SINGULARITY VANISHING THEOREM FOR NEUROVARIETIES AND CAUSALITY

Anonymous authors
Paper under double-blind review

ABSTRACT

We analyze the identifiability problem in general machine learning and establish the singularity-vanishing theorem. We first extend the classic Alexander theorem to a foundational double-point theorem for generic polynomial networks by characterizing the singular locus of their associated ideal sheaves. Building upon these algebro-geometric insights, we revisit and reformulate causal representation learning (CRL) to provide the first formal proof of identifiability under strictly non-injective transformations. We introduce algebraic extension modules paired with models or a hypothetical functional space, and show that valid interventions yield a sufficient number of independent conditions to obstruct the double-point scheme, thereby establishing a general singularity vanishing theorem and resolving a long-standing conjecture in identifiable representation learning. Furthermore, by anchoring representation learning in rigid algebraic geometry, our work implicitly echoes the recent call by Fluri et al. for the "Hilbert Problems of AI"[1] from geometry. We envision that our results provide a novel, grounded theoretical perspective on causality and world models—the ultimate objectives of sciences such as biology and physics—while potentially opening new avenues for theoretical guarantees across broader machine learning domains.

1 INTRODUCTION

Identifiability underpins the accurate, ground-truth recovery of features and representations in modern machine learning, capitalizing on the learnability of hypothetical models. While existing literature addresses identification in settings spanning neural networks with a limited number of activations and latent variable modeling, an overarching geometric understanding of the underlying mechanisms remains elusive, leaving the connection between generic functional identifiability in neural networks and latent representation unresolved. The geometry of learning models (i.e., statistical or graphical) plays a pivotal role in their behavior in identification, computation, and optimization, and has recently received increasing attention (Bronstein et al., 2021; Marchetti et al., 2025a). A large class of models is equipped with subtle geometric properties that reflect the underlying generative mechanism as a general principle and that provide useful information for curating the optimization process and suggesting better algorithm design. At their heart, neural networks are representative of several learning models with desirable structural and geometric properties and have largely spurred deep study over a decade. Neural networks have witnessed astounding success in modern deep learning tasks that feature learning (Radhakrishnan et al., 2024), language model-based reasoning, and applications across various scientific fields (Wei et al., 2025; Li et al., 2025). Driven by the growing interest in large language models (LLMs), it has continuously received more attention to understanding how language models tend to think and respond, via chain-of-thought(CoT) reasoning (Huang et al., 2025; Duan et al., 2022). The generative mechanism is learned by the neural network architecture and has served as the bedrock for various machine learning areas, such as general representation learning and causal disentanglement (Schölkopf et al., 2021).

Identifiability theory of unobservable causal variables, processes, and parameters contributes a major part of causal representation learning, and, for decades, scientists have long believed that identifi-

[1] A recent white paper by Fluri et al. advocates for transforming artificial intelligence into a theory-rigorous scientific field, rather than one merely driven by empirical applications. For a non-exhaustive list outlining these "Hilbert Problems of AI," we refer the readers to (Felsberg et al.)

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

ability should be researcher-led or coupled with expertise in theoretical domains. Once a model is hypothesized, the standard regime is to prove identifiability through *either interventions or passive variability* and then harness the *inductive bias* to design an algorithm. The resulting algorithms, with such an identifiability guarantee, unintentionally produce a set of inductive biases that span abstraction of plausible assumptions, actionable interventions, and necessary variability in datasets. Another neglected utility of identifiability lies in probing inductive biases for world models from a performance-driven perspective. By probing hidden or implicit inductive biases, scientists aim to extract machine-compressed, interpretable physical laws (Liu et al., 2026). Crucially, such learning targets must be strictly identifiable to ensure that any proposed machine learning models yield physically meaningful and reliable representations.

Despite recent progress, a rigorous mathematical characterization of the generative mechanism and representation in the data remains elusive. In particular, closing the gap between the algebraic structure of feature embeddings and the geometry of the neural networks remains a fundamental open problem.

1.1 Main Contribution

In this paper, we consider the problem of identifying a unique representation of an arbitrary neural network from the response data under both generic and non-generic conditions. Then, we extend our results to the non-generic conditions of random variables and give the first formal proof of identifiability under sufficiently multiple interventions. We present an advanced, systematic theoretical framework for neural network Theory, representation learning, and causality through a consistent mathematical chain in algebraic geometry. The core theoretical contributions of this work are crystallized in a triptych of theorems that bridge classical algebraic geometry with the frontier of neural representation theory.

Beyond Alexander-Hirschowitz. As a foundational milestone, our work establishes a novel path to global identifiability for PNNs. By analyzing the *zero-dimensional schemes* of double points within homogeneous polynomial spaces, we provide a proof technique that diverges from classical Kruskal-rank arguments and the standard application of the tensor decomposition. This approach allows us to handle structured dependencies that previously fell into the defective gaps of established geometric methods.

Algebro-geometric foundation for causality and world models The global identifiability conjecture often encounters a "bottleneck" where non-generic configurations lead to algebraic breakdown. Our theorems show that this phenomenon is not an idiosyncrasy of neural networks but a fundamental property of feature learning in latent variable models. We provide a unified geometric characterization of this *intrinsic defectiveness*, offering a formal language to describe the interplay between generic identifiability of latent representation learning and algebraic collapse across diverse architectural paradigms.

Autonomous identifiability verification. Leveraging algebraic geometry and sheaf cohomology, we establish the first vanishing-singularity theorem for neurovarieties, moving strictly beyond classical tensor/tensor-ring decompositions and heuristic Kruskal rank conditions to rigorously characterize the identifiability of mechanisms, parameters, and structures across deep network layers. Importantly, we accomplish a fundamental conceptual leap by recasting the elusive identifiability dilemma of machine learning into a closed-form, pure mathematical problem; by modeling layer-wise propagation as morphisms between ideal sheaves, we bridge empirical neural dynamics with exact algebraic structures to form a complete theoretical system. Finally, by formulating deep learning identifiability into the language of commutative algebra and vanishing theorems, our framework transforms an ambiguous empirical challenge into a mathematically tractable system, providing a rigorous structural foundation with high potential to be tackled, verified, and automated by AI agents.

2 BACKGROUND

2.1 BASIC NOTATIONS

We are now ready to give notations and formally revisit the concept introduced in the first section. For the remainder of this paper, we work over a complex field $\mathbb{C}^n$, unless further stated, because generic complex results imply generic real results. In contrast, the reverse direction is much more challenging, posing open problems. As a reflection of the Introduction, technical terms for homogeneous polynomials in algebraic geometry will be introduced in the sequel.

We mainly study a general point $P \in \mathbb{C}^n$ at different levels of geometry. A point P is general if it ranges over the Zariski closure of irreducible components of X, denoted by $\overline{X}^{\text{Zar}}$. A projective space $\mathbb{P}^r_{\mathbb{C}}$, defined on the complex field, is identified as the quotient of an affine space $\mathbb{A}^r \setminus \{0\}$ modulo the symmetry group action $\mathbb{G}_m$, i.e., each general point P is identified up to non-zero scalar multiplication. A projective variety $X \subset \mathbb{P}^r$ is *reduced, irreducible, and pure* if and only if it does not contain any nilpotent elements (Chiantini & Ciliberto, 2010). X is called *not* degenerate if it is not contained in any sub-hypersurfaces and its component is of the full dimension in X. As a convention in algebraic geometry, for reduced, irreducible varieties $X_1, X_2, \cdots, X_k$, we obtain the joint variety $J(X_1, X_2, \cdots, X_k)$ of the irreducible forms (we stress that the intersection $\bigcap_{i=1}^{k} X_i$ is generally not irreducible).

