omega_2$ is connected.

Now let $W$ be the collection of all open sets $\Omega' \subset \mathbb{C}$ such that $\mathbb{R} \subset \Omega'$, $\mathbb{C} \setminus \Omega'$ is countable, and $f$ continues analytically to $\Omega'$. If $\Omega'$, $\Omega'' \in W$, and if $F$, $G$ denote the analytic continuations of $f$ to $\Omega'$, $\Omega''$ respectively, then $F = G$ in $\Omega' \cap \Omega''$ by (5). It follows that $f$ continues analytically to $\Omega_* = \cup_{\Omega' \in W} \Omega'$. Since $\Omega \in W$ by hypothesis, properties (2) and (3) are obvious, and we know that

(6) $f$ continues analytically to any open set $\Omega' \subset \Omega_*$.

Next, suppose $\Omega' \subset \mathbb{C}$ is open and connected, and meets $\mathbb{R}$; and assume $f$ continues analytically to an analytic function $G$ on $\Omega'$. Let $F$ denote the analytic continuation of $f$ to $\Omega_*$. Then $F = G$ on $\Omega' \cap \Omega_*$ by (5), so that $f$ continues analytically to $\Omega' \cup \Omega_*$. We have shown that

(7) If $\Omega' \subset \mathbb{C}$ is open, connected and meets $\mathbb{R}$, and if $f$ continues analytically to $\Omega'$, then $\Omega' \subset \Omega_*$.

Assertions (6) and (7) complete the proof of (4). It remains only to prove the uniqueness of $\Omega_*$. Thus, suppose $\Omega_*^1$ and $\Omega_*^2$ both have properties (2), (3), (4). Then $\Omega_*^1$ is an open, connected set that meet $\mathbb{R}$, and $f$ continues analytically to $\Omega_*^1$. Since $\Omega_*^2$ has property (4), it follows that $\Omega_*^1 \subset \Omega_*^2$. Similarly, $\Omega_*^2 \subset \Omega_*^1$, which proves that $\Omega_*$ is unique.

(8) **Definition.** *Let $f$ be a function on $\mathbb{R}$. If $f$ continues analytically to an open set with countable complement, then we define the sets* $\mathrm{Sing}(\ell, f) \subset \mathbb{C}$ *for* $\ell \geq 0$ *by the following induction:*

(9) $\mathrm{Sing}(0, f)$ *is the complement of the natural domain of* $f$,

(10) $\mathrm{Sing}(\ell+1, f)$ *is the set of accumulation points of* $\mathrm{Sing}(\ell, f)$.

We will take $f$ to be the output of a neural net. The next two lemmas help us to show that $f$ continues analytically to an open set with countable complement, and to understand the sets $\mathrm{Sing}(\ell, f)$.

(11) **Lemma.** *Let $F$ be analytic on a connected, open set $\Omega \subset \mathbb{C}$. Let $\Pi(\omega, \beta)$ be an arithmetic progression. Suppose that $F$ is either nonconstant, or else identically equal to a constant not belonging to $\Pi(\omega, \beta)$. Then the set $E = \{t \in \Omega\colon\ F(t) \in \Pi(\omega, \beta)\}$ has no accumulation points in $\Omega$. In particular, $E$ is countable.*

PROOF. Suppose $t_\nu \to t_*$ as $\nu \to \infty$, with $t_\nu \in E$ and $t_* \in \Omega$. Then $F(t_\nu) \to F(t_*)$ and $F(t_\nu) \in \Pi(\omega, \beta)$. It follows that $F(t_\nu)$ is eventually constant: $F(t_\nu) = b$ for all $\nu \geq \nu_0$, with $b \in \Pi(\omega, \beta)$. Since $\{t_\nu\}$ accumulate at $t_*$ and $\Omega$ is connected, it follows in turn that $F(t) = b$ for all $t \in \Omega$, contradicting our hypothesis on $F$. Thus, $E$ has no accumulation points in $\Omega$. This implies that $E_N = \{t \in E\colon\ |t| \leq N \text{ and } \operatorname{dist}\{t, \mathbb{C} \setminus \Omega\} \geq 1/N\}$ is a bounded set without accumulation points. Hence $E_N$ is finite, so that $E = \cup_{N \geq 1} E_N$ is countable.

(12) **Lemma.** *Let $U$ be a disc centered at $z_0 \in \mathbb{C}$. Suppose $\Phi$ is meromorphic and $\Psi$ analytic in a neighborhood of $\bar{U}$. Assume $\Phi$ has a single pole at $z_0$ (not necessarily simple). Let $\Pi(\omega, \beta)$ be an arithmetic progression. Then the set*

$$E = \{\Phi(t) + \Psi(t)\colon\ t \in U \setminus \{z_0\} \quad \textit{and} \quad \Phi(t) \in \Pi(\omega, \beta)\} \tag{13}$$

*approximates the arithmetic progression*

$$\Pi(\omega, \beta + \Psi(z_0)). \tag{14}$$

PROOF. Suppose that $\Phi$ has a pole of order $m$ at $z_0$. Then for large enough $\zeta \in \mathbb{C}$, the solutions of

$$\Phi(z) = \zeta \tag{15}$$

are given by a Puiseux expansion. That is, the solutions to (15) are

$$z = H(\zeta^{-1/m}), \tag{16}$$

where $H$ is analytic in a neighborhood of the origin, and $\zeta^{-1/m}$ runs over all the $m^{\text{th}}$ roots of $\zeta^{-1}$. Also, $H(0) = z_0$ and $H'(0) \neq 0$ (see [H]).

Let $\varepsilon > 0$ be given. We have to verify

(17) All but finitely many $\xi \in E$ lie within distance $\varepsilon$ of $\Pi(\omega, \beta + \Psi(z_0))$, and

(18) All but finitely many $\xi \in \Pi(\omega, \beta + \Psi(z_0))$ lie within distance $\varepsilon$ of $E$.

Pick $\delta > 0$ so small that

$$|z - z_0| \leq \delta \quad \text{implies} \quad z \in U \text{ and } |\Psi(z) - \Psi(z_0)| < \varepsilon. \tag{19}$$

To verify (17), suppose

$$0 < |z - z_0| \leq \delta \quad \text{and} \quad \Phi(z) \in \Pi(\omega, \beta). \tag{20}$$

Then dist $\{\Phi(z) + \Psi(z), \Pi(\omega, \beta + \Psi(z_0))\} \leq |\Psi(z) - \Psi(z_0)| < \varepsilon$. On the other hand, $\Phi$ is analytic on a neighborhood of the closure of $\widehat{U} = \{z \in U \colon |z - z_0| > \delta\}$, and $\Phi$ is non-constant on $\widehat{U}$ since $\Phi$ has a pole at $z_0$. Hence, $\Phi$ is bounded on $\widehat{U}$, and $\{z \in \widehat{U}\colon\ \Phi(z) = \xi\}$ is finite for any $\xi$. It follows that $\{z \in \widehat{U}\colon\ \Phi(z) \in \Pi(\omega, \beta)\}$ is finite. Thus, all but finitely many points of $E$ arise as $\xi = \Phi(z) + \Psi(z)$ for some $z$ satisfying (20), and therefore satisfy dist $\{\xi, \Pi(\omega, \beta + \Psi(z_0))\} < \varepsilon$. This proves (17).

To verify (18), let $\xi \in \Pi(\omega, \beta + \Psi(z_0))$ be sufficiently large, and let $z = H((\xi - \Psi(z_0))^{-1/m})$ for any choice of the $m^{\text{th}}$ root. Thus, $z \in U \setminus \{z_0\}$ and $\Phi(z) = \xi - \Psi(z_0) \in \Pi(\omega, \beta)$, so that $\zeta = \Phi(z) + \Psi(z) = \xi - \Psi(z_0) + \Psi(z)$ belongs to $E$. Moreover, if $\xi$ is large enough, then $(\xi - \Psi(z_0))^{-1/m}$ will be so small that $|z - z_0| = |H((\xi - \Psi(z_0))^{-1/m}) - H(0)| < \delta$. Therefore, $|\Psi(z) - \Psi(z_0)| < \varepsilon$ by (19), so that $|\zeta - \xi| = |\Psi(z) - \Psi(z_0)| < \varepsilon$, and thus dist $\{\xi, E\} < \varepsilon$. This completes the proof of (18).

