GLOBAL IDENTIFIABILITY OF CONSTANT-WIDTH DEEP POLYNOMIAL NETWORKS

Abstract. Multihomogeneity accounts for the generic fibers of deep polynomial neural networks when the activation degree is sufficiently large. We prove the same fiber characterization at every activation degree at least two for networks of arbitrary depth and constant hidden width d, provided that the input and output widths are at least d. The degrees may vary by layer. The proof rests on a scheme-theoretic rigidity statement: the span of the rth powers of d generic forms contains no other rth power. Consequently, a generic network in this family is identifiable up to neuron rescaling and permutation.

1. Introduction

A homogeneous polynomial neural network parametrizes a family of vector-valued forms by weight matrices and coordinate-wise monomial activations [5, 6]. Its parametrization is multihomogeneous: permuting the neurons in a hidden layer or rescaling them, with a compensating change in the adjacent weight matrices, does not change the network function [3, Lemma 4]. The identifiability problem asks when these transformations account for the entire fiber.

Finkel, Rodriguez, Wu, and Yahl answered this question when the common activation degree is sufficiently large [3, Theorem 18]. Their proof uses a Newman–Slater-type result on linear relations among powers [3, Proposition 16]. With $M := -1 + 2\max\{d_1, \dots, d_{L-1}\}$, their hypothesis is

$$r > 6M^2 - 6M + 1.$$

This bound is quadratic in the largest hidden width and does not cover small degrees such as $r = 2$. For equi-width architectures, Finkel et al. separately proved that every $r \geq 2$ gives the expected neurovariety dimension [3, Theorem 22]. Expected dimension, however, does not imply that multihomogeneity exhausts the generic fiber.

Independently, Crăciun obtained another high-degree linear-independence bound using a generalized Mason inequality, which is a polynomial *abc* theorem [2, Theorem 3]. Combined with the induction of Finkel et al., this yields the same fiber characterization by a different high-degree route.

Usevich, Borsoi, Dérand, and Clausel subsequently proved finite identifiability for broad families, including nonincreasing hidden widths and layerwise degrees at least two [8, Theorem 11 and Corollary 16]. Here finite identifiability means that a generic fiber contains only finitely many multihomogeneity orbits, whereas global identifiability means that it contains exactly one. Thus their result still allows several inequivalent parameterizations of a generic function, and they stated the corresponding global result as a conjecture [8, Conjecture 40]. Other global-identifiability results require additional numerical or geometric hypotheses [7, Theorem 4.13].

We remove the high-degree hypothesis for networks of constant hidden width and allow the activation degree to vary by layer. We work over an algebraically

closed field $\Bbbk$ of characteristic zero. A property holds *generically* if it holds on a nonempty Zariski-open subset of the relevant parameter space.

Theorem 1.1 (Generic fiber)**.** *Let $n, m \geq d \geq 2$ and $L \geq 2$. Fix the architecture $\boldsymbol{d} = (n, d, \ldots, d, m)$ and activation degrees $r_1, \ldots, r_{L-1} \geq 2$. There is a nonempty Zariski-open subset $\mathcal{U}$ of the parameter space such that, for every $W \in \mathcal{U}$, the fiber through W is one multihomogeneity orbit. Equivalently, a generic network is globally identifiable up to hidden-neuron rescaling and permutation.*

Thus, for constant hidden width, global identifiability holds whenever every layerwise degree is at least two. This strengthens both the equi-width dimension theorem of Finkel et al. and the finite-identifiability theorem of Usevich et al. in this setting.

The proof reduces the network problem to rigidity of spans of powers. For $n, e \geq 1$, define

$$H_{n,e} := \Bbbk[x_1, \ldots, x_n]_e, \qquad X_{n,e,r} := \{[f^r] : 0 \neq f \in H_{n,e}\} \subseteq \mathbb{P}(H_{n,er}).$$

Thus $X_{n,e,r}$ is the projective variety of rth powers of degree-e forms, endowed with its reduced induced scheme structure. The required geometric statement is the following scheme-theoretic form of power-span rigidity. Its set-theoretic content is related to the generalized trisecant lemma [1, Proposition 2.6].