Dual forms of polynomials. Let V be an $(n+1)$-dimensional vector space over $\mathbb{C}$, and let $\mathbb{P}^n = \mathbb{P}(V)$ denote the associated projective space. We denote by $S = \text{Sym}(V^\vee) \cong \mathbb{C}[x_0, \ldots, x_n]$ the symmetric algebra of the dual space $V^\vee$, where the d-th graded component $S_d = S^d V^\vee$ represents the vector space of homogeneous polynomials of degree d. Concretely, the elements of the dual space $V^\vee$ act as coordinate functions on V. With a basis $(e_0, \cdots, e_n)$, the space of degree d polynomials S_d is canonically isomorphic to the d-th symmetric power $S^d V^\vee$, where each monomial is formed by the symmetric product of these linear dual functionals. To ground the preceding abstract definitions, we present a latent variable model as a running example in Example A at the end of this paper. This model not only serves as the reader's intuition but also provides a concrete template that will be revisited throughout the paper to illustrate our core technical arguments.

Remark 2.1. In the literature of machine learning, the duality between the symmetric power and its dual is frequently conflated, treating them as effectively equivalent through some implicit choice of basis or inner product. In this work, however, we intentionally preserve this distinction. Maintaining this duality is crucial for our framework, as it allows us to formally distinguish between the algebraic equations (elements of the symmetric algebra) and the geometric objects (such as tensors or secant varieties) upon which they act.

Double points. For a configuration of k distinct points $P = \{p_1, \ldots, p_k\} \in (\mathbb{P}^n)^k$, we define the subscheme $X \subset \mathbb{P}^n$ of a set of closed points supported on P via the ideal sheaf:

$$\mathcal{I}_X = \bigcap_{i=1}^{k} \mathfrak{m}_{p_i}$$

where $\mathfrak{m}_{p_i} \subset \mathcal{O}_{\mathbb{P}^n}$ is the maximal ideal sheaf corresponding to the point p_i. let $\mathcal{O}_{X,p}$ denote the local ring at p_i, and let $\mathfrak{m}_{p_i}$ be its maximal ideal, consisting of all germs of functions vanishing at p_i. The notation $\mathcal{O}_X$ is written for a sheaf of $\mathcal{O}$-module on a topological space X. The quotient $\mathfrak{m}_{P_i}/\mathfrak{m}^2_{P_i}$ captures the first-order vanishing directions of these functions. We write p_i^2 for the first infinitesimal neighborhood of the closed point p_i, which we call *double point* throughout this paper. A double point subscheme encodes both the functional evaluation and the infinitesimal data of its vanishing the ordinary power (symbolic square) (Araújo Dos Santos et al., 2018, Symbolic Square of Ideals)

$$\mathfrak{m}_p^2 := \{f \in S_d \mid f(p) = 0; \nabla f(p) = 0\};$$

and accordingly,

$$\mathcal{I}_X^2 = \bigcap_{i=1}^{k} \mathfrak{m}_{p_i}^2$$

denotes an ideal sheaf supported on the union of points $\{p_1^2, \ldots, p_k^2\} \in (\mathbb{P}^n)^k$ and identifies hypersurfaces through and also singular at all points of P. Numerically, each double point behaves as a zero-dimensional scheme of length $n+1$: one dimension is taken up by the value at the node, and the remaining n dimensions capture the first-order derivatives.

2.2 Neuromanifold and Varieties

Let Ψ be the map that maps the L- tuple of matrices $(W_1, \cdots, W_L)$ to a tuple of a degree r_{L-1} homogeneous polynomial space $(\mathrm{Sym}_{\mathbf{d},\mathrm{r}})^{\times d_L}$ in d_0 unknowns of length d_L. The modulo effect of the group G describes the natural multi-homogeneity that reflects the minimal number of equivalent classes due to the permutation and scaling. In what follows, we provide a formal definition of neuromanifold and its varieties.

Definition 2.2 (Neurovariety)**.** Let $\mathcal{W} \cong \mathbb{C}^{d_1 \times d_0}$ and $\mathcal{C} \cong \mathbb{C}^{d_L \times d_{L-1}}$ be the parameter spaces for the hidden and output layers, respectively. The polynomial neural network defines a parametrization map:

$$\Psi : \mathcal{W} \times \mathcal{C} \longrightarrow \mathrm{Sym}^r(\mathbb{C}^{d_0})^{d_L},$$

which sends the weights $\{W_i, c_{jk}\}$ to the coefficients of the resulting polynomial $F(\mathbf{x})$. The neurovariety $\mathcal{V}_{\mathbf{d},\mathbf{r}}$ is defined as the Zariski closure of the image of this map:

$$\mathcal{V}_{\mathbf{d},\mathbf{r}} = \overline{\mathrm{Im}(\Psi)}.$$

Definition 2.3 (Global identifiability of PNN)**.** A polynomial neural network (PNN) with architecture $\mathbf{d} = (d_0, d_1, \ldots, d_L)$ and degree r is said to be *globally identifiable* if, for a generic output f of the map $\Psi_{\mathbf{d},r}$:

$$\mathbb{C}^{d_1 \times d_0} \times \mathbb{C}^{d_2 \times d_1} \times \cdots \times \mathbb{C}^{d_{L-1} \times d_{L-2}} \times \mathbb{C}^{d_L \times d_{L-1}} \to (\mathrm{Sym}_{d_0,r})^{\times d_L}$$

the fiber

$$\mathcal{F}_f := \Psi_{\mathbf{d},r}^{-1}(f) = \{w \mid \Psi_{\mathbf{d},r}(w) = f\}.$$

modulo the natural symmetry group G is finite, i.e., $\dim\left(\Psi_{\mathbf{d},r}^{-1}(f)/G\right) = 0$

3 Results of Vanishing Singularity Theorem

In this section, we present our main results for neurovarieties and their transcendental extensions.

3.1 Vanishing Singularity for Algebraic Networks

Global identifiability of polynomial neural networks, suggested by Definition 2.3, indicates that the neurovariety attains its expected dimension (Massarenti & Mella, 2025; Usevich et al., 2025). This relationship can be revealed by studying the tangent spaces of the corresponding neurovarieties. However, as an independent contribution of our work, we focus on the double points of neurovarieties; thus, we will not discuss the tangent space in detail and instead intend to work on establishing *Vanishing Theorem* (Alexander & Hirschowitz, 1997; Chandler, 2000; Alexander & Hirschowitz, 2000) of generic double points for algebraic neurovarieties.

Important Remark. Polynomial neural networks (PNNs) exhibit two fundamental distinctions from classical rank-1 Veronese embeddings introduced in Eq. (D.1). First, a PNN produces a vector of homogeneous polynomials where each component is defined by a linear combination of points on a Veronese variety $\mathbb{P}^N$, rather than a single global embedding. Second, the compositional structure implies that the intermediate-layer polynomials are constrained to a locus of points chosen generically only to the preceding layer's image (e.g., within specific twisted curves). Consequently, the resulting coefficients are not generic in the ambient parameter space of all polynomials, instead confined to a structured subvariety determined by the network's architecture $\mathbf{d}$ and degrees $\mathbf{r}$.

At present, we adjust all notations to neurovarieties. The double-point subscheme $\mathcal{I}_X$ has Euclidean dimension k and the polynomial dual space has projective dimension N. Therefore, the polynomial neural network lies in the final projective dual space to S_r, and can be compactly represented as

a summation of a finite k number of degree r dual forms in $n+1$ unknowns $[x_0 : \cdots : x_n]$ of projective dimension n:

$$F(x) \in \mathcal{O}_{\mathbb{P}^n}(d) = \sum_{i=1}^{k} \lambda_i L_i(x)^r, \tag{3.1}$$

where each $\lambda_i L_i^r := \lambda_i \langle \mathbf{a}^{\otimes r}, x^{\otimes r} \rangle = \langle \lambda_i \mathbf{a}^{\otimes r}, x^{\otimes r} \rangle$ is a generic dual point to k-th Veronese. Consider a generic collection X of double points and its ideal sheaf $\mathcal{I}_X^2$, if X imposes k independent conditions on $\mathcal{O}_{\mathbb{P}^n}(r)$, then $F(x)$ is identified with the global section $H^0(\mathcal{I}_X(r))$ derived by a long exact sequence in Eq. (D.3). Alexander and Hirschowitz, in their successive works (Alexander, 1988; 1992; Alexander & Hirschowitz, 1995; Alexander & Hirschowitz, 2000) proved that imposing independent conditions on the same number of singular points suffices for the identifiability of such a polynomial, except for a few very unlikely cases. Therefore, the result can be generalized to neurovarieties $\mathcal{V}_{\mathbf{d},\mathbf{r}}$, where $\mathbf{d}, \mathbf{r}$ characterizes the network architecture.