## B. Continuing the Output to the Complement of a Countable Set.

Fix a neural net $\mathcal{N} = [(D_0, \dots, D_L), (\omega_{jk}^\ell), (\theta_j^\ell)]$. Recall that $D_0 = D_L = 1$. By induction on $\ell$ $(0 \le \ell \le L)$ we will define for $1 \le j \le D_\ell$ a set $\Omega_j^\ell \subset \mathbb{C}$ and a function $x_j^\ell(t, \mathcal{N})$ on $\Omega_j^\ell$. For $\ell = 0$, we set

(1) $\Omega_1^0 = \mathbb{C}$, and

(2) $x_1^0(t, \mathcal{N}) = t\,.$

Assume we have defined the $\Omega_j^{\ell-1}$ and $x_j^{\ell-1}(t, \mathcal{N})$ for a fixed $\ell$, $1 \le \ell \le L$. Then set

$$\Omega_*^{\ell-1} = \bigcap_{1 \le j \le D_{\ell-1}} \Omega_j^{\ell-1}, \tag{3}$$

$$E_j^\ell = \left\{ t \in \Omega_*^{\ell-1} \colon \sum_{k=1}^{D_{\ell-1}} \omega_{jk}^\ell\, x_k^{\ell-1}(t, \mathcal{N}) + \theta_j^\ell \in \Pi(2\pi i, \pi i) \right\}, \tag{4}$$

$$\Omega_j^\ell = \Omega_*^{\ell-1} \setminus E_j^\ell\,, \tag{5}$$

and

$$x_j^\ell(t, \mathcal{N}) = \sigma\left( \sum_{k=1}^{D_{\ell-1}} \omega_{jk}^\ell\, x_k^{\ell-1}(t, \mathcal{N}) + \theta_j^\ell \right), \qquad \text{for } t \in \Omega_j^\ell\,. \tag{6}$$

Note that (6) makes sense because we need not evaluate $\sigma(\,\cdot\,)$ at one of its poles. The poles of $\sigma$ are precisely $\Pi(2\pi i, \pi i)$. Note also that $\mathbb{R} \subset \Omega_j^\ell$, since $\sum_{k=1}^{D_{\ell-1}} \omega_{jk}^\ell x_k^{\ell-1}(t, \mathcal{N}) + \theta_j^\ell$ is real for $t$ real, so that $\mathbb{R} \cap E_j^\ell = \varnothing$. For real $t$, our formulas (2) and (6) agree with the definition of $x_j^\ell(t, \mathcal{N})$ given in Section I. Hence we have extended the outputs of the nodes from $\mathbb{R}$ to subsets $\Omega_j^\ell$ of the complex plane.

Note that (3) leaves $\Omega_*^L$ undefined. We make the natural definition

$$\Omega_*^L = \Omega_1^L \tag{7}$$

(recall that $D_L = 1$ and compare with (3)). Our definitions have the obvious consequences

$$\Omega_*^\ell = \Omega_*^{\ell-1} \setminus \bigcup_{j=1}^{D_\ell} E_j^\ell\,, \qquad 1 \le \ell \le L\,, \tag{8}$$

and

$$\mathbb{C} \setminus \Omega_*^\ell = \bigcup_{1 \le \ell' \le \ell} \bigcup_{j=1}^{D_{\ell'}} E_j^{\ell'}, \qquad 1 \le \ell \le L. \tag{9}$$

(10) **Lemma.** *The following properties hold for* $1 \le \ell \le L$, $1 \le j \le D_\ell$.

(11) $\Omega_j^\ell$ *is open, and* $\mathbb{C} \setminus \Omega_j^\ell$ *is countable.*

(12) $E_j^\ell$ *is a countable subset of* $\Omega_*^{\ell-1}$, *with no accumulation points in* $\Omega_*^{\ell-1}$.

(13) $x_j^\ell(t, \mathcal{N})$ *is analytic on* $\Omega_j^\ell$.

(14) $x_j^\ell(t, \mathcal{N})$ *has poles at the points of* $E_j^\ell$.

PROOF. We use induction on $\ell$. Fix $\ell$, $1 \le \ell \le L$, and assume

(15) $\Omega_j^{\ell-1}$ is open, and $\mathbb{C} \setminus \Omega_j^{\ell-1}$ is countable, $1 \le j \le D_{\ell-1}$, and

(16) $x_j^{\ell-1}(t, \mathcal{N})$ is analytic on $\Omega_j^{\ell-1}$, $1 \le j \le D_{\ell-1}$.

Note that (15), (16) are obvious for $\ell = 1$ by (1), (2).

We will show that (15), (16) imply (11)-(14). This will imply Lemma (10). From (15), (16) and (3), we see that $\Omega_*^{\ell-1}$ is open, that $\mathbb{C} \setminus \Omega_*^{\ell-1}$ is countable, and that

$$X_j^\ell(t) = \sum_{k=1}^{D_{\ell-1}} \omega_{jk}^\ell x_j^{\ell-1}(t, \mathcal{N}) + \theta_j^\ell, \qquad 1 \le j \le D_\ell, \tag{17}$$

is analytic on $\Omega_*^{\ell-1}$. Moreover, $\mathbb{R} \subset \Omega_*^{\ell-1}$, and $X_j^\ell(t)$ is real for $t \in \mathbb{R}$. Hence, if $X_j^\ell(t)$ is constant on $\Omega_*^{\ell-1}$, then that constant is real. In particular, $X_j^\ell(t)$ is either non-constant on $\Omega_*^{\ell-1}$, or else identically equal to a constant not in $\Pi(2\pi i, \pi i)$. Therefore, by Lemma A.11 and (4), the set $E_j^\ell \subset \Omega_*^{\ell-1}$ is countable and has no accumulation points in $\Omega_*^{\ell-1}$. This proves (12), from which (11) follows at once by virtue of (5), since $\Omega_*^{\ell-1}$ is open and has countable complement. Assertion (13) follows from the formula $x_j^\ell(t, \mathcal{N}) = \sigma(X_j^\ell(t))$, since $X_j^\ell(t) \notin \Pi(2\pi i, \pi i)$ for $t \notin E_j^\ell$. To verify (14), let $t_0 \in E_j^\ell \subset \Omega_*^{\ell-1}$. By (12), we can find a disc $U_\delta = \{t \in \mathbb{C}: |t - t_0| < \delta\} \subset \Omega_*^{\ell-1}$ such that $U_\delta \setminus \{t_0\}$ does not meet $E_j^\ell$. Thus, $U_\delta \setminus \{t_0\} \subset \Omega_*^\ell$, so $X_j^\ell(t) \notin \Pi(2\pi i, \pi i)$ and $x_j^\ell(t, \mathcal{N}) = \sigma(X_j^\ell(t))$ for $t \in U_\delta \setminus \{t_0\}$. Moreover, $X_j^\ell(t)$ is analytic on $\Omega_*^{\ell-1}$, hence on $U_\delta$;

and we have $X_j^\ell(t_0) \in \Pi(2\pi i, \pi i)$ since $t_0 \in E_j^\ell$. These remarks show that $x_j^\ell(t, \mathcal{N})$ has a pole at $t_0$, proving (14). The proof of (11)-(14) is complete.

Lemma (10) shows in particular that the output of the neural net $t \mapsto x_1^L(t, \mathcal{N})$ continues analytically from $\mathbb{R}$ to an open subset of $\mathbb{C}$ with countable complement. Hence, the natural domain of $x_1^L(t, \mathcal{N})$ and the sequence of singular sets $\operatorname{Sing}(\ell, x_1^L(t, \mathcal{N}))$ are well-defined. We write $\operatorname{Sing}(\ell, \mathcal{N})$ for $\operatorname{Sing}(\ell, x_1^L(t, \mathcal{N}))$, and note that

(18) The sets $\operatorname{Sing}(\ell, \mathcal{N})$, $\ell \geq 0$, are determined completely by the output $t \mapsto x_1^L(t, \mathcal{N})$ $(t \in \mathbb{R})$ of the neural net $\mathcal{N}$.

In a similar spirit, we see at once from the definitions (1)-(7) that

(19) For each $\ell$, $1 \leq \ell \leq L$, the sets $\Omega_j^\ell$, $\Omega_*^\ell$, $E_j^\ell$ and the functions $x_j^\ell(t, \mathcal{N})$ are determined completely by the $D_{\ell'}$, $\omega_{jk}^{\ell'}$, $\theta_j^{\ell'}$ with $1 \leq \ell' \leq \ell$.

## C. The Structure of the Singular Sets.

In this section, we will study the sets $\operatorname{Sing}(\ell, \mathcal{N})$ associated to a neural net $\mathcal{N} = [(D_0, \ldots, D_L), (\omega_{jk}^\ell), (\theta_j^\ell)]$, in terms of the sets $\Omega_*^\ell$, $\Omega_j^\ell$, $E_j^\ell$ defined in the previous section.

(1) **Lemma.** $E_1^L \subset \operatorname{Sing}(0, \mathcal{N}) \subset \bigcup_{\ell=1}^{L} \bigcup_{j=1}^{D_\ell} E_j^\ell$.

PROOF. Let $\Omega_*$ be the natural domain of $x_1^L(t, \mathcal{N})$ and let $X(t)$ be the analytic continuation of $x_1^L(t, \mathcal{N})$ to $\Omega_*$. Lemma B.10 shows that $x_1^L(t, \mathcal{N})$ continues analytically from $\mathbb{R}$ to $\Omega_1^L$. Hence the defining property (A.4) for the natural domain tells us that

(2) $\Omega_1^L \subset \Omega_*$, and

(3) $X(t) = x_1^L(t, \mathcal{N})$ for $t \in \Omega_1^L$.

From (2), (B.7), (B.9) and (A.9), we get

$$\operatorname{Sing}(0, \mathcal{N}) = \mathbb{C} \setminus \Omega_* \subset \mathbb{C} \setminus \Omega_1^L = \mathbb{C} \setminus \Omega_*^L = \bigcup_{1 \leq \ell \leq L} \bigcup_{j=1}^{D_\ell} E_j^\ell ,$$

which is half of Lemma 1.