Theorem 1.2 (Power-span rigidity)**.** *Let $n \geq d \geq 1$, $e \geq 1$, and $r \geq 2$. For a generic ordered d-tuple $(f_1, \ldots, f_d) \in H_{n,e}^d$, the forms $f_1^r, \ldots, f_d^r$ are linearly independent, and*

$$X_{n,e,r} \cap \mathbb{P}\operatorname{span}\{f_1^r, \ldots, f_d^r\} = \{[f_1^r], \ldots, [f_d^r]\},$$

as a reduced zero-dimensional scheme.

For a generic network, the span of the output coordinates determines the span of the powers of the last hidden preactivations. Theorem 1.2 recovers those preactivations up to rescaling and permutation. Applying the same argument to the prefix network proves Theorem 1.1 by induction on the depth.

Section 2 introduces polynomial networks and their multihomogeneity. Section 3 proves Theorem 1.2, and Section 4 applies it to the generic fiber. Section 5 discusses exceptional fibers and the roles of the width assumptions; the final section summarizes the remaining cases.

2. Polynomial networks and multihomogeneity

Let $\Bbbk$ be an algebraically closed field of characteristic zero. For $n, e \geq 1$, define

$$H_{n,e} := \Bbbk[x_1, \ldots, x_n]_e,$$

for the vector space of homogeneous degree-e polynomials. If $v = (v_1, \ldots, v_s)^T$ is a vector of polynomials and $r \geq 1$, set

$$\sigma_r(v) := (v_1^r, \ldots, v_s^r)^T.$$

Fix an architecture $\boldsymbol{d} = (d_0, \ldots, d_L)$ and activation degrees $\boldsymbol{r} = (r_1, \ldots, r_{L-1})$. A parameter is a tuple

$$W = (W_1, \ldots, W_L), \qquad W_\ell \in \Bbbk^{d_\ell \times d_{\ell-1}}.$$

For $1 \leq \ell \leq L-1$, let $\sigma_\ell := \sigma_{r_\ell}$. Define the associated network recursively by

$$h_0(x) = x, \qquad z_\ell(x) = W_\ell h_{\ell-1}(x), \qquad h_\ell(x) = \sigma_\ell(z_\ell(x)),$$

for $1 \le \ell \le L-1$, and

$$p_W(x) := W_L h_{L-1}(x). \tag{1}$$

Thus p_W is a d_L-tuple of forms of degree $r_1 \cdots r_{L-1}$.

Define the parameter map by

$$\Psi_{\boldsymbol{d},\boldsymbol{r}}(W) := p_W.$$

For $1 \le \ell \le L-1$, let D_ℓ be an invertible diagonal matrix and let P_ℓ be a permutation matrix of size d_ℓ. Multihomogeneity refers to the replacements

$$W_1 \longmapsto P_1 D_1 W_1, \tag{2}$$

$$W_\ell \longmapsto P_\ell D_\ell W_\ell D_{\ell-1}^{-r_{\ell-1}} P_{\ell-1}^T \qquad (2 \le \ell \le L-1), \tag{3}$$

$$W_L \longmapsto W_L D_{L-1}^{-r_{L-1}} P_{L-1}^T. \tag{4}$$

Indeed,

$$\sigma_\ell(P_\ell D_\ell v) = P_\ell D_\ell^{r_\ell} \sigma_\ell(v),$$

so the factors cancel between consecutive layers and leave p_W unchanged. This is precisely neuron permutation and rescaling [5, Lemma 13]. A fiber is *characterized by multihomogeneity* if all its parameters lie in one orbit under the replacements (2)–(4). The continuous family of such replacements has dimension $\sum_{\ell=1}^{L-1} d_\ell$.

3. Power-span rigidity

We first isolate a standard special-fiber principle in a form suited to the power incidence used below.