Theorem 3.1 (Fixed Level Veronese Theorem)**.** *For a network architecture* $(\mathbf{d}, \mathbf{r})$ *and a generic collection of* k *double points* $X \subset \mathbb{P}^{d_{L-1}}$*, the conditions imposed on the structure sheaf* $\mathcal{O}_{\mathbb{P}^{d_{L-1}}}(r_{L-1})$ *are independent, implying the vanishing of the first cohomology group:* $H^1(\mathcal{I}_X^2(r_{L-1})) = 0$*. Consequently, except for the exceptional cases classified by the Alexander-Hirschowitz Theorem, the following sequence is exact:*

$$0 \longrightarrow H^0(\mathcal{I}_X^2(r_{L-1})) \longrightarrow H^0(\mathcal{O}_{\mathbb{P}^{d_{L-1}}}(r_{L-1})) \xrightarrow{\phi} H^0(\mathcal{O}_X^{(2)}(r_{L-1})) \longrightarrow 0 \tag{3.2}$$

Proof. By interpreting the network output $F(x)$ as a generic linear combination of linear forms, we identify the basis with the dual of the L-th layer Veronese variety. The weight matrix W_L maps these generic vectors to a subspace of the projective space. Since the parameters are generic and d_{L-1} avoids the AH-exceptional dimensions, the map ϕ (the evaluation at fat points) is surjective, completing the exact sequence. □

For a classic Veronese embedding $\mathcal{V} \subset \mathbb{P}^N$, Theorem 3.1 yields the same result for generic hyperplanes. Notably, the gluing sheaves ensure the identification of global sections $\lambda_i \mathbf{a}^{\otimes r_{L-1}} = (\lambda_i \mathbf{a})^{\otimes r_{L-1}}$ for generic λ_i. This global identifiability is shown in our motivating Example A. However, a crucial caveat for multi-layer neurovarieties remains: while the global section $H^0(\mathcal{O}_X^{(2)}(d_{L-1}))$ is fixed by the dual representation $\langle \mathbf{a}^{\otimes}, \mathbf{x}^{\otimes} \rangle$, this does *not* guarantee the uniqueness of the constituent weight matrices $W_{L-1}, \ldots, W_1$, because evaluation vector $\mathbf{a}$ is obtained from column composition of preceding layer weight matrix W_l for $l < L$. Such ambiguities, modulo the group symmetry, arise from the composition effects of the *levels of Veronese*, which we explain as follows. Thanks to the construction of examples in Massarenti & Mella (2025), we take this opportunity to interpret this term from the perspective of double points and their induced exact diagram. For better illustration, we provide a concrete Example C.1 to show how layer-wise composition of neurovarieties lead to the identifiability problem.

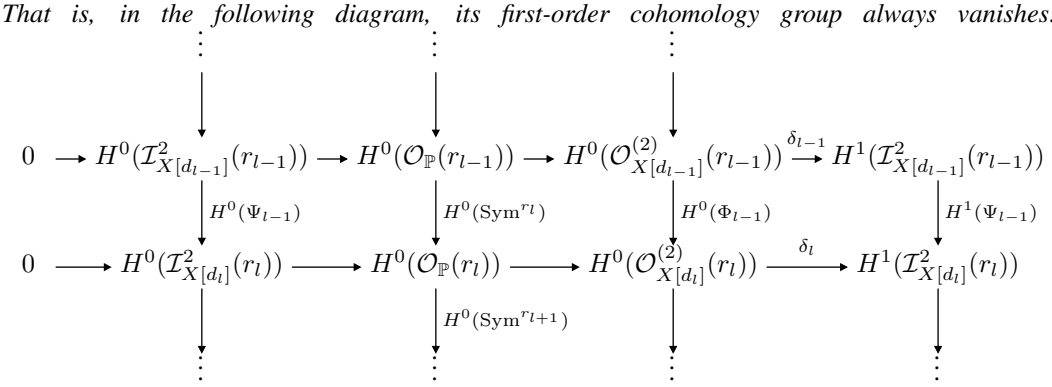
Working over the universal coordinate ring R that encompasses all potential hidden dimensions, Proposition 3.2 is the key to understanding the evolution of network dimensions, and it lays out the foundation of our main results.

Proposition 3.2 (Morphisms between Layer-wise Ideal Sheaf)**.** *The layer-wise propagation defines a sequence of morphisms between ideal sheaves* $\mathcal{I}^2_{X[d_l]}$*, given by a generic collection of* d_{l-1} *double points. The diagram is commutative.*

Proof. Proof of Proposition 3.2 requires knowledge and terms beyond simple algebra, and thus we delegate the formal version to Appendix D. □

Theorem 3.3 (Cohomology Diagram for Layer-wise Neurovarieties)**.** *Fixing a level of Veronese embedding, the global identifiability of a polynomial neural network is attained if, for any two-layer Veronese, the double-point subscheme* $\mathcal{I}^2_{X[d_{l-1}]}(r_{l-1})$ *sufficiently imposes independent conditions to the neurovariety sheaves* $\mathcal{O}_{\mathbb{P}}(r_{l-1})$*.*

That is, in the following diagram, its first-order cohomology group always vanishes:



4 From Vanishing Singularity to Causality and Representation

In Section 3, we constructed an end-to-end theory of generic neurovarieties for polynomial neural networks, securing their global identifiability. Crucially, this problem boils down to determining whether the layer-wise double-point subscheme X imposes independent conditions on the global sections of twisted sheaves $\Gamma(\mathcal{O}_{\mathbb{P}}, \cdot)$—a condition elegantly expressible as a sheaf cohomology vanishing theorem via proper diagrammatic extensions. However, recovering finer structural details exposes a fundamental "genericity gap": observations $[z]$ governed by data distributions are fundamentally non-free and distribution-bound, whereas abstract neurovarieties operate on free, generic coordinates $[x]$ over sheaves. Relying solely on standard vanishing theorems is insufficient to bridge this transition from constrained distributions to general algebraic structures. We demonstrate that the recovery of causal mechanisms is fundamentally a quest for higher-order vanishing theorems. By lifting distribution-bound variables into higher-order algebraic and analytic clusters, we prove a series of rigorous theorems resolving the identifiability of causal mechanisms.

4.1 Partial Genericity from Module Extensions

The procedure we discussed remains valid as long as the evaluation map induces a dual function with free indeterminates such that the δ-functor can glue localized information together at any generic point. However, a more complex scenario arises when causal variables evolve: a set of random (causal) variables is sampled from a distribution $\mathcal{D}$ via some parameterization, by Definition 4.1, which fails to acquire full genericity over a module $\mathcal{O}_X$.

Definition 4.1 (Partial Genericity)**.** Let $\mathscr{D}$ be a set of samples drawn from a distribution $\mathcal{D}$ representing an unknown transformation $X = \mathcal{T}(Z)$. Z denotes a set of random (causal) variables, and the map $\mathcal{T}$ is generic. The system

$$\mathcal{S} := \{\mathscr{D}, \mathcal{D}, \mathcal{T}\}$$

is said to be of partial genericity if there exists an extension of its underlying modules such that $\mathcal{S}$ can be embedded into a fully generic system $\mathcal{S}_{\mathrm{Exd}}$ and the evaluation operator ev_P from the variable-constrained field to the corresponding extension is surjective.

Partially generic systems delineate a connection between the form $F(x) = \sum_{j=0}^{2} a_j y_j(\mathcal{P}_{generic}) = 0$ (See Eq.(D.1)) associated with types of neural network architectures and another popular problem in representation learning in the form of $X = \mathcal{T}(Z)$ considering a non-injective transformation.