To verify the other half of Lemma 1, suppose $t_0 \in E_1^L \cap \Omega_*$. Since $E_1^L \subset \Omega_*^{L-1}$ and $E_1^L$ has no accumulation points in $\Omega_*^{L-1}$, we know that a small enough disc

$$U_\delta = \{t \in \mathbb{C}:\ |t - t_0| < \delta\}$$

is contained in $\Omega_*^{L-1}$, and that $U_\delta \setminus \{t_0\}$ does not meet $E_1^L$. Then (3) shows that $X(t) = x_1^L(t, \mathcal{N})$ in $U_\delta \setminus \{t_0\}$, while (B.14) gives

$$\lim_{\substack{t \to t_0 \\ t \neq t_0}} |x_1^L(t, \mathcal{N})| = \infty\,.$$

Hence, $\lim_{t\to t_0,\ t\neq t_0} |X(t)| = \infty$, which contradicts the fact that $X(t)$ is analytic on an open set $\Omega_*$ containing $t_0$. Therefore, $E_1^L \cap \Omega_*$ is empty, *i.e.* $E_1^L \subset \mathbb{C} \setminus \Omega_* = \mathrm{Sing}\,(0, \mathcal{N})$.

Next we study the $E_j^\ell$, as well as

(4) $$\overset{\circ}{E}{}_j^\ell = E_j^\ell \setminus \bigcup_{j' \neq j} E_{j'}^\ell\,.$$

To do so, we impose the following hypothesis on the weights $(\omega_{jk}^\ell)$.

(5) **Assumption.** *$\omega_{jk}^\ell \neq 0$, and for $j \neq j'$, the ratio $\omega_{jk}^\ell / \omega_{j'k}^\ell$ is not equal to any fraction of the form $p/q$ with $p$, $q$ integers and $1 \le q \le 100\, D_\ell^2$.*

Immediately from (B.1)-(B.4), we see that $E_j^1$ consists of all $t \in \mathbb{C}$ such that $\omega_{j1}^1 t + \theta_j^1 \in \Pi(2\pi i, \pi i)$. In other words,

(6) $$E_j^1 = \Pi\left(\frac{2\pi i}{\omega_{j1}^1}, \frac{\pi i - \theta_j^1}{\omega_{j1}^1}\right), \qquad 1 \le j \le D_1\,.$$

From (4), (5), (6) and Corollary II.D.33, we get

(7) $\overset{\circ}{E}{}_j^1$ is infinite. $1 \le j \le D$.

The following lemma shows how $E_j^{\ell+1}$ and $\overset{\circ}{E}{}_j^{\ell+1}$ look near a point of $\overset{\circ}{E}{}_j^{\ell}$.

(8) **Lemma.** *Fix* $t_0 \in \overset{\circ}{E}{}^{\ell}_{k_0}$, $1 \le \ell \le L-1$, $1 \le k_0 \le D_\ell$. *For* $\delta > 0$, *set*

$$U_\delta = \{t \in \mathbb{C}\colon\ |t - t_0| < \delta\}. \tag{9}$$

*If* $\delta$ *is small enough, then the following properties hold.*

(10) $U_{2\delta} \subset \Omega_*^{\ell-1}$.

(11) $U_{2\delta} \setminus \{t_0\} \subset \Omega^\ell_{k_0}$; *thus,* $x^\ell_{k_0}(t, \mathcal{N})$ *is analytic on* $U_{2\delta} \setminus \{t_0\}$, *with a pole at* $t_0$.

(12) $U_{2\delta} \subset \Omega^\ell_k$ *for* $k \ne k_0$; *thus* $x^\ell_k(t, \mathcal{N})$ *is analytic on* $U_{2\delta}$.

$$\begin{aligned}(U_\delta \setminus \{t_0\}) \cap E_j^{\ell+1} \\ = \left\{ t \in U_\delta \setminus \{t_0\}\colon \sum_{k=1}^{D_\ell} \omega_{jk}^{\ell+1} x_k^\ell(t, \mathcal{N}) + \theta_j^{\ell+1} \in \Pi(2\pi i, \pi i) \right\},\end{aligned} \tag{13}$$

*for* $1 \le j \le D_{\ell+1}$.

(14) *The set* $F^\ell_j = \{x^\ell_{k_0}(t, \mathcal{N})\colon t \in (U_\delta \setminus \{t_0\}) \cap E_j^{\ell+1}\}$ *approximates the arithmetic progression* $\Pi\left(2\pi i/\omega_{jk_0}^{\ell+1}, \beta_{jk_0}^{\ell+1}\right)$ *for some complex number* $\beta^{\ell+1}_{jk_0}$, $1 \le j \le D_{\ell+1}$.

(15) *For each* $j_0$, $1 \le j_0 \le D_{\ell+1}$, $t_0$ *is an accumulation point of* $\overset{\circ}{E}{}^{\ell+1}_{j_0}$.

PROOF. We know that $t_0 \in \overset{\circ}{E}{}^\ell_{k_0} \subset E^\ell_{k_0} \subset \Omega_*^{\ell-1}$ by (B.12), so (10) holds simply because $\Omega_*^{\ell-1}$ is open. Another application of (B.12) shows that $U_{2\delta} \setminus \{t_0\}$ meets none of the $E^\ell_k$ $1 \le k \le D_\ell$, if $\delta$ is small. Since $t_0 \in \overset{\circ}{E}{}^\ell_{k_0}$, it follows that $U_{2\delta} \setminus \{t_0\} \subset \Omega^\ell_{k_0}$ and $U_{2\delta} \subset \Omega^\ell_k$ for $k \ne k_0$, by (B.5) and (10). Therefore, (11) and (12) follow from (B.13), (B.14).

Next note that (11), (12) yield $U_{2\delta} \setminus \{t_0\} \subset \Omega^\ell_*$. Hence (13) follows at once from the definition (B.4).

We set $U = U_\delta$,

$$\Phi(t) = \sum_{k=1}^{D_\ell} \omega_{jk}^{\ell+1}\, x_k^\ell(t, \mathcal{N}) + \theta_j^{\ell+1},$$

$$\Psi(t) = -\sum_{\substack{1 \le k \le D_\ell \\ (k \ne k_0)}} \omega_{jk}^{\ell+1}\, x_k^\ell(t, \mathcal{N}) - \theta_j^{\ell+1},$$

$z_0 = t_0$, $\Pi(\omega,\beta) = \Pi(2\pi i, \pi i)$. Then (11) and (12) show that the hypotheses of Lemma A.12 are satisfied. In view of (13), that lemma shows that $\{\omega_{jk_0}^{\ell+1} x_{k_0}^{\ell}(t,\mathcal{N})\colon t \in (U \setminus \{t_0\}) \cap E_j^{\ell+1}\}$ approximates an arithemtic progression of the form $\Pi(2\pi i, \beta_j)$. This yields (14) at once.

It remains to verify (15). By (5), (14) and Corollary II.D.33, we can find a sequence $(x_\nu)_{\nu \geq 1}$ satisfying

(16) $|x_\nu| \to \infty\,, \quad \text{as} \;\; \nu \to \infty\,,$

(17) $x_\nu \in F_{j_0}\,,$

(18) $x_\nu \notin F_j\,, \quad \text{for} \;\; j \neq j_0\,, \quad 1 \leq j \leq D_{\ell+1}\,.$

By definition of $F_j$, (17) means that

(19) $x_\nu = x_{k_0}^{\ell}(t_\nu, \mathcal{N})\,, \quad \text{with}$

(20) $t_\nu \in (U_\delta \setminus \{t_0\}) \cap E_{j_0}^{\ell+1}\,.$

If we had $t_\nu \in E_j^{\ell+1}$ for some $n \neq j_0$, then (19), (20) would imply $x_\nu \in F_j$, contradicting (18). Hence $t_\nu \notin E_j^{\ell+1}$, $j \neq j_0$, so that (20) can be sharpened to

$$t_\nu \in (U_\delta \setminus \{t_0\}) \cap \mathring{E}_{j_0}^{\ell+1}\,. \tag{21}$$

Also, (11), (19) and (16) show that $t_\nu \to t_0$ as $\nu \to \infty$. Therefore, (21) shows that $t_0$ is an accumulation point of $\mathring{E}_{j_0}^{\ell+1}$, which is (15).

(22) **Corollary.** *The set $\mathring{E}_j^{\ell}$ is infinite, for $1 \leq \ell \leq L$, $1 \leq j \leq D_\ell$.*

PROOF. We use induction on $\ell$. For $\ell = 1$, the Corollary is already known (see (7)). If $\mathring{E}_{k_0}^{\ell}$ is non-empty, then (15) shows that $\mathring{E}_j^{\ell+1}$ must be infinite, completing the induction.

(23) **Corollary.** *The output function $x_1^L(t,\mathcal{N})$ is non-constant.*

PROOF. If $x_1^L(t,\mathcal{N})$ were constant, its natural domain would be all of $\mathbb{C}$, so that $\text{Sing}(0,\mathcal{N})$ would be empty. However, we know that $\mathring{E}_1^L = E_1^L \subset \text{Sing}(0,\mathcal{N})$ by Lemma C.1, and $\mathring{E}_1^L$ is infinite, by the preceding corollary.