Lemma 3.1 (Persistence of an exhausted reduced fiber)**.** *Let B be a noetherian variety, let $\pi : Z \to B$ be a projective morphism of finite type, and let $s_1, \dots, s_d : B \to Z$ be sections. Suppose that, over a point $b_0 \in B$, the fiber Z_{b_0} is the reduced union of the d distinct points $s_1(b_0), \dots, s_d(b_0)$. Then there is an open neighborhood U of b_0 such that, for every $b \in U$, the fiber Z_b is the reduced union of the d distinct points $s_1(b), \dots, s_d(b)$.*

Proof. After shrinking B, the sections are pairwise disjoint. Upper semicontinuity of fiber dimension and the zero-dimensionality of Z_{b_0} allow us to shrink once more so that every fiber is zero-dimensional. The restricted morphism is projective and quasi-finite, hence finite [4, Chapter II].

Set $\mathcal{A} := \pi_* \mathcal{O}_Z$, a coherent $\mathcal{O}_B$-module. For a finite morphism, the length of the scheme-theoretic fiber over b is

$$\dim_{\kappa(b)} \big(\mathcal{A} \otimes_{\mathcal{O}_B} \kappa(b) \big).$$

For a finitely presented module this dimension is upper semicontinuous, as is seen locally from a presentation matrix and its ranks. At b_0 the length is d, so after another shrinking every fiber has length at most d. The d disjoint sections contribute d distinct points to every fiber, each with length at least one. Hence every fiber has length exactly d, has no further points or nilpotent structure, and is exhausted by the sections. □

For $e, r \ge 1$, consider the morphism

$$\nu_{e,r} : \mathbb{P}(H_{n,e}) \longrightarrow \mathbb{P}(H_{n,er}), \qquad [f] \longmapsto [f^r],$$

Its image is $X_{n,e,r}$, endowed with the reduced induced scheme structure.

Lemma 3.2 (The power embedding)**.** *The morphism $\nu_{e,r}$ is a closed immersion onto $X_{n,e,r}$.*

Proof. The morphism is universally injective. Indeed, after every algebraically closed field extension, proportionality of f^r and g^r implies proportionality of f and g by unique factorization. At a geometric point $[f]$, the differential sends the tangent class represented by h to the class represented by

$$rf^{r-1}h.$$

If this class vanishes, then $rf^{r-1}h$ is proportional to f^r. Since the characteristic is zero, cancellation gives $h \in \operatorname{span}\{f\}$. The differential is therefore injective at every geometric point, so the morphism is unramified.

A universally injective unramified morphism is a monomorphism: its diagonal is a surjective open immersion and hence an isomorphism. The morphism $\nu_{e,r}$ is also proper because its source is projective and its target is separated. A proper monomorphism is a closed immersion. Its image is reduced because its source is reduced, so the image scheme is $X_{n,e,r}$ with the specified scheme structure. □

Lemma 3.3 (The coordinate power plane)**.** *Let $n \geq d \geq 1$, $e \geq 1$, and $r \geq 2$. Set $f_i := x_i^e$ for $1 \leq i \leq d$ and*

$$\Lambda_0 := \mathbb{P}\operatorname{span}\{x_1^{er}, \ldots, x_d^{er}\} \subseteq \mathbb{P}(H_{n,er}).$$

Then

$$\nu_{e,r}^{-1}(\Lambda_0) = \{[x_1^e], \ldots, [x_d^e]\},$$

as a reduced scheme.

Proof. Suppose that

$$g^r = \sum_{i=1}^{d} c_i x_i^{er}, \tag{5}$$

for a nonzero $g \in H_{n,e}$. If two coefficients, say c_i and c_j, are nonzero, restrict (5) to the coordinate plane on which all variables except x_i, x_j vanish. The right-hand side is a nonzero binary form

$$c_i x_i^{er} + c_j x_j^{er},$$

with er distinct linear factors, because $\operatorname{char} \Bbbk = 0$. Every irreducible factor of the rth power on the left has multiplicity divisible by r, a contradiction. Thus exactly one c_i is nonzero, and unique factorization gives $g \propto x_i^e$. This proves the set-theoretic assertion.