Recall the motivating Example A, which demonstrates a map to one of the trivial extensions:

Corollary 4.2. *Let $X = FZ$ be a proper subvariety $\mathcal{V}^*$ of a ring $K(\cdot)$, then $\mathcal{V}^* \mapsto \kappa_d(\mathcal{V}^*)$ is a natural map by introducing the infinite-order generic coordinates over the ring $K_n(\kappa_1, \kappa_2, ..., \kappa_N), n \to \infty$. We identify the cumulant operator $\kappa_d(\cdot)$ as the realization of an extension functor $\mathcal{V}^* \otimes_K$ that maps a constrained subvariety into a formal power series ring $K[[\kappa_1, \kappa_2, \dots]]$.*

This functor effectively replaces the lack of genericity in the samples with infinite-order algebraic independence in the moment space, thereby satisfying the transversality conditions for global identifiability. Thanks to this extension, we can restate a more general identifiability problem for linear causal disentanglement.

Proposition 4.3. *Given a linear causal disentanglement problem $X = FZ$, where the mixing matrix $F \in \mathbb{R}^{p\times q}$ is generic but unnecessarily injective, and the variable Z follows a causal graph with parameters $Z = (I - \Lambda)^{-1}x$ with $\Lambda \in \mathbb{R}^{q\times q}$ and $x \in \mathbb{R}^q$ a zero-mean noise variable. If the* $\mathrm{Spec}(Z)$ *and* $\mathrm{Spec}(x)$ *are not Artinian, then the LSD identifiability coincides with the identifiability of its modules of the related algebraic extension system $\mathcal{S}_{\mathrm{Ext}}$ over the ring K_n:*

$$\begin{array}{ccccccc} 0 & \longrightarrow & H^0(\mathcal{I}_X^2 \otimes \mathcal{O}_{\mathcal{S}_{\mathrm{Ext}}}) & \longrightarrow & H^0(\mathcal{O}_{\mathcal{S}_{\mathrm{Ext}}}) & \xrightarrow{\phi} & H^0(\mathcal{O}_X^{(2)} \otimes \mathcal{O}_{\mathcal{S}_{\mathrm{Ext}}}) \\ & & & & & \overset{\delta}{\searrow} & \\ & & H^1(\mathcal{I}_X^2 \otimes \mathcal{O}_{\mathcal{S}_{\mathrm{Ext}}}) & \longrightarrow & \cdots & & \end{array} \tag{4.1}$$

where $\mathcal{I}_X^2$ is a new *zero-dimensional subscheme of k double points and $\mathcal{O}_{\mathcal{S}_{\mathrm{Ext}}}$ is a structure sheaf on the dual function space to $\mathcal{S}_{\mathrm{Ext}}$.*

Proof. The non-Artinian nature of $\mathrm{Spec}(A)$ and $\mathrm{Spec}(x)$ ensures that the coordinate ring of the causal variables possesses an infinite transcendence degree over the base field K. We construct the extension $K_n(\kappa_1, \ldots, \kappa_N)$, which interprets the random noise x as a set of formal indeterminates. This extension induces a morphism of sheaves where the original variety is lifted to its jet scheme representation. Consequently, the structure sheaf $\mathcal{O}_{\mathcal{S}_{\mathrm{Ext}}}$ admits a representation as a Symmetric Algebra $\mathrm{Sym}_{\mathcal{S}_{\mathrm{Ext}}}(V)$, whose global sections represent the identifiable parameters $F(I-\Lambda)^{-1}$. The sequence (4.1) remains exact under the extension. The vanishing of the first cohomology group $H^1(\mathcal{I}_X^2 \otimes \mathcal{O}_{\mathcal{S}_{\mathrm{Ext}}}) = 0$ is guaranteed by the Alexander-Hirschowitz theorem applied to the extended generic coordinates. This vanishing implies the surjectivity or injectivity of the evaluation map $H^0(\mathcal{O}_{\mathcal{S}_{\mathrm{Ext}}}) \to H^0(\mathcal{O}_X^{(2)} \otimes \mathcal{O}_{\mathcal{S}_{\mathrm{Ext}}})$ in the long exact sequence. Since δ maps the parameter space to the full observed cumulant space with no kernel (up to the defined equivalence), the coefficients of F and Λ are uniquely determined as sections of $\mathcal{O}_{\mathcal{S}_{\mathrm{Ext}}}$. □

Using Proposition 4.3, we are ready to establish an end-to-end vanishing theorem.

Theorem 4.4. *Let $\mathcal{S}_{\mathrm{Ext}}^{(k)\vee}$ be a collection of extended structure sheaves representing varied causal contexts or interventions. For the evaluation morphism $\bigoplus_k \phi^{(k)}$ induced by the subscheme of double points $X^{(2)}$, there exists a sufficiently large k such that the map:*

$$\bigoplus_k \phi^{(k)} : H^0\left(\bigoplus_k \mathcal{O}_{\mathcal{S}_{\mathrm{Ext}}}\right) \longrightarrow H^0\left(\bigoplus_k \mathcal{O}_X^{(2)} \otimes \mathcal{O}_{\mathcal{S}_{\mathrm{Ext}}}\right) \tag{4.2}$$

is of maximal rank, implying the vanishing of the global section space $H^0(\mathcal{I}_X^2 \otimes \bigoplus_k \mathcal{S}_{\mathrm{Ext}}^{(k)\vee})$ and the local identifiability of the causal parameters.

The proof relies on the dimension count between the parameter space and the constraints imposed by the subscheme $X^{(2)}$. As noted in (Carreno et al., 2024), the "intrinsic degree" of a single observational module eventually saturates, failing to provide additional independent constraints for full disentanglement. Leveraging the distributivity of the cohomology functor over direct sums, we demonstrate that each intervention k contributes a transversal slice in the dual tensor space. By adding double points to these extended modules, each context k provides $q(p+1)$ independent conditions. Since these conditions are additive under the direct sum $\bigoplus_k$, the total rank of the aggregate morphism $\bigoplus_k \phi^{(k)}$ grows linearly with k. For a sufficiently large collection of contexts, the first cohomology group H^1 vanishes (via the Alexander-Hirschowitz theorem), forcing the kernel of the identification problem to become trivial.

4.2 A LINEAR CRITERION FOR PARTIAL IDENTIFIABILITY UNDER MULTI-TARGET INTERVENTIONS

Leveraging the distributivity of the cohomology functor $H^k(\cdot)$, we first apply the tensor product with $\mathcal{I}_X^2$ followed by the extension of $\mathcal{S}_{\mathrm{Ext}}$, which yields the reduced linear system $A = F(I-\Lambda)^{-1}$. We then incorporate interventions such that their induced hypersurfaces are transversal to the base variety defined by $\mathcal{I}_X^2$. This ensures that each additional intervention provides linearly independent

constraints, effectively reducing the dimension of the solution space. We now formulate a precise version of the conjectural statement in the multi-intervention setting, regardless type and number of interventions. As in Section 3.3 of (Carreno et al., 2024), we work after the common permutation and scaling ambiguities across contexts have been fixed, so that the matrices

$$A^{(k)} = F(I - \Lambda^{(k)})^{-1}, \qquad k = 0, 1, \dots, m,$$

are regarded as known.

Proposition 4.5. *Let $\mathcal{A} = \{A^{(0)}, A^{(1)}, \dots, A^{(k)}\}$ be a collection of reduced linear systems $A^{(k)} = F(I - \Lambda^{(k)})^{-1}$ derived from k distinct interventions. Let $\Phi = \bigoplus_k \phi^{(k)}$ be the aggregate evaluation morphism. If the set of interventions is transversal in the sense that the intersection of their respective sheaves*

$$\bigcap_{k=0}^{n} \Gamma(\mathcal{I}_X^2 \otimes \mathcal{S}_{\mathrm{Ext}}^{(k)\vee})$$

consists of a discrete set of points, then the system is globally identifiable for n large enough such that the total codimension matches the dimension of the parameter space $(p \times q) + (q \times q)$

To obtain a mathematically meaningful zero-dimensionality statement, one must fix both the support pattern of the observational model and the intervention design.