(24) **Corollary.** *All the functions* $x_j^\ell(t,\mathcal{N})$, $1 \le \ell \le L$, $1 \le j \le D_\ell$, *are non-constant.*

PROOF. Fix $\bar\ell$, $\bar\jmath$. Then $x_{\bar\jmath}^{\bar\ell}(t,\mathcal{N})$ is the output of a simpler neural net $\mathcal{N}_\# = [(D_0^\#,\dots,D_{L_\#}^\#), \omega_{jk}^{\#\ell}, (\theta_j^{\#\ell})]$, defined by

$$\begin{gathered} L_\# = \bar\ell,\ \ D_\ell^\# = D_\ell\,, \qquad \text{for } \ \ell < L_\#\ ,\ D_{L_\#} = 1\,, \\ \omega_{jk}^{\#\ell} = \omega_{jk}^\ell \quad \text{and} \quad \theta_j^{\#\ell} = \theta_j^\ell\ , \qquad \text{for } \ \ell < L_\#\ , \\ \omega_{1k}^{\#\bar\ell} = \omega_{\bar\jmath k}^{\bar\ell} \quad \text{and} \quad \theta_1^{\#\bar\ell} = \theta_{\bar\jmath}^{\bar\ell}\,. \end{gathered}$$

The net $\mathcal{N}_\#$ again satisfies (5), so Corollary (23) applies to $\mathcal{N}_\#$. Thus $x_1^{L\#}(t,\mathcal{N}_\#)$ is non-constant, and we observed that $x_{\bar\jmath}^{\bar\ell}(t,\mathcal{N}) = x_1^{L\#}(t,\mathcal{N}_\#)$.

Next, we relate $\operatorname{Sing}(\ell,\mathcal{N})$ for $\ell \ge 1$ to the sets $E_j^{\ell'}$.

(25) **Lemma.** $\mathring{E}_j^\ell \subset \operatorname{Sing}(L-\ell,\mathcal{N})$ *for* $1 \le \ell \le L$, $1 \le j \le D_\ell$.

PROOF. We use downward induction on $\ell$. When $\ell = L$, (25) is contained in (1). For the induction step, fix $\ell$ $(1 \le \ell \le L-1)$, and assume

$$\mathring{E}_j^{\ell+1} \subset \operatorname{Sing}(L-\ell-1,\mathcal{N})\,. \tag{26}$$

We shall prove that

$$\mathring{E}_{k_0}^\ell \subset \operatorname{Sing}(L-\ell,\mathcal{N})\,, \qquad \text{for } \ 1 \le k_0 \le D_\ell\,. \tag{27}$$

In fact, (26) and (15) show that every point of $\mathring{E}_{k_0}^\ell$ is an accumulation point of $\operatorname{Sing}(L-\ell-1,\mathcal{N})$ and thus belongs to $\operatorname{Sing}(L-\ell,\mathcal{N})$ by Definition A.10. Hence, (26) implies (27), completing the induction.

(28) **Lemma.** $\operatorname{Sing}(L-\bar\ell,\mathcal{N}) \subset \bigcup_{1 \le \ell \le \bar\ell} \ \bigcup_{1 \le j \le D_\ell} E_j^\ell$, *for* $1 \le \bar\ell \le L$.

PROOF. Again we use downward induction on $\bar\ell$. When $\bar\ell = L$, (28) is contained in (1). For the induction step, fix $\bar\ell$, $2 \le \bar\ell \le L$, and assume

$$\operatorname{Sing}(L-\bar\ell,\mathcal{N}) \subset \bigcup_{1 \le \ell \le \bar\ell} \ \bigcup_{j=1}^{D_\ell} E_j^\ell\,. \tag{29}$$

We shall prove that

$$\text{(30)} \qquad \operatorname{Sing}(L-\bar{\ell}+1,\mathcal{N}) \subset \bigcup_{1\le \ell<\bar{\ell}} \bigcup_{j=1}^{D_\ell} E_j^\ell .$$

In fact, (B.9) shows that the right-hand side of (29) is a closed set. Hence, any accumulation point of $\operatorname{Sing}(L-\bar{\ell},\mathcal{N})$ is again contained in the right-hand side of (29). By definition (A.10), this implies that

$$\operatorname{Sing}(L-\bar{\ell}+1,\mathcal{N}) \subset \bigcup_{1\le \ell\le\bar{\ell}} \bigcup_{j=1}^{D_\ell} E_j^\ell .$$

Therefore, (30) will follow if we can prove

(31) No point of $E_{j_0}^{\bar{\ell}}$ is an accumulation point of $\bigcup_{1\le \ell\le\bar{\ell}} \bigcup_{j=1}^{D_\ell} E_j^\ell$, $1 \le j_0 \le D_{\bar{\ell}}$.

Assertion (31) is equivalent to

(32) $E_{j_0}^{\bar{\ell}}$ contains no accumulation points of $E_j^\ell$ $1\le\ell\le\bar{\ell}$, $1\le j_0\le D_{\bar{\ell}}$, $1\le j\le D_\ell$.

Thus, (30) follows from (32). To prove (32), we distinguish two cases.

*CASE* 1: $\ell<\bar{\ell}$. From (B.9) and (B.12) we see that

$$\mathbb{C}\setminus\Omega_*^{\bar{\ell}-1} = \bigcup_{1\le \ell<\bar{\ell}} \bigcup_{j=1}^{E_\ell} E_j^\ell$$

and that $E_{j_0}^{\bar{\ell}} \subset \Omega_*^{\bar{\ell}-1}$. Since $\Omega_*^{\bar{\ell}-1}$ is open, these remarks imply (32) for $\ell<\bar{\ell}$.

*CASE* 2: $\ell=\bar{\ell}$. Then (B.12) shows that $E_{j_0}^{\bar{\ell}} \subset \Omega_*^{\bar{\ell}-1}$, and that $E_j^{\bar{\ell}}$ has no accumulation points in $\Omega_*^{\bar{\ell}-1}$. These remarks imply (32) for $\ell=\bar{\ell}$.

Thus, (32) holds in either case, which completes the proof of (30). We have shown that (29) implies (30), completing the downward induction.

(33) **Corollary.** $\operatorname{Sing}(\ell,\mathcal{N})$ *is empty for* $\ell\ge L$.

PROOF. Lemma (28) yields $\mathrm{Sing}(L-1,\mathcal{N}) \subset \bigcup_{j=1}^{D_\ell} E_j^1$. From (6) we see that $\bigcup_{j=1}^{D_1} E_j^1$ has no accumulation points. Hence, $\mathrm{Sing}(L,\mathcal{N})$ is empty, from which (33) is obvious.

(34) **Lemma.** *Fix* $t_0 \in \overset{\circ}{E}{}_{k_0}^{\ell}$, $1 \le \ell \le L-1$, $1 \le k_0 \le D_\ell$, *and set*

$$U_\delta = \{t \in \mathbb{C}\colon\ |t-t_0| < \delta\}, \qquad \text{for } \delta > 0. \tag{35}$$

*If $\delta$ is small enough, then* $\mathrm{Sing}(L-\ell-1,\mathcal{N}) \cap (U_\delta \setminus \{t_0\})$ *is the approximate union of the sets*

$$E_j^{\ell+1} \cap (U_\delta \setminus \{t_0\}), \qquad j = 1,\ldots,D_{\ell+1}.$$

PROOF. We must prove two assertions:

$$\mathrm{Sing}(L-\ell-1,\mathcal{N}) \cap (U_\delta \setminus \{t_0\}) \subset \bigcup_{j=1}^{D_{\ell+1}} E_j^{\ell+1} \cap (U_\delta \setminus \{t_0\}) \tag{36}$$

and

(37) Any point belonging to exactly one of the sets $E_j^{\ell+1} \cap (U_\delta \setminus \{t_0\})$, $1 \le j \le D_{\ell+1}$, belongs also to $\mathrm{Sing}(L-\ell-1,\mathcal{N}) \cap (U_\delta \setminus \{t_0\})$.

However, (37) is immediate from (25), so it remains only to prove (36). From (28) we have
(38)

$$\mathrm{Sing}(L-\ell-1,\mathcal{N}) \cap (U_\delta \setminus \{t_0\}) \subset \bigcup_{1 \le \ell' \le \ell+1} \bigcup_{j=1}^{D_{\ell'}} \left(E_j^{\ell'} \cap (U_\delta \setminus \{t_0\})\right).$$

On the other hand, since $t_0 \in E_{k_0}^{\ell}$, (32) shows that $E_j^{\ell'} \cap (U_\delta \setminus \{t_0\})$ is empty if $\ell' \le \ell$ and $\delta$ is small enough. Therefore, (38) implies (36).

(39) **Lemma.** $\mathrm{Sing}(L-1,\mathcal{N})$ *is the approximate union for the sets* $E_j^1$ *for* $j = 1,\ldots,D_1$.

PROOF. Immediate from Lemmas 25 and 28.