It remains to check reducedness. At $[x_i^e]$, a projective tangent vector represented by $h \in H_{n,e}$ maps under the differential of $\nu_{e,r}$ to the class of

$$rx_i^{e(r-1)}h.$$

If this lies in $\operatorname{span}\{x_1^{er}, \ldots, x_d^{er}\}$, divisibility by $x_i^{e(r-1)}$ forces all terms except the x_i^{er} term to vanish, and then h is proportional to x_i^e. Hence the tangent space of the fiber is zero at every point, so the finite fiber is reduced. □

Proof of Theorem 1.2. Let $B \subseteq H_{n,e}^d$ be the open subset on which $f_1^r, \ldots, f_d^r$ are linearly independent. Over B, form the closed incidence correspondence

$$\mathcal{I} := \left\{ \big((f_1, \ldots, f_d), [g]\big) \in B \times \mathbb{P}(H_{n,e}) : g^r \in \operatorname{span}(f_1^r, \ldots, f_d^r) \right\}.$$

Closedness follows by imposing the vanishing of the $(d+1)\times(d+1)$ minors of the coefficient matrix with columns $f_1^r,\dots,f_d^r,g^r$. The projection $\pi:\mathcal{I}\to B$ is projective. It has d pairwise disjoint tautological sections

$$s_i(f_1,\dots,f_d)=\big((f_1,\dots,f_d),[f_i]\big).$$

The coordinate tuple $(x_1^e,\dots,x_d^e)$ belongs to B, and Lemma 3.3 says that its fiber is precisely the reduced union of the d section points. Lemma 3.1 therefore gives a nonempty open subset of B over which every fiber is exactly the reduced union of the tautological sections. Lemma 3.2 identifies these fibers with the corresponding scheme-theoretic intersections in $X_{n,e,r}$ and proves the theorem. □

Remark 3.4. Theorem 1.2 concerns a generic secant plane arising from d generic points of the power variety, not an arbitrary $(d-1)$-plane in the ambient projective space. The coordinate fiber supplies an explicit unisecant witness, and projectivity rules out branches of additional power points appearing near that witness.

4. The generic fiber

We now apply power-span rigidity one layer at a time.

Proof of Theorem 1.1. We induct on L, the number of linear layers. For a prefix ending just before the last activation, write

$$q_W=(q_1,\dots,q_d)^T:=W_{L-1}h_{L-2}.$$

Its entries are forms of degree

$$R:=r_1\cdots r_{L-2},$$

with the convention $R=1$ when $L=2$. The full output is

$$p_W=W_L\sigma_{L-1}(q_W). \tag{6}$$

Let $\mathcal{V}_{R,r_{L-1}}\subseteq H^d_{n,R}$ be the open neighborhood of the coordinate tuple constructed in the proof of Theorem 1.2. Its inverse image under the prefix parameter map is nonempty: the choice

$$W_1=\begin{bmatrix}I_d & 0\end{bmatrix},\qquad W_2=\cdots=W_{L-1}=I_d,$$

produces $q_W=(x_1^R,\dots,x_d^R)^T$. Hence the inverse image is a nonempty open subset of the irreducible prefix parameter space.

For $L>2$, intersect this open set with the nonempty induction open set for the prefix architecture $(n,d,\dots,d,d)$ and degrees $r_1,\dots,r_{L-2}$. For $L=2$, no induction condition is needed. Finally require W_L to have column rank d, require W_1 to have row rank d, and require the square matrices $W_2,\dots,W_{L-1}$ to be invertible. These rank conditions are nonempty and open. They also ensure that the multihomogeneous rescalings and permutations act freely. The resulting set $\mathcal{U}_L$ is nonempty and open.