Corollary 4.6 (Generic zero-dimensionality criterion)**.** *Assume, in the situation of Theorem E.2, that for every $j \in [q]$ there exists at least one $|E_j| \times |E_j|$ minor of $M[j]$ which is not identically zero on the admissible model class. Then the locus of admissible data*

$$\left(A^{(0)}, A^{(1)}, \dots, A^{(m)}\right)$$

for which the baseline fiber $\mathcal{F}_G\left(A^{(0)}, \dots, A^{(m)}\right)$ is zero-dimensional is Zariski open and dense in the model class. Equivalently, for a generic admissible family of multi-target interventions, the baseline parameters (F, Λ) are uniquely determined.

Proof. For fixed j, the condition $\operatorname{rank} M[j] = |E_j|$ is the non-vanishing of at least one maximal minor of $M[j]$, hence it defines a Zariski-open subset of the admissible model class. By assumption that this subset is nonempty, it is dense. Intersecting these open dense sets over $j = 1, \dots, q$ yields a Zariski-open dense locus on which all matrices $M[j]$ have full column rank. The conclusion now follows from Theorem E.2. □

Remark 4.7 (Insight). The theorem isolates the only genuinely identifiable object in the multi-target soft-intervention regime, namely the *baseline* pair (F, Λ). The key point is that, once the common mixing matrix is eliminated via

$$F = A^{(0)}(I - \Lambda),$$

the compatibility problem decouples *column by column* in Λ. For each source node j, every context contributes only the requirement that an explicit vector lie in the span of the intervened columns of $A^{(k)}$. Projecting onto the corresponding left kernel removes the context-specific nuisance parameters and produces a linear system in the baseline coefficients $(\lambda_{ij})_{i \in \mathrm{ch}_G(j)}$. Proceeding with the reduced linear system, the conjectural "almost zero-dimensionality" is controlled exactly by a rank condition: one gets uniqueness precisely when the accumulated linear constraints on every column are of full rank.

This does not contradict the positive-dimensionality statements in Section 3.3 of (Carreno et al., 2024): those results concern the *full* soft-intervention parameter space, whereas Theorem E.2 concerns only its projection to the observational pair (F, Λ).

Corollary 4.8 (Soft interventions: identifiability of the baseline pair)**.** *Assume that, for each context $k \in [m]$, the intervention is soft on the target set S_k, namely $\Lambda^{(k)}$ is allowed to vary arbitrarily on the rows indexed by S_k and satisfies*

$$\lambda_{ij}^{(k)} = \lambda_{ij} \qquad \text{for all } i \notin S_k,\ j \in [q].$$

Then Theorem E.2 applies without modification. In particular, the set of baseline pairs (F, Λ) compatible with

$$A^{(0)}, A^{(1)}, \dots, A^{(m)}$$

is an affine linear space of dimension

$$\sum_{j=1}^{q} \big(|E_j| - \operatorname{rank} M[j]\big),$$

and the baseline pair (F, Λ) *is uniquely determined if and only if*

$$\operatorname{rank} M[j] = |E_j| \qquad \textit{for all } j \in [q].$$

Moreover, this conclusion concerns only the observational pair (F, Λ)*. It does not imply identifiability of the full family*

$$(F, \Lambda^{(1)}, \ldots, \Lambda^{(m)}),$$

which may still vary in a positive-dimensional compatible set.

5 Conclusion and Discussion

In this work, we investigated the fundamental identifiability of machine learning frameworks. By leveraging polynomial neural networks as a stepping stone, we established the first proof of the vanishing-singularity theorem for neurovarieties. Unlike prior works that rely on direct applications of classical tensor or tensor-ring decompositions or naive Kruskal rank conditions, our geometric framework provides a rigorous departure. Furthermore, we extended these foundational results on neurovarieties to address broader problems in causal representation learning. Beyond foundational machine learning, our geometric framework offers a rigorous stepping stone toward causal world models in scientific discoveries, where addressing non-identifiability under noisy, incomplete data is paramount. By moving beyond static representations, this approach provides the necessary theoretical machinery to capture complex dynamics—such as perturbation responses and temporal trajectories—thereby uncovering latent scientific causal rules and ensuring structural interpretability in high-order systems. Therefore, opening new avenues for the theoretical foundations of machine learning, this work also points to several promising directions for future exploration.

Identifiability of broader physical priors Current paradigms in representation learning and world models largely either hinge on restrictive assumptions—such as injectivity via a diffeomorphism—to guarantee identifiability with presumed existing or viable interventions and variability (Zhang et al., 2024; Ahuja et al., 2023), or rely on specific external, predefined physical laws such as spatial smoothness and stability to avoid learning nonsensical representations (Vafa et al., 2025; Qiu et al., 2026). However, fundamental scientific laws governing phenomena from thermodynamic collapse to symmetry breaking in physics and biology are inherently non-injective, and operate on a more generic hypothetical space beyond human experience and definition (Xing et al., 2026). Furthermore, the requisite inductive biases for such laws remain elusive and often demand independent theoretical development, rendering existing frameworks inapplicable to a vast range of natural phenomena.

Theoretical foundation for world models. Another promising direction lies in probing inductive biases for world models from a performance-driven perspective. By probing hidden or implicit inductive biases, scientists aim to extract machine-compressed, interpretable physical laws (Liu et al., 2026). Crucially, such learning targets must be strictly identifiable to ensure that any proposed machine learning models yield physically meaningful and reliable representations.

A compelling example in biology is the development of cellular foundation models involving data from multiple perturbation donors; indeed, such models are often agnostic as to whether a specific donor is truly informative for constructing a world model (ElSheikh et al., 2026; Zhang et al., 2026; Wang et al.). The relentless inclusion of data donor sources merely increases the volume of data yet fails to enrich the biological context (Wang et al.; Dibaeinia et al.). Superficial data diversity has been criticized for being incapable of mapping the observational data to a landscape where no *identifiable* scientific information is recovered (Xing et al., 2026; Xing & Song; Song et al., 2026). Furthermore, since biological systems are governed by evolutionary dynamics, a true world model must consistently infer identifiable principles regardless of whether incoming data scaling stems from mere volume expansion or genuine diversity and richness.

ACKNOWLEDGMENTS

AI USE STATEMENT

In this work, we used generative AI tools for polishing writing and detecting typos and grammatical issues. We have not used generative AI tools for full generation of theoretical results, and synthetic data generation and experiment-related usage do not apply to this work because our work is completely theory-focused and intensive. Additionally, we used generative AI tools for generating proper non-essential content, such as asking AI to modify the format of footnotes and citation. We have reviewed all AI-assisted work. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

REFERENCES

[2507.05169] Critique of World Model. URL https://arxiv.org/abs/2507.05169.

Kartik Ahuja, Divyat Mahajan, Yixin Wang, and Yoshua Bengio. Interventional Causal Representation Learning. In *International Conference on Machine Learning*, pp. 372–407. PMLR, July 2023. URL https://proceedings.mlr.press/v202/ahuja23a.html. ISSN: 2640-3498.

J. Alexander. Singularités imposables en position générale à une hypersurface projective. *Compositio Mathematica*, 68(3):305–354, 1988. URL http://eudml.org/doc/89939.

J. Alexander. La méthode d'horace éclatée: application à l'interpolation en degré quatre. *Inventiones mathematicae*, 110:585–602, 1992. URL http://dml.mathdoc.fr/item/GDZPPN002109484.

James Alexander and André Hirschowitz. Polynomial interpolation in several variables. *Journal of Algebraic Geometry*, 4(2):201–222, 1995. URL https://staff.math.su.se/shapiro/ProblemSolving/Hirschowitz.pdf.

James Alexander and André Hirschowitz. Generic hypersurface singularities. *Proceedings - Mathematical Sciences*, 107(2):139–154, May 1997. ISSN 0973-7685. doi: 10.1007/BF02837722.

James Alexander and André Hirschowitz. An asymptotic vanishing theorem for generic unions of multiple points. *Inventiones mathematicae*, 140(2):303–325, 2000. doi: 10.1007/s002220000057.