**D. Summary.**

Let $\mathcal{N} = [(D_0, \ldots, D_L), (\omega^\ell_{jk}), (\theta^\ell_j)]$ be a neural net. We make the following

(1) **Assumption.** *$\omega^\ell_{jk} \neq 0$, and for $j \neq j'$, the ratio $\omega^\ell_{jk}/\omega^\ell_{j'k}$ is not equal to any fraction of the form $p/q$ with $p$, $q$ integers and $1 \le q \le 100\, D^2_\ell$.*

(2) **Lemma.** *For each $\ell$ $(1 \le \ell \le L)$, the sets $\Omega^\ell_j$, $\Omega^\ell_*$, $E^\ell_j$, $\mathring{E}^\ell_j$ and the functions $x^\ell_j(t, \mathcal{N})$ are determined entirely by the $D_{\ell'}$, $\omega^{\ell'}_{jk}$ and $\theta^{\ell'}_j$ for $\ell' \le \ell$ (see* (B.19)*).*

(3) **Lemma.** *For each $\ell \ge 0$, the set* $\mathrm{Sing}(\ell, \mathcal{N})$ *is determined entirely by the output $t \mapsto x^L_1(t, \mathcal{N})$ $(t \in \mathbb{R})$ of the neural net (see* (B.18)*).*

(4) **Lemma.** *For $1 \le \ell \le L$, $1 \le j \le D_\ell$, the function $x^\ell_j(t, \mathcal{N})$ is analytic on $\Omega^\ell_j$, with poles at the points of $E^\ell_j$ (see* (B.13) *and* (B.14)*).*

(5) **Lemma.** $\mathrm{Sing}(\ell, \mathcal{N})$ *is empty for $\ell \ge L$ (see* (C.33)*).*

(6) **Lemma.** $\mathrm{Sing}(L-1, \mathcal{N})$ *is the approximate union of the arithmetic progressions $\Pi\left(2\pi i/\omega^1_{j1}, (\pi i - \theta^1_j)/\omega^1_{j1}\right)$ for $j = 1, \ldots, D_1$ (see* (C.6) *and* (C.39)*).*

(7) **Lemma.** *For $1 \le \ell \le L$, $1 \le j \le D_\ell$, the set $\mathring{E}^\ell_j$ is infinite (see* (C.22)*).*

(8) **Lemma.** *Fix $t_0 \in \mathring{E}^\ell_{k_0}$, $1 \le \ell \le L-1$, $1 \le k_0 \le D_\ell$. For $\delta > 0$, set*

$$U_\delta = \{t \in \mathbb{C} : |t - t_0| < \delta\}, \quad V_\delta = U_\delta \setminus \{t_0\}. \tag{9}$$

*Then the following properties hold if $\delta$ is small enough.*

(10) $V_{2\delta} \subset \Omega^\ell_{k_0}$ *(see* (C.11)*).*

(11) $U_{2\delta} \subset \Omega^\ell_k$ *for $k \neq k_0$ $(1 \le k \le D_\ell)$ (see* (C.12)*).*

(12) *Sing* $(L-\ell-1, \mathcal{N}) \cap V_\delta$ *is the approximate union of the sets* $E_j^{\ell+1} \cap V_\delta$ *for* $j = 1, \dots, D_{\ell+1}$, (*see* (C.34)).

(13) $$E_j^{\ell+1} \cap V_\delta = \left\{ t \in V_\delta \colon \sum_{k=1}^{D_\ell} \omega_{jk}^{\ell+1} x_k^\ell(t, \mathcal{N}) + \theta_j^{\ell+1} \in \Pi(2\pi i, \pi i) \right\},$$

$1 \le j \le D_{\ell+1}$, (*see* (C.13)).

(14) *The set* $F_j = \{x_{k_0}^\ell(t, \mathcal{N}) \colon t \in E_j^{\ell+1} \cap V_\delta\}$ *approximates an arithmetic progression of the form* $\Pi\left(2\pi i/\omega_{jk_0}^{\ell+1}, \beta_{jk_0}^{\ell+1}\right)$ *for some complex number* $\beta_{jk_0}^{\ell+1}$, (*see* (C.14)).

(15) **Lemma.** *For* $1 \le \ell \le L$, $1 \le j \le D_\ell$, *the function* $t \mapsto x_j^\ell(t, \mathcal{N})$ $(t \in \mathbb{R})$ *is non-constant* (*see* (C.24)).

## IV. Proof of the Uniqueness Theorem.

### A. Setting up the Induction.

In this section, we start the proof of Theorem I.B.9. We begin with some preliminary remarks. Let $\mathcal{N} = [(D_0, \dots, D_L), (\omega_{jk}^\ell), (\theta_j^\ell)]$ be a neural net satisfying condition (I.B.10). By (III.D.6) and Corollary II.D.33, the set Sing $(L-1, \mathcal{N})$ is non-empty. On the other hand, (III.D.5) shows that Sing $(\ell, \mathcal{N})$ is empty for $\ell \ge L$. Hence, the depth $L$ of the neural net can be inferred from a knowledge of the sets Sing $(\ell, \mathcal{N})$, $\ell \ge 0$. Lemma III.D.3 therefore shows that $L$ can be inferred from knowledge of the output function $t \mapsto x_1^L(t, \mathcal{N})$. Thus, if two neural nets produce the same output, then they have the same depth.

Now let $\mathcal{N} = [(D_0, \dots, D_L), (\omega_{jk}^\ell), (\theta_j^\ell)]$, $\widetilde{\mathcal{N}} = [(\widetilde{D}_0, \dots, \widetilde{D}_{\widetilde{L}}), (\widetilde{\omega}_{jk}^\ell), (\widetilde{\theta}_j^\ell)]$ be neural nets satisfying the hypotheses of Theorem I.B.9. We must show that $\mathcal{N}$ and $\widetilde{\mathcal{N}}$ are isomorphic. If Theorem I.B.9 were false, then we could find a counterexample with max{Size $(\mathcal{N})$, Size $(\widetilde{\mathcal{N}})$} as small as possible. (Recall that the size of $\mathcal{N}$ is defined as the sum $D_0 + \cdots + D_L$.) Thus, we may assume that

(1) max{Size $(\mathcal{N})$, Size $(\widetilde{\mathcal{N}})$} $= S$, and that

(2) Theorem I.B.9 holds for any two neural nets $\mathcal{N}'$, $\widetilde{\mathcal{N}}$ of size strictly less than $S$.

Also, from the preceding paragraph, we know that

$$L = \widetilde{L}\,. \tag{3}$$

By induction on $\bar{\ell}$, $1 \le \bar{\ell} \le L$, we will prove that by subjecting $\mathcal{N}$, $\widetilde{\mathcal{N}}$ to isomorphisms we can achieve

$$D_\ell = \widetilde{D}_\ell\,, \qquad \text{for } \ell \le \bar{\ell}\,, \tag{4}$$

$$\omega^\ell_{jk} = \widetilde{\omega}^\ell_{jk}\,, \qquad \text{for } 1 \le \ell \le \bar{\ell},\ 1 \le j \le D_\ell\,,\ 1 \le k \le D_{\ell-1} \tag{5}$$

$$\theta^\ell_j = \widetilde{\theta}^\ell_j\,, \qquad \text{for } 1 \le \ell \le \bar{\ell},\ 1 \le j \le D_\ell\,. \tag{6}$$

If we can prove this for $\bar{\ell} = L$, then Theorem I.B.9 is established.

We will prove (4), (5), (6) by induction on $\bar{\ell}$. In this section, we treat the case $\bar{\ell} = 1$, while the next section gives the induction step. Thus, suppose $\bar{\ell} = 1$. Lemmas III.D.3 and III.D.6 show that the set $\mathrm{Sing}\,(L-1, \mathcal{N})$ is the approximate union of the arithmetic progressions $\Pi\left(2\pi i/\omega^1_{j1}, (\pi i - \theta^1_j)/\omega^1_{j1}\right)$, $1 \le j \le D_1$, and also the approximate union of the progressions $\Pi\left(2\pi i/\widetilde{\omega}^1_{j1}, (\pi i - \widetilde{\theta}^1_j)/\widetilde{\omega}^1_{j1}\right)$, $1 \le j \le \widetilde{D}_1$. The Deconstruction Lemma therefore tells us that $D_1 = \widetilde{D}_1$, and that

$$\Pi\left(\frac{2\pi i}{\omega^1_{j1}}, \frac{\pi i - \theta^1_j}{\omega^1_{j1}}\right) = \Pi\left(\frac{2\pi i}{\widetilde{\omega}^1_{(\gamma j)1}}, \frac{\pi i - \widetilde{\theta}^1_{(\gamma j)}}{\widetilde{\omega}^1_{(\gamma j)1}}\right)\,, \qquad 1 \le j \le D_1\,,$$

for a permutation $\gamma\colon \{1, \ldots, D_1\} \to \{1, \ldots, D_1\}$. Remark I.A.7 yields

$$\omega^1_{j1} = \varepsilon_j\, \widetilde{\omega}^1_{(\gamma j)1}\,, \quad \theta^1_j = \varepsilon_j\, \widetilde{\theta}^1_{(\gamma j)} + 2\pi i\, m_j\,, \qquad 1 \le j \le D_1\,, \tag{7}$$

for $\varepsilon_j = \pm 1$ and integers $m_j$. Since $\theta^1_j$ and $\widetilde{\theta}^1_j$ are real, we must have $m_j = 0$. Also, by subjecting $\widetilde{\mathcal{N}}$ to an isomorphism that permutes the nodes of layer 1, we can achieve $\gamma = \text{identity}$ in (7). Thus,

$$\omega^1_{j1} = \varepsilon_j\, \widetilde{\omega}^1_{j1}\,, \quad \theta^1_j = \varepsilon_j\, \widetilde{\theta}^1_j\,, \quad \varepsilon_j = \pm 1\,, \qquad 1 \le j \le D_1\,. \tag{8}$$

If $L > 1$, then we can subject $\widetilde{\mathcal{N}}$ to an isomorhism that changes the signs of the nodes at layer 1, to achieve $\varepsilon_j = 1$, $1 \le j \le D_1$, in (8).