Fix $W\in\mathcal{U}_L$, and suppose that $p_{\widetilde{W}}=p_W$. Set $q:=q_W$, $\widetilde{q}:=q_{\widetilde{W}}$, and $r:=r_{L-1}$. Define the span of the output coordinates by

$$S(p_W):=\operatorname{span}\{(p_W)_1,\dots,(p_W)_m\}.$$

Since W_L has column rank d and $q_1^r,\dots,q_d^r$ are linearly independent, this span satisfies

$$S(p_W)=\operatorname{span}\{q_1^r,\dots,q_d^r\},\qquad \dim S(p_W)=d. \tag{7}$$

The same output is a linear combination of $\widetilde{q}_1^r, \dots, \widetilde{q}_d^r$. Thus

$$S(p_W) \subseteq \operatorname{span}\{\widetilde{q}_1^r, \dots, \widetilde{q}_d^r\}.$$

The right-hand side has dimension at most d, so equality holds and the $\widetilde{q}_j^r$ are linearly independent. Each projective point $[\widetilde{q}_j^r]$ therefore lies in

$$X_{n,R,r} \cap \mathbb{P}S(p_W).$$

By Theorem 1.2, this intersection consists exactly of $[q_1^r], \dots, [q_d^r]$. Independence prevents repetitions, so there are a permutation matrix P and nonzero scalars $\lambda_1, \dots, \lambda_d$ such that

$$\widetilde{q} = PDq, \qquad D := \operatorname{diag}(\lambda_1, \dots, \lambda_d).$$

Here we used unique factorization once more to pass from proportional rth powers to proportional forms. Comparing (6) and using the independence of $q_1^r, \dots, q_d^r$ gives

$$\widetilde{W}_L = W_L D^{-r} P^T. \tag{8}$$

Normalize the competing prefix by replacing only its final matrix with

$$\widehat{W}_{L-1} := (PD)^{-1} \widetilde{W}_{L-1}.$$

The resulting prefix output is exactly q. If $L = 2$, this gives $\widehat{W}_1 = W_1$. If $L > 2$, the induction hypothesis shows that the normalized prefix is obtained from $(W_1, \dots, W_{L-1})$ by multihomogeneity. Restoring the factor PD in the last hidden layer and using (8) gives the replacements (2)–(4), with $P_{L-1} = P$ and $D_{L-1} = D$. Hence every representation of p_W in the fiber is obtained by multihomogeneity. □

Corollary 4.1 (Equi-width fibers)**.** *Let $d \geq 2$ and $L \geq 2$. For every choice of activation degrees $r_1, \dots, r_{L-1} \geq 2$, the generic fiber of the equi-width architecture $(d, d, \dots, d)$ is characterized by multihomogeneity. In particular, a generic equi-width network is globally identifiable up to hidden-neuron rescaling and permutation.*

Proof. Apply Theorem 1.1 with $n = m = d$. □

Corollary 4.2 (Dimension of the neurovariety)**.** *Under the hypotheses of Theorem 1.1, the generic fiber has dimension $(L-1)d$, and the affine neurovariety has dimension*

$$nd + (L-2)d^2 + md - (L-1)d. \tag{9}$$

In particular, the parametrization is nondefective relative to the multihomogeneity count.

Proof. The generic fiber is characterized by multihomogeneity. On the rank-open set used in the proof, these replacements act freely: $P_1 D_1 W_1 = W_1$ and the full row rank of W_1 force $P_1 D_1 = I_d$. The invertibility of $W_2, \dots, W_{L-1}$ then forces $P_\ell D_\ell = I_d$ successively. The finite permutation choices do not affect dimension, while the diagonal rescalings form $(\Bbbk^\times)^{(L-1)d}$. The parameter space has dimension

$$nd + (L-2)d^2 + md.$$

The fiber-dimension theorem now gives (9) by subtracting the orbit dimension. □

5. Exceptional fibers and limitations

The word "generic" in Theorem 1.1 cannot be replaced by "every."