Raimundo Nonato Araújo Dos Santos, Aurélio Menegon Neto, David Mond, Marcelo J. Saia, and Jawad Snoussi (eds.). *Singularities and Foliations. Geometry, Topology and Applications: BMMS 2/NBMS 3, Salvador, Brazil, 2015*, volume 222 of *Springer Proceedings in Mathematics & Statistics*. Springer International Publishing, Cham, 2018. ISBN 978-3-319-73638-9 978-3-319-73639-6. doi: 10.1007/978-3-319-73639-6.

Federico Barbero, Cristian Bodnar, Michael Bronstein, Petar Velic˘kovic, and Pietro Lio. Sheaf Neural Networks with Connection Laplacians.

Michael M Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021.

Ferran Hernandez Caralt and Adrian Bazaga. ON THE NECESSITY OF LEARNABLE SHEAF LAPLA-. 2026.

Paula Leyes Carreno, Chiara Meroni, and Anna Seigal. Linear causal disentanglement via higher-order cumulants, July 2024. URL http://arxiv.org/abs/2407.04605. arXiv:2407.04605 [stat].

Phyllis J. Cassidy. Differential Algebraic Groups. *American Journal of Mathematics*, 94(3):891–954, 1972.

Karen A. Chandler. A brief proof of a maximal rank theorem for generic double points in projective space. *Transactions of the American Mathematical Society*, 353(5):1907–1920, December 2000. ISSN 0002-9947, 1088-6850. doi: 10.1090/S0002-9947-00-02732-X. URL http://www.ams.org/tran/2001-353-05/S0002-9947-00-02732-X/.

Luca Chiantini and Ciro Ciliberto. On the dimension of secant varieties. *Journal of the European Mathematical Society*, 12(5):1267–1291, August 2010. ISSN 1435-9855, 1435-9863. doi: 10.4171/jems/229. URL https://ems.press/doi/10.4171/jems/229.

Steven Dale Cutkosky. *Introduction to algebraic geometry*. Number volume 188 in Graduate studies in mathematics. American Mathematical Society, Providence, Rhode Island, 2018. ISBN 978-1-4704-3518-9.

Payam Dibaeinia, Sudarshan Babu, Mei Knudson, Ali ElSheikh, Yibo Wen, Han Liu, Jason Perera, and Aly A Khan. Position: Virtual Cells Need Context, Not Just Scale.

Xuguang Duan, Xin Wang, Ziwei Zhang, and Wenwu Zhu. Parametric Visual Program Induction with Function Modularization. In *International Conference on Machine Learning*, pp. 5643–5658. PMLR, June 2022. URL https://proceedings.mlr.press/v162/duan22c.html. ISSN: 2640-3498.

Ali ElSheikh, Rui-Xi Wang, Weimin Wu, Yibo Wen, Payam Dibaeinia, Jennifer Yuntong Zhang, Jerry Yao-Chieh Hu, Mei Knudson, Sudarshan Babu, Shao-Hua Sun, Aly A. Khan, and Han Liu. Cell-JEPA: Latent Representation Learning for Single-Cell Transcriptomics, February 2026. URL http://arxiv.org/abs/2602.02093. arXiv:2602.02093 [cs.CE].

Michael Felsberg, Fredrik Heintz, Fredrik Johansson, Danica Kragic, Fredrik Lindsten, and Amy Loutfi. What would be hilbert's problems of ai?

Kelsey Fu, Luna Lyu, Shuzhen Li, Suozhi Huang, Suheng Xu, Joy Zheng, Xuefeng Liu, Shilong Liu, Greg Barbone, Ying Liu, Linxi Jim Fan, Yuke Zhu, Matthew Davis, Kaiyi Jiang, Sherry Yang, Xu Kuang, Yuan Luo, Francesca Dominici, Christina Curtis, Xiaojie Qiu, Alberto Abadie, Caroline Uhler, Mary Tang, Joseph C. Wu, Ju Li, Jared Toettcher, José L. Avalos, Rebecca Rojansky, Huimin Zhao, John P. A. Ioannidis, Barry Rand, Emma Lundberg, Andrew S. Rosen, Zhuang Liu, Shana Kelley, Chenyan Xiong, Barbara E. Engelhardt, Thomas Montine, Christina V. Theodoris, Katherine S. Pollard, Fuxin Li, Xueyue Zhang, Tianyi Peng, Dmitri Basov, Xiang Qian, Eric Xing, Ali Yazdani, Jure Leskovec, Nigam Shah, Zhenan Bao, Pietro Perona, Sanfeng Wu, Clifford Brangwynne, Le Cong, and Mengdi Wang. Agentic Laboratories of the Future: Towards World Models for Scientific Discovery. August 2026. doi: 10.20944/preprints202608.0213.v1. URL https://www.preprints.org/manuscript/202608.0213.

Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

Huy Tai Ha and Paolo Mantero. The Alexander-Hirschowitz theorem and related problems, August 2021.

Jakob Hansen and Thomas Gebhart. Sheaf neural networks, 2020. URL https://arxiv.org/abs/2012.06333.

Robin Hartshorne. Varieties. In *Algebraic Geometry*, pp. 1–59. Springer, New York, NY, 1977. ISBN 978-1-4757-3849-0. doi: 10.1007/978-1-4757-3849-0_1. URL https://link.springer.com/chapter/10.1007/978-1-4757-3849-0_1.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Yu Huang, Zixin Wen, Aarti Singh, Yuejie Chi, and Yuxin Chen. Transformers Provably Learn Chain-of-Thought Reasoning with Length Generalization, November 2025. URL https://arxiv.org/abs/2511.07378v1.

Ilyes Khemakhem, Diederik Kingma, Ricardo Monti, and Aapo Hyvarinen. Variational autoencoders and nonlinear ica: A unifying framework. In *International conference on artificial intelligence and statistics*, pp. 2207–2217. PMLR, 2020.

Askold G. Khovanskii. Fewnomials and Pfaff Manifolds. *Selecta Mathematica*, 10(3):269–293, 1991.

Joe Kileel, Matthew Trager, and Joan Bruna. On the expressive power of deep polynomial neural networks. *Advances in neural information processing systems*, 32, 2019a.

Joe Kileel, Matthew Trager, and Joan Bruna. On the Expressive Power of Deep Polynomial Neural Networks. *Advances in Neural Information Processing Systems*, 32, 2019b. URL https://proceedings.neurips.cc/paper/2019/hash/a0dc078ca0d99b5ebb465a9f1cad54ba-Abstract.html.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Bohdan Kivva, Goutham Rajendran, Pradeep Ravikumar, and Bryon Aragam. Identifiability of deep generative models without auxiliary information. *Advances in Neural Information Processing Systems*, 35:15687–15701, 2022.

Kathlén Kohn, Thomas Merkh, Guido Montúfar, and Matthew Trager. Geometry of Linear Convolutional Networks, August 2021. URL https://arxiv.org/abs/2108.01538v2.

Kathlén Kohn, Guido Montúfar, Vahid Shahverdi, and Matthew Trager. Function Space and Critical Points of Linear Convolutional Networks, April 2023. URL https://arxiv.org/abs/2304.05752v2.

Kaie Kubjas, Jiayi Li, and Maximilian Wiesmann. Geometry of Polynomial Neural Networks, February 2024. URL https://arxiv.org/abs/2402.00949v2.

Zongxia Li, Xiyang Wu, Hongyang Du, Fuxiao Liu, Huy Nghiem, and Guangyao Shi. A Survey of State of the Art Large Vision Language Models: Alignment, Benchmark, Evaluations and Challenges, April 2025. URL http://arxiv.org/abs/2501.02189. arXiv:2501.02189 [cs].

Ziming Liu, Sophia Sanborn, Surya Ganguli, and Andreas Tolias. From Kepler to Newton: Inductive Biases Guide Learned World Models in Transformers, February 2026. URL https://arxiv.org/abs/2602.06923v1.

Francesco Locatello, Stefan Bauer, Mario Lucic, Gunnar Raetsch, Sylvain Gelly, Bernhard Schölkopf, and Olivier Bachem. Challenging common assumptions in the unsupervised learning of disentangled representations. In *international conference on machine learning*, pp. 4114–4124. PMLR, 2019.