Thus, we have achieved (4), (5), (6) with $\bar{\ell} = 1$, unless $L = 1$. If $L = 1$, then there is no isomorphism that changes the signs at layer 1, since layer 1 is the output layer. In this case we argue as follows. For $L = 1$, the outputs of the nets $\mathcal{N}$, $\widetilde{\mathcal{N}}$ are

$$(9) \qquad x_1^L(t,\mathcal{N}) = \sigma(\omega_{11}^1 t + \theta_1^1) \qquad x_1^L(t,\widetilde{\mathcal{N}}) = \sigma(\widetilde{\omega}_{11}^1 t + \widetilde{\theta}_1^1)\,.$$

Equation (8) says that

$$(10) \qquad \omega_{11}^1 = \varepsilon\, \widetilde{\omega}_{11}^1\,, \quad \theta_1^1 = \varepsilon\, \widetilde{\theta}_1^1\,, \quad \varepsilon = \pm 1\,.$$

From (9), (10) we get $x_1^L(t,\mathcal{N}) = \varepsilon x_1^L(t,\widetilde{\mathcal{N}})$ for $t \in \mathbb{R}$. On the other hand, hypothesis of Theorem I.B.9 gives $x_1^L(t,\mathcal{N}) = x_1^L(t,\widetilde{\mathcal{N}})$, and $x_1^L(t,\mathcal{N})$ is not identically zero. Hence, $\varepsilon = +1$, so that we have achieved (4), (5), (6).

## B. The Inductive Step.

Suppose $\bar{\ell}$ is given, $1 \le \bar{\ell} \le L-1$, and the nets $\mathcal{N}$, $\widetilde{\mathcal{N}}$ in the previous section satisfy

$$(1) \qquad D_\ell = \widetilde{D}_\ell\,, \qquad \text{for } 0 \le \ell \le \bar{\ell},$$

$$(2) \qquad \omega_{jk}^\ell = \widetilde{\omega}_{jk}^\ell\,, \qquad \text{for } 1 \le \ell \le \bar{\ell},\ 1 \le j \le D_\ell,\ 1 \le k \le D_{\ell-1}\,,$$

$$(3) \qquad \theta_j^\ell = \widetilde{\theta}_j^\ell\,, \qquad \text{for } 1 \le \ell \le \bar{\ell},\ 1 \le j \le D_\ell\,.$$

Then we will prove that

$$(4) \qquad D_{\bar{\ell}+1} = \widetilde{D}_{\bar{\ell}+1}\,,$$

and that we can subject $\mathcal{N}$, $\widetilde{\mathcal{N}}$ to isomorphisms to achieve

$$(5) \qquad \omega_{jk}^\ell = \widetilde{\omega}_{jk}^\ell\,, \qquad \text{for } 1 \le \ell \le \bar{\ell}+1,\ 1 \le j \le D_\ell,\ 1 \le k \le D_{\ell-1}\,,$$

$$(6) \qquad \theta_j^\ell = \widetilde{\theta}_j^\ell\,, \qquad \text{for } 1 \le \ell \le \bar{\ell},\ 1 \le j \le D_\ell\,.$$

This inductive step will complete the proof of Theorem I.B.9.

Let $E_j^\ell$, $\mathring{E}_j{}^\ell$, $\Omega_j^\ell$, $\Omega_*^\ell$ be the sets constructed from $\mathcal{N}$ in Section III, and let $\widetilde{E}_j^\ell$, $\mathring{\widetilde{E}}_j^\ell$, $\widetilde{\Omega}_j^\ell$, $\widetilde{\Omega}_*^\ell$ be the analogous sets arising from $\widetilde{\mathcal{N}}$. By (1), (2), (3) and Lemma III.D.2, we have

$$(7) \qquad E_j^\ell = \widetilde{E}_j^\ell\,, \quad \mathring{\widetilde{E}}_j^\ell = \mathring{E}_j^\ell\,, \quad \widetilde{\Omega}_j^\ell = \Omega_j^\ell\,, \quad \widetilde{\Omega}_*^\ell = \Omega_*^\ell\,, \quad \text{for } \ell \le \bar{\ell},$$

and

(8) $$x_j^\ell(t,\mathcal{N}) = x_j^\ell(t,\widetilde{\mathcal{N}}), \qquad \text{for } \ell \leq \bar{\ell},\ t \in \Omega_j^\ell .$$

Set

(9) $$k_0 = D_{\bar{\ell}} .$$

Lemma III.D.7 shows that $\overset{\circ}{E}{}_{k_0}^{\bar{\ell}} = \overset{\circ}{\widetilde{E}}{}_{k_0}^{\bar{\ell}}$ is infinite. Fix any $t_0 \in \overset{\circ}{E}{}_{k_0}^{\bar{\ell}}$. For $\delta > 0$, set

(10) $$U_\delta = \{t \in \mathbb{C}\colon\ |t - t_0| < \delta\}, \qquad V_\delta = U_\delta \setminus \{t_0\},$$

and take $\delta$ so small that (III.D.10)-(III.D.14) hold with $\bar{\ell}$ in place of $\ell$, both for $\mathcal{N}$ and $\widetilde{\mathcal{N}}$. Lemma III.D.3 gives

(11) $$\operatorname{Sing}(L - \bar{\ell} - 1, \mathcal{N}) = \operatorname{Sing}(L - \bar{\ell} - 1, \widetilde{\mathcal{N}}).$$

We will check that

$$F = \{x_{k_0}^{\bar{\ell}}(t,\mathcal{N})\colon\ t \in V_\delta \cap \operatorname{Sing}(L - \bar{\ell} - 1, \mathcal{N})\}$$

is the approximate union of the sets $F_j$ defined in (III.D.14). This amounts to showing that

(12) $F \subset \bigcup_{j=1}^{D_{\bar{\ell}+1}} F_j$, and

(13) Any point $x$ belonging to exactly one of the $F_j$ must belong to $F$.

To see (12), let $x \in F$. Then $x = x_{k_0}^{\bar{\ell}}(t,\mathcal{N})$ with $t \in V_\delta \cap \operatorname{Sing}(L - \bar{\ell} - 1, \mathcal{N})$, so that $t \in V_\delta \cap E_j^{\bar{\ell}+1}$ for some $j$, by (III.D.12). Since $x = x_{k_0}^{\bar{\ell}}(t,\mathcal{N})$ with $t \in V_\delta \cap E_j^{\bar{\ell}+1}$, we have $x \in F_j$, proving (12). To check (13), suppose $x$ belongs to $F_{j_0}$ but not to any other $F_j$. Then since $x \in F_{j_0}$, we have $x = x_{k_0}^{\bar{\ell}}(t,\mathcal{N})$ with $t \in E_{j_0}^{\bar{\ell}+1} \cap V_\delta$. If $t \in E_j^{\bar{\ell}+1}$ for some $j \neq j_0$, then it would follow that $x \in F_j$, contradicting our assumption. Hence, $t$ belongs to $E_{j_0}^{\bar{\ell}+1}$ but not to $E_j^{\bar{\ell}+1}$ for $j \neq j_0$. Since also $t \in V_\delta$, (III.D.12) implies $t \in V_\delta \cap \operatorname{Sing}(L - \bar{\ell} - 1, \mathcal{N})$, so that $x = x_{k_0}^{\bar{\ell}}(t,\mathcal{N})$ belongs to $F$. This completes the proof of (13), and shows that $F$ is the approximate union of the $F_j$, $1 \leq j \leq D_{\bar{\ell}+1}$. In view of (11), an analogous argument shows that $F$ is also the approximate union of the

$\widetilde{F}_j$ , $1 \le j \le \widetilde{D}_{\bar{\ell}+1}$ , where $\{\widetilde{F}_j\}$ are the analogues of the $\{F_j\}$ arising from $\widetilde{\mathcal{N}}$. Moreover, $F_j$ approximates $\Pi\left(2\pi i/\omega_{jk_0}^{\bar{\ell}+1}, \beta_j\right)$ for suitable $\beta_j$, while $\widetilde{F}_j$ approximates $\Pi\left(2\pi i/\widetilde{\omega}_{jk_0}^{\bar{\ell}+1}, \tilde{\beta}_j\right)$ for suitable $\tilde{\beta}_j$, by (III.D.14).