Example 5.1 (A nongeneric collision)**.** The following collision, due to Usevich, Borsoi, Dérand, and Clausel [8, Example 41], occurs already for width two and two quadratic activations. Let

$$H := \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \qquad C := \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix}.$$

Consider the two parameter tuples

$$(I_2, H, I_2) \qquad \text{and} \qquad \left(H, \frac{1}{2}H, C\right).$$

They define the same polynomial map

$$\begin{bmatrix} (x_1^2 + x_2^2)^2 \\ (x_1^2 - x_2^2)^2 \end{bmatrix},$$

but they are not related by multihomogeneity: the first-layer replacement would require $H = PD$ for a permutation matrix P and a diagonal matrix D. Define the exceptional last preactivations by

$$q_1 := x_1^2 + x_2^2, \qquad q_2 := x_1^2 - x_2^2.$$

They satisfy

$$(2x_1x_2)^2 = q_1^2 - q_2^2.$$

Thus the secant line $\mathbb{P}\,\mathrm{span}\{q_1^2, q_2^2\}$ contains a third point of the power variety. This pinpoints the exceptional incidence excluded by Theorem 1.2; it does not contradict generic identifiability.

The inequalities $n, m \geq d$ have different roles. The input inequality supplies the explicit coordinate witness $(x_1^R, \ldots, x_d^R)$ at every depth. The output inequality makes the generic output matrix injective on the d hidden coordinates, so that the span of the output entries equals the entire span of the hidden powers. Neither condition is asserted to be necessary. When $m < d$, the generic fiber may still be characterized by multihomogeneity, but the output no longer reveals the full hidden power span. This is precisely the regime in which tensor-identifiability and varieties-of-powers methods become essential [8, 7].

The proof also uses the pure-power activation essentially. For an arbitrary polynomial activation, the image of a hidden form is not a point of a single projective power variety, and the multihomogeneity replacements change. Extending the secant-plane argument to shifted or mixed-degree power varieties is a natural problem.

6. Conclusion

Finkel et al. characterize generic fibers by multihomogeneity above a quadratic Newman–Slater bound, while Crăciun's polynomial-*abc* argument supplies another high-degree route. Usevich et al. prove finite identifiability much more broadly at degrees at least two, but this still permits several inequivalent parameterizations of a generic function. For constant hidden width, Theorem 1.1 rules out that remaining ambiguity at every layerwise degree at least two: the generic fiber is exactly one multihomogeneity orbit.

The constant-width and endpoint assumptions ensure that the full hidden power span is visible at every layer. Varying hidden widths requires compatible power-span statements of different cardinalities, while a smaller output width requires recovery of a hidden secant plane from a proper subspace. These are natural next cases for combining the present projective argument with the localization methods of Usevich, Borsoi, Dérand, and Clausel [8].

References

1. Luca Chiantini and Ciro Ciliberto, *Weakly defective varieties*, Transactions of the American Mathematical Society **354** (2002), no. 1, 151–178.
2. Alexandru Crăciun, *Linear independence of powers for polynomials*, 2025, arXiv:2507.10163v1, 14 July 2025.
3. Bella Finkel, Jose Israel Rodriguez, Chenxi Wu, and Thomas Yahl, *Activation degree thresholds and expressiveness of polynomial neural networks*, Algebraic Statistics **16** (2025), no. 2, 113–130.
4. Robin Hartshorne, *Algebraic geometry*, Graduate Texts in Mathematics, vol. 52, Springer-Verlag, New York, 1977.
5. Joe Kileel, Matthew Trager, and Joan Bruna, *On the expressive power of deep polynomial neural networks*, Advances in Neural Information Processing Systems, vol. 32, 2019.
6. Kaie Kubjas, Jiayi Li, and Maximilian Wiesmann, *Geometry of polynomial neural networks*, Algebraic Statistics **15** (2024), no. 2, 295–328.
7. Alex Massarenti and Massimiliano Mella, *The Alexander–Hirschowitz theorem for neurovarieties*, 2025, arXiv:2511.19703v1, 24 November 2025.
8. Konstantin Usevich, Ricardo Borsoi, Clara Dérand, and Marianne Clausel, *Identifiability of deep polynomial neural networks*, Advances in Neural Information Processing Systems, vol. 38, 2025, arXiv:2506.17093v3, 30 January 2026.