Giovanni Luca Marchetti, Vahid Shahverdi, Stefano Mereta, Matthew Trager, and Kathlén Kohn. An invitation to neuroalgebraic geometry. *arXiv e-prints*, pp. arXiv–2501, 2025a.

Giovanni Luca Marchetti, Vahid Shahverdi, Stefano Mereta, Matthew Trager, and Kathlén Kohn. Algebra Unveils Deep Learning – An Invitation to Neuroalgebraic Geometry, January 2025b. URL https://arxiv.org/abs/2501.18915v2.

A. Massarenti and M. Mella. The Alexander-Hirschowitz theorem for neurovarieties, November 2025.

Sebastian Musslick, Laura K. Bartlett, Suyog H. Chandramouli, Marina Dubova, Fernand Gobet, Thomas L. Griffiths, Jessica Hullman, Ross D. King, J. Nathan Kutz, Christopher G. Lucas, Suhas Mahesh, Franco Pestilli, Sabina J. Sloman, and William R. Holmes. Automating the practice of science: Opportunities, challenges, and implications. *Proceedings of the National Academy of Sciences*, 122(5):e2401238121, February 2025. doi: 10.1073/pnas.2401238121. URL https://www.pnas.org/doi/abs/10.1073/pnas.2401238121.

Mufan Qiu, Genhui Zheng, Yinuo Xu, Ruichen Zhang, Ying Ding, Qi Long, and Tianlong Chen. Chreode: A Cell World Model for One-Step Temporal Dynamics and Perturbation Prediction, May 2026. URL http://arxiv.org/abs/2605.28111. arXiv:2605.28111 [cs.LG].

Adityanarayanan Radhakrishnan, Daniel Beaglehole, Parthe Pandit, and Mikhail Belkin. Mechanism for feature learning in neural networks and backpropagation-free machine learning models. *Science*, 383(6690):1461–1467, 2024. doi: 10.1126/science.adi5639. URL https://www.science.org/doi/abs/10.1126/science.adi5639.

Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *International conference on machine learning*, pp. 1530–1538. PMLR, 2015.

Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio. Toward causal representation learning. *Proceedings of the IEEE*, 109(5):612–634, 2021.

Vahid Shahverdi. Algebraic Complexity and Neurovariety of Linear Convolutional Networks, January 2024. URL https://arxiv.org/abs/2401.16613v1.

Vahid Shahverdi, Giovanni Luca Marchetti, Georg Bökman, and Kathlén Kohn. Identifiable Equivariant Networks are Layerwise Equivariant, January 2026a. URL http://arxiv.org/abs/2601.21645. arXiv:2601.21645 [cs].

Vahid Shahverdi, Giovanni Luca Marchetti, and Kathlén Kohn. LEARNING ON A RAZOR'S EDGE: IDENTIFIABILITY AND SINGULARITY OF POLYNOMIAL NEURAL NETWORKS. 2026b.

Le Song, Eran Segal, and Eric Xing. How to build an AI-driven digital organism. *Nature Medicine*, pp. 1–10, August 2026. ISSN 1546-170X. doi: 10.1038/s41591-026-04595-0. URL https://www.nature.com/articles/s41591-026-04595-0.

Konstantin Usevich, Ricardo Borsoi, Clara Dérand, and Marianne Clausel. Identifiability of Deep Polynomial Neural Networks, June 2025. URL https://arxiv.org/abs/2506.17093v2.

Keyon Vafa, Peter G. Chang, Ashesh Rambachan, and Sendhil Mullainathan. What Has a Foundation Model Found? Using Inductive Bias to Probe for World Models, December 2025. URL http://arxiv.org/abs/2507.06952. arXiv:2507.06952 [cs.LG].

Chloe Wang, Mehran Karimzadeh, Neal G Ravindra, Lexi R Bounds, Nader Alerasool, Ann C Huang, Shihao Ma, Daniel R Gulbranson, Haotian Cui, Yongju Lee, Anusuya Arjavalingam, Elliot J MacKrell, Matthew S Wilken, Jieming Chen, Benjamin W Herken, Jesse A Weber, Massimo M Onesto, Barbara Gonzalez-Teran, Nicole F Leung, Sally Yu Shi, Byron J Smith, Sharon K Lam, Adam Barner, Philip Wright, Elizabeth M Rumsey, Soohong Kim, Rene V Sit, Adam J Litterman, Ci Chu, and Bo Wang. X-Cell: Scaling Causal Perturbation Prediction Across Diverse Cellular Contexts via Diffusion Language Models.

Zhiting Wei, Yiheng Wang, Yicheng Gao, Shuguang Wang, Ping Li, Duanmiao Si, Yuli Gao, Siqi Wu, Danlu Li, Kejing Dong, Xingbo Yang, Chen Tang, Shaliu Fu, Xiaohan Chen, Wannian Li, Yuzhou You, Chen Zhang, Aibin Liang, Guohui Chuai, and Qi Liu. Benchmarking algorithms for generalizable single-cell perturbation response prediction. *Nature Methods*, pp. 1–14, December 2025. ISSN 1548-7105. doi: 10.1038/s41592-025-02980-0. URL https://www.nature.com/articles/s41592-025-02980-0. Publisher: Nature Publishing Group.

Eric Xing and Le Song. A World Model of the Virtual Cell.

Eric Xing, Mingkai Deng, and Jinyu Hou. Critique of world model, 2026. URL https://arxiv.org/abs/2507.05169.

Han Zhang, Guo-Hua Yuan, Chaohao Yuan, Tingyang Xu, Tian Bian, Hong Cheng, Wenbing Huang, Deli Zhao, and Yu Rong. Lingshu-Cell: A generative cellular world model for transcriptome modeling toward virtual cells, March 2026. URL http://arxiv.org/abs/2603.25240. arXiv:2603.25240 [q-bio.QM].

Kun Zhang, Shaoan Xie, Ignavier Ng, and Yujia Zheng. Causal Representation Learning from Multiple Distributions: A General Setting, February 2024. URL https://arxiv.org/abs/2402.05052v1.

A PRELIMINARIES FOR IDENTIFIABILITY OF NEURAL NETWORKS

A neural network (or an equivalent architecture) is said to be identifiable if the representation can be obtained by a set of general points representing the weight matrices up to permutation and scaling. In geometry, this immediately implies that, for a polynomial neural network, a set of weight matrices is equivalent up to a nonzero scalar multiple, and it also establishes a birational map between the parameter space and the closure of the image of the function space. For a class of feedforward PNN architecture $(\mathbf{d}, \mathbf{n})$, where $\mathbf{n} = (d_0, \cdots, d_L)$ the weight matrix size of each layer and $(r_1, \cdots, r_{L-1})$ the degree of activation function. $f_i(x) = W_i x + b_i$ the i-th layer network with the weight matrix $W_i \in \mathbb{C}^{d_i \times d_{i-1}}$ and the (potential) bias vector $b_i \in \mathbb{C}^{d_i}$. Each ρ_i is a component-wise activation function in the form of a monomial $\rho_l = x^{r_l}$ of a finite number of unknowns x. Geometrically, the identification of a deep fully polynomial neural network reduces to determining the smallest algebraic variety $\mathcal{V}_{d_0,d_1,\cdots,d_L,r}$, called the *neurovariety*, which is the Zariski closure of the ambient neuromanifold $\mathcal{M}_{d_0,d_1,\cdots,d_L,r}$ such that

$$\overline{\mathrm{Im}(\mathcal{M}_{d_0,d_1,\cdots,d_L,r})}^{\mathrm{Zar}} := \mathcal{V}_{d_0,d_1,\cdots,d_L,r} \supseteq \mathcal{M}_{d_0,d_1,\cdots,d_L,r},$$

succeeded by a motivating theorem and a conjecture. Together, these constitute a core theoretical framework for the study of Polynomial Neural Networks (PNNs).