Since also the $(\omega_{jk_0}^{\bar{\ell}+1})$ and $(\widetilde{\omega}_{jk_0}^{\bar{\ell}+1})$ satisfy (I.B.10) and (I.B.11), the Deconstruction Lemma applies. Hence, (4) holds, and for some permutation $\gamma: \{1, \ldots, D_{\bar{\ell}+1}\} \to \{1, \ldots, D_{\bar{\ell}+1}\}$ we have

$$\Pi\left(\frac{2\pi i}{\omega_{jk_0}^{\bar{\ell}+1}}, \beta_j\right) = \Pi\left(\frac{2\pi i}{\widetilde{\omega}_{(\gamma j)k_0}^{\bar{\ell}+1}}, \tilde{\beta}_{(\gamma j)}\right), \qquad 1 \le j \le D_{\bar{\ell}+1}.$$

In particular,

$$(14) \qquad \omega_{jk_0}^{\bar{\ell}+1} = \varepsilon_j \, \widetilde{\omega}_{(\gamma j)k_0}^{\bar{\ell}+1}, \quad 1 \le j \le D_{\bar{\ell}+1}, \qquad \text{with } \varepsilon_j = \pm 1.$$

By subjecting $\widetilde{\mathcal{N}}$ to an isomorphism that permutes the nodes of layer $(\bar{\ell}+1)$, we can preserve (1)-(4) and (7), (8), (11), and bring about $\gamma =$ identity in (4). Thus we may assume

$$(15) \qquad \omega_{jk_0}^{\bar{\ell}+1} = \varepsilon_j \, \widetilde{\omega}_{jk_0}^{\bar{\ell}+1}, \qquad 1 \le j \le D_{\bar{\ell}+1},$$

with $\varepsilon_j = \pm 1$ and $\gamma =$ identity. Recall that $k_0 = D_{\bar{\ell}}$ (see (9)).

The next step is to establish the following result.

(16) **Lemma.** *$\omega_{jk}^{\bar{\ell}+1} = \varepsilon_j \, \widetilde{\omega}_{jk}^{\bar{\ell}+1}$, $1 \le j \le D_{\bar{\ell}+1}$, $1 \le k \le D_{\bar{\ell}}$, and $\theta_j^{\bar{\ell}+1} = \varepsilon_j \, \widetilde{\theta}_j^{\bar{\ell}+1}$, $1 \le j \le D_{\bar{\ell}+1}$.*

Note that the proof of (15) applies to other $k$, not just $k_0$, and shows that $\omega_{jk}^{\bar{\ell}+1} = \varepsilon_{jk} \widetilde{\omega}_{(\gamma' j)k}^{\bar{\ell}+1}$ with $\varepsilon_{jk} = \pm 1$, and $\gamma'$ depending on $k$. However, (16) gives sharper restrictions on the $\omega$'s and $\widetilde{\omega}$'s.

PROOF OF LEMMA 16. We return to the Deconstruction Lemma applied to $F$, $F_j$, $\widetilde{F}_j$, $\Pi\left(2\pi i/\omega_{jk_0}^{\bar{\ell}+1}, \beta_j\right)$, $\Pi\left(2\pi i/\widetilde{\omega}_{jk_0}^{\bar{\ell}+1}, \tilde{\beta}_j\right)$. Since $\gamma =$ identity, the Deconstruction Lemma yields for each fixed $j_0$, $1 \le j_0 \le D_{\bar{\ell}+1}$, a sequence $(x_\nu)_{\nu \ge 1}$ with the properties

$$(17) \qquad |x_\nu| \to \infty, \qquad \text{as } \nu \to \infty,$$

$$(18) \qquad x_\nu \in F_{j_0} \setminus \bigcup_{j \ne j_0} F_j,$$

$$(19) \qquad x_\nu \in \widetilde{F}_{j_0} \setminus \bigcup_{j \ne j_0} \widetilde{F}_j.$$

Since $x_\nu \in F_{j_0}$, we have

$$x_\nu = x^{\bar\ell}_{k_0}(t_\nu, \mathcal{N}), \tag{20}$$

with

$$t_\nu \in V_\delta \cap E^{\bar\ell+1}_{j_0}. \tag{21}$$

Observe that

$$t_\nu \notin V_\delta \cap E^{\bar\ell+1}_{j}, \qquad \text{for } j \neq j_0. \tag{22}$$

In fact, if (22) were false, then (20) would show that $x_\nu \in F_j$ with $j \neq j_0$, contradicting (18). Similarly, we know from (19) that

$$t_\nu \notin V_\delta \cap \widetilde{E}^{\bar\ell+1}_{j}, \qquad \text{for } j \neq j_0. \tag{23}$$

From (21), (22) and (III.D.12), we see that $t_\nu \in V_\delta \cap \mathrm{Sing}(L-\bar\ell-1, \mathcal{N})$. Hence also $t_\nu \in V_\delta \cap \mathrm{Sing}(L-\bar\ell-1, \widetilde{\mathcal{N}})$ by (11), so that another application of (III.D.12) yields $t_\nu \in V_\delta \cap \widetilde{E}^{\bar\ell+1}_{j_1}$ for some $j_1$. In view of (23), we must have $j_1 = j_0$. Hence,

$$t_\nu \in V_\delta \cap \widetilde{E}^{\bar\ell+1}_{j_0}. \tag{24}$$

From (17), (20), $t_\nu \in V_\delta$, and (III.D.4), (III,D.10), we see that

$$t_\nu \to t_0, \qquad \text{as } \nu \to \infty, \quad t_\nu \neq t_0. \tag{25}$$

From (21), (24) and (III.,D.13), we get

$$t_\nu \in V_\delta, \qquad \sum_{k=1}^{D_{\bar\ell}} \omega^{\bar\ell+1}_{j_0 k} x^{\bar\ell}_k(t_\nu, \mathcal{N}) + \theta^{\bar\ell}_{j_0} \in \Pi(2\pi i, \pi i) \tag{26}$$

and

$$t_\nu \in V_\delta, \qquad \sum_{k=1}^{D_{\bar\ell}} \widetilde{\omega}^{\bar\ell+1}_{j_0 k} x^{\bar\ell}_k(t_\nu, \widetilde{\mathcal{N}}) + \theta^{\bar\ell+1}_{j_0} \in \Pi(2\pi i, \pi i). \tag{27}$$

In view of (15), (26), (27), we obtain

$$\left(\sum_{k=1}^{D_{\bar\ell}-1} \omega_{j_0k}^{\bar\ell+1}\, x_k^{\bar\ell}(t_\nu,\mathcal{N}) + \theta_{j_0}^{\bar\ell+1}\right) - \left(\sum_{k=1}^{D_{\bar\ell}-1} (\varepsilon_{j_0}\,\widetilde\omega_{j_0k}^{\bar\ell+1})\, x_k^{\bar\ell}(t_\nu,\widetilde{\mathcal{N}}) + (\widetilde\theta_{j_0}^{\bar\ell+1}\varepsilon_{j_0})\right) \in \Pi(2\pi i,0).$$

That is,

$$F(t_\nu) \in \Pi(2\pi i, 0)\,, \tag{28}$$

with

$$F(t) = \left(\sum_{k=1}^{D_{\bar\ell}-1} \omega_{j_0}^{\bar\ell+1}\, x_k^{\bar\ell}(t,\mathcal{N}) + \theta_{j_0}^{\bar\ell+1}\right) - \left(\sum_{k=1}^{D_{\bar\ell}-1} (\varepsilon_{j_0}\,\widetilde\omega_{j_0k}^{\bar\ell+1})\, x_k^{\bar\ell}(t,\widetilde{\mathcal{N}}) + (\varepsilon_{j_0}^{\bar\ell+1}\widetilde\theta_{j_0}^{\bar\ell+1})\right). \tag{29}$$

Since $t_0 \in \overset{\circ}{E}{}_{k_0}^{\bar\ell}$ with $k_0 = D_{\bar\ell}$, (III.D.4) and (III.D.11) show that $F(t)$ is analytic on $U_\delta$. Also, (25) yields $F(t_\nu) \to F(t_0)$ as $\nu \to \infty$. This shows that $F(t_\nu)$ is eventually constant, by (28). Another application of (25) shows that $F(t)$ is constant on $U_\delta$. In view of (28), there is an integer $m$ such that $F(t) = 2\pi i m$ for all $t \in U_\delta$. However, $F(t)$ is analytic on $\Omega_*^{\bar\ell}$ (by (7), (29) and (IIII.D.4)). Since $\Omega_*^{\bar\ell}$ contains $V_\delta$ (by (III.D.10), (III,D.11)), and since $\Omega_*^{\bar\ell}$ is connected, it follows by analytic continuation that $F(t) = 2\pi i m$ for all $t \in \Omega_*^{\bar\ell}$. In particular, $F(t) = 2\pi i m$ for $t \in \mathbb{R}$. A glance at (29) shows that $F(t)$ is real for $t \in \mathbb{R}$. Hence, $m = 0$, so that $F(t) = 0$ for $t$ real, *i.e.*

$$\sum_{k=1}^{D_{\bar\ell}-1} \omega_{j_0k}^{\bar\ell+1}\, x_k^{\bar\ell}(t,\mathcal{N}) + \theta_{j_0}^{\bar\ell+1} = \sum_{k=1}^{D_{\bar\ell}-1} (\widetilde\omega_{j_0k}^{\bar\ell+1}\,\varepsilon_{j_0})\, x_k^{\bar\ell}(t,\widetilde{\mathcal{N}}) + (\widetilde\theta_{j_0}^{\bar\ell+1}\varepsilon_{j_0})\,, \tag{30}$$

for all $t \in \mathbb{R}$.

To complete the proof of (16), we distinguish two cases.

*CASE* 1: $D_{\bar\ell} = 1$. Then already (15) shows that $\omega_{j_0k}^{\bar\ell+1} = \varepsilon_{j_0}\,\widetilde\omega_{j_0k}^{\bar\ell+1}$ for $1 \le k \le D_{\bar\ell}$, and (30) shows that $\theta_{j_0}^{\bar\ell+1} = \varepsilon_{j_0}\widetilde\theta_{j_0}^{\bar\ell+1}$. Since $j_0$ is arbitrary, $1 \le j_0 \le D_{\bar\ell+1}$, the proof of (16) is complete in *Case* 1.

*CASE* 2: $D_{\bar{\ell}} > 1$. Then the left and right-hand sides of (30), composed with $\sigma(\,\cdot\,)$, are the outputs of auxiliary neural nets $\widehat{\mathcal{N}}$ and $\check{\mathcal{N}}$ respectively. Specifically, we set

(31) $\widehat{\mathcal{N}} = [(D_0, D_1, \ldots, D_{\bar{\ell}-1}, D_{\bar{\ell}} - 1, 1), (\widehat{\omega}^{\ell}_{jk}), (\widehat{\theta}^{\ell}_{j})]$, with

(32) $\widehat{\omega}^{\ell}_{jk} = \omega^{\ell}_{jk}$, $\widehat{\theta}^{\ell}_{j} = \theta^{\ell}_{j}$ for $\ell \leq \bar{\ell}$, and

(33) $\widehat{\omega}^{\bar{\ell}+1}_{1k} = \omega^{\bar{\ell}+1}_{j_0 k}$, $\widehat{\theta}^{\bar{\ell}+1}_{1} = \theta^{\bar{\ell}+1}_{j_0}$; and we set

(34) $\check{\mathcal{N}} = [(D_0, D_1, \ldots, D_{\bar{\ell}-1}, D_{\bar{\ell}} - 1, 1), (\check{\omega}^{\ell}_{jk}), (\check{\theta}^{\ell}_{j})]$, with

(35) $\check{\omega}^{\ell}_{jk} = \widetilde{\omega}^{\ell}_{jk}$, $\check{\theta}^{\ell}_{j} = \widetilde{\theta}^{\ell}_{j}$ for $\ell \leq \bar{\ell}$, and

(36) $\check{\omega}^{\bar{\ell}+1}_{1k} = \varepsilon_{j_0} \widetilde{\omega}^{\bar{\ell}+1}_{j_0 k}$, $\check{\theta}^{\bar{\ell}+1}_{1} = \varepsilon_{j_0} \widetilde{\theta}^{\bar{\ell}+1}_{j_0}$.

That is, $\widehat{\mathcal{N}}$ is made from $\mathcal{N}$ by deleting the following nodes:

(a) Node $k_0$ at level $\bar{\ell}$;

(b) All nodes except node $j_0$ at level $\bar{\ell} + 1$;

(c) All nodes at levels higher than $\bar{\ell} + 1$.

For the surviving nodes in $\widehat{\mathcal{N}}$, the weights and thresholds are the same as those of $\mathcal{N}$. Thus, $x^{\bar{\ell}+1}_{j_0}(t, \mathcal{N})$ is the output of the net $\widehat{\mathcal{N}}$.

Similarly, $\check{\mathcal{N}}$ is made from $\widetilde{\mathcal{N}}$ by deleting the same nodes as in (a), (b), (c) above. For the surviving nodes in $\check{\mathcal{N}}$, the weights and thresholds are the same as those of $\widetilde{\mathcal{N}}$, except that we multiply the weights and thresholds at the output level of $\check{\mathcal{N}}$ by $\varepsilon_{j_0}$. Thus, $\varepsilon_{j_0} x^{\bar{\ell}+1}_{j_0}(t, \widetilde{\mathcal{N}})$ is the output of the net $\check{\mathcal{N}}$. Note that our assumption $D_{\bar{\ell}} > 1$ was needed to define $\widehat{\mathcal{N}}$, $\check{\mathcal{N}}$ as neural nets.

Equation (30) shows that the nets $\widehat{\mathcal{N}}$ and $\check{\mathcal{N}}$ produce the same output. Also, $\widehat{\mathcal{N}}$ and $\check{\mathcal{N}}$ satisfy (I.B.10) and (I.B.11). Moreover, the size of $\widehat{\mathcal{N}}$ is strictly less than that of $\mathcal{N}$, and the size of $\check{\mathcal{N}}$ is strictly less than that of $\widetilde{\mathcal{N}}$. (For, one node at level $\bar{\ell}$, and possible additional nodes, are deleted from $\mathcal{N}$, $\widetilde{\mathcal{N}}$ to make $\widehat{\mathcal{N}}$, $\check{\mathcal{N}}$). Hence, by (A.1) and (A.2), the uniqueness Theorem I.B.9 applies to $\widehat{\mathcal{N}}$, $\check{\mathcal{N}}$. Therefore, $\widehat{\mathcal{N}}$ is isomorphic to $\check{\mathcal{N}}$. Also, by definition of $\widehat{\mathcal{N}}$, $\check{\mathcal{N}}$ and by (1), (2), (3), the nets $\widehat{\mathcal{N}}$, $\check{\mathcal{N}}$

are identical below their output level, *i.e.* $\widehat{\omega}^\ell_{jk} = \check{\omega}^\ell_{jk}$, $\widehat{\theta}^\ell_j = \check{\theta}^\ell_j$ for $\ell \le \bar{\ell}$. Also, for fixed $\ell$, $k$, we have $\widehat{\omega}^\ell_{jk} \neq 0$ and $|\widehat{\omega}^\ell_{jk}| \neq |\widehat{\omega}^\ell_{j'k}|$ for $j \neq j'$, by (32) and hypothesis (I.B.10). Therefore, Lemma I.C.1 applies, and shows that the nets $\widehat{\mathcal{N}}$, $\check{\mathcal{N}}$ are identical. In particular,

$$(37) \quad \omega^{\bar{\ell}+1}_{j_0} = \varepsilon_{j_0}\, \widetilde{\omega}^{\bar{\ell}+1}_{j_0 k} \quad \text{for} \quad 1 \le k \le D_{\bar{\ell}} - 1\,, \quad \text{and} \quad \theta^{\bar{\ell}+1}_{j_0} = \varepsilon_{j_0}\, \widetilde{\theta}^{\bar{\ell}+1}_{j_0}\,.$$

Since $j_0$, $1 \le j_0 \le D_{\bar{\ell}+1}$, was arbitrary, Lemma 16 follows from (9), (15) and (37).

To finish the proof of Theorem I.B.9, we distinguish two cases.

*CASE* 1: $L > \bar{\ell} + 1$. Then by subjecting $\widetilde{\mathcal{N}}$ to an isomorphism that changes the signs of the nodes at level $\bar{\ell} + 1$, we can achieve (5) and (6). (That is obvious from (1), (2), (3), (16).) We already proved (4), so we have completed the inductive step in the proof of Theorem I.B.9 in Case 1.

*CASE* 2: $L = \bar{\ell} + 1$. Then (1), (2), (3), (16) show that $D_\ell = \widetilde{D}_\ell$, $0 \le \ell \le L$, and that

$$(38) \qquad \omega^\ell_{jk} = \widetilde{\omega}^\ell_{jk}\,, \quad \theta^\ell_j = \widetilde{\theta}^\ell_j\,, \qquad \text{if } \ell < L\,,$$

$$(39) \qquad \omega^L_{1k} = \varepsilon\, \widetilde{\omega}^L_{1k}\,, \quad \theta^L_1 = \varepsilon \widetilde{\theta}^L_1\,, \qquad \text{with } \varepsilon = \pm 1\,.$$

It follows at once that $x^L_1(t, \mathcal{N}) = \varepsilon\, x^L_1(t, \widetilde{\mathcal{N}})$ for all $t \in \mathbb{R}$. However, from the hypothesis of Theorem I.B.9, we know that $x^L_1(t, \mathcal{N}) = x^L_1(t, \widetilde{\mathcal{N}})$ since $L = \widetilde{L}$. Since also $x^L_1(t, \mathcal{N})$ is a non-constant function (see (III.D.15)), it follows that $\varepsilon = +1$, hence (38), (39) show that the nets $\mathcal{N}$, $\widetilde{\mathcal{N}}$ are identical. In particular, we have achieved (4), (5), (6), completing the inductive step in the proof of theorem (I.B.9) in *Case* 2.

The proof of Theorem I.B.9 is complete.

## References.

[CLH] Hecht-Nielson, R. et al., *Hecht-Nielson transformations.*

[H] Hille, E., *Analytic Function Theory* I. Chelsea, 1976.

[MS] MacIntyre, A. and Sontang, E., to appear.

[N] Moody, J., Hanson, S. and Lippmann, R., eds., *Neural Information Processing Systems* **4**. Morgan Kaufmann, 1992.

[So] Albertini, F. and Sontag, E., Uniqueness of weights for neural networks, to appear.

[Su] Sussman, H., Uniqueness of the weights for minimal feedforward nets with a given input-output map. *Neural Networks* **5** (1992), 589-593.

[W] Weyl, H., *Collected Works.* Springer-Verlag, 1968.

*Recibido:* 18 de junio de 1.993

Charles Fefferman*
Department of Mathematics
Princeton University
Princeton, NJ 08544, U.S.A.

* This research was supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Air Force Office of Scientific Research under Contract F49620-92-C-0072. The United States Goverment is authorized to reproduce and distribute reprints for governmental purposes notwithstanding and copyright notation hereon. This work was also supported by the National Science Foundation.