Theorem A.1 (Theorem 11 of (Usevich et al., 2025)). *Let $((d_0, \cdots, d_L), (r_1, \cdots, r_{L-1}))$ be the homogeneous Polynomial Neural Network format. For $l = 0, \cdots, L-2$ denote $\tilde{d}_l := \min\{d_0, \cdots, d_l\}$. Then the following holds true: if for all $l = 1, \cdots, L-1$ the two-layer architecture $\mathrm{hPNN}_{(\tilde{d}_{l-1}, d_l, d_{l+1}), r_l[\cdot]}$ is finitely identifiable, then the L-layer architecture $\mathrm{hPNN}_{d, r_l[\cdot]}$ is finitely identifiable as well.*

Conjecture A.2 (Conjecture 40 of (Usevich et al., 2025)). *Under the assumptions of Theorem A.1, the L-layer hPNN is globally identifiable.*

For a shallow polynomial neural network with depth $L \leq 2$, Theorem A.2 asserts that generic identifiability holds, whereas failure may occur only on a measure-zero set of pathological parameter configurations. Consequently, under the Kruskal rank conditions, these results are generally constrained to finite-to-one identifiability, as evidenced by several immediate counterexamples. Recently, Massarenti and Mella (Massarenti & Mella, 2025) proved the global and generic identifiability via extending an important Alexander-Hirschowitz theorem to polynomial neurovarieties. The classic Alexander-Hirschowitz theorem resolves a major problem about secant varieties of the usual Veronese embeddings in $\mathbb{P}_n$. Their results apply to most cases of secant varieties using a refined, elementary inductive procedure.

A.1 EXAMPLES FOR LAYER-WISE NEURAL NETWORK COMPOSITION

Example A.3. Take an architecture $(\mathbf{d} = (2, 3, 3, 2), \mathbf{r} = (2, 3))$, we generate the final homogeneous polynomial functions via layer-wise composition:

- $F^{(1)} = W_1[x_0, x_1]$

$$\longrightarrow F^{(1)} = \begin{bmatrix} w_{11}^{(1)} x_0 + w_{12}^{(1)} x_1 \\ w_{21}^{(1)} x_0 + w_{22}^{(1)} x_1 \\ w_{31}^{(1)} x_0 + w_{32}^{(1)} x_1 \end{bmatrix}$$

- $F^{(2)} = W_2[(F_j^{(1)})^2], j = 0, 1, 2$

$$\longrightarrow F^{(2)} = \begin{bmatrix} w_{11}^{(2)}(w_{11}^{(1)} x_0 + w_{12}^{(1)} x_1)^2 + w_{12}^{(2)}(w_{21}^{(1)} x_0 + w_{22}^{(1)} x_1)^2 + w_{13}^{(2)}(w_{31}^{(1)} x_0 + w_{32}^{(1)} x_1)^2 \\ w_{21}^{(2)}(w_{11}^{(1)} x_0 + w_{12}^{(1)} x_1)^2 + w_{22}^{(2)}(w_{21}^{(1)} x_0 + w_{22}^{(1)} x_1)^2 + w_{23}^{(2)}(w_{31}^{(1)} x_0 + w_{32}^{(1)} x_1)^2 \\ w_{31}^{(2)}(w_{11}^{(1)} x_0 + w_{12}^{(1)} x_1)^2 + w_{32}^{(2)}(w_{21}^{(1)} x_0 + w_{22}^{(1)} x_1)^2 + w_{33}^{(2)}(w_{31}^{(1)} x_0 + w_{32}^{(1)} x_1)^2 \end{bmatrix}$$

- $F^{(3)} = W_3[(F_j^{(2)})^3], j = 0, 1, 2$

$$\longrightarrow F^{(3)} = \begin{bmatrix} w_{11}^{(3)}\left[\sum_{k=1}^3 w_{1k}^{(2)}(w_{k1}^{(1)}x_0 + w_{k2}^{(1)}x_1)^2\right]^3 + w_{12}^{(3)}\left[\sum_{k=1}^3 w_{2k}^{(2)}(w_{k1}^{(1)}x_0 + w_{k2}^{(1)}x_1)^2\right]^3 \\ +w_{13}^{(3)}\left[\sum_{k=1}^3 w_{3k}^{(2)}(w_{k1}^{(1)}x_0 + w_{k2}^{(1)}x_1)^2\right]^3 \\ w_{21}^{(3)}\left[\sum_{k=1}^3 w_{1k}^{(2)}(w_{k1}^{(1)}x_0 + w_{k2}^{(1)}x_1)^2\right]^3 + w_{22}^{(3)}\left[\sum_{k=1}^3 w_{2k}^{(2)}(w_{k1}^{(1)}x_0 + w_{k2}^{(1)}x_1)^2\right]^3 \\ +w_{23}^{(3)}\left[\sum_{k=1}^3 w_{3k}^{(2)}(w_{k1}^{(1)}x_0 + w_{k2}^{(1)}x_1)^2\right]^3 \end{bmatrix}$$

By this construction, we can show the effects on the Veronese levels by fixing the level-wise dual-basis monomials. Since the linear transformation at level $l = 1$ is trivial, we start with level $l = 2$. Viewing each entry of S_l as a dual form, the de-dualization corresponds to the extraction of the base vectors $(W_{k:}^{(1)})$ from the intersection of the polar hyperspaces. Geometrically, this is equivalent to recovering the base points of the linear system defined by H^0.

$$S_2 = \begin{bmatrix} w_{11}^{(2)}(W_{1:}^{(1)})^{\otimes 2} + w_{12}^{(2)}(W_{2:}^{(1)})^{\otimes 2} + w_{13}^{(2)}(W_{3:}^{(1)})^{\otimes 2} \\ w_{21}^{(2)}(W_{1:}^{(1)})^{\otimes 2} + w_{22}^{(2)}(W_{2:}^{(1)})^{\otimes 2} + w_{23}^{(2)}(W_{3:}^{(1)})^{\otimes 2} \\ w_{31}^{(2)}(W_{1:}^{(1)})^{\otimes 2} + w_{32}^{(2)}(W_{2:}^{(1)})^{\otimes 2} + w_{33}^{(2)}(W_{3:}^{(1)})^{\otimes 2} \end{bmatrix} \tag{A.1}$$

Level 3 becomes more complex because the dual basis S_l of fixed level l is not a global Veronese embedding for a single vector. As a consequence, the symmetric tensor algebra embeds the preceding polynomials into the Veronese variety $v_2(\mathbb{P}^1) \subset \mathbb{P}^2$. Let $\mathbf{y} = [x_0^2, x_0x_1, x_1^2]^T$ be the dual basis. The coefficients vector $\mathbf{a}_k$ represents the coordinates of $F_k^{(2)}$ in this lifted projective space.

For $w_{jk}, j = 1$, the coefficients corresponding to each dual basis are

$$\begin{aligned} x_0^2 &: w_{11}^{(2)}(w_{11}^{(1)})^2 + w_{12}^{(2)}(w_{21}^{(1)})^2 + w_{13}^{(2)}(w_{31}^{(1)})^2 \Rightarrow a_0 \\ x_0x_1 &: 2(w_{11}^{(2)}w_{11}^{(1)}w_{12}^{(1)} + w_{12}^{(2)}w_{21}^{(1)}w_{22}^{(1)} + w_{13}^{(2)}w_{31}^{(1)}w_{32}^{(1)}) \Rightarrow a_1 \\ x_1^2 &: w_{11}^{(2)}(w_{12}^{(1)})^2 + w_{12}^{(2)}(w_{22}^{(1)})^2 + w_{13}^{(2)}(w_{32}^{(1)})^2 \Rightarrow a_2 \end{aligned}$$

Collecting all coefficients in a vector $\mathbf{a}_k := [a_0, a_1, a_2]_k^T$, and then taking the dual to the polynomial space yields

$$F^{(3)} = \begin{bmatrix} w_{11}^{(3)}\mathbf{a}_1^{\otimes 3} + w_{12}^{(3)}\mathbf{a}_2^{\otimes 3} + w_{13}^{(3)}\mathbf{a}_3^{\otimes 3} \\ w_{21}^{(3)}\mathbf{a}_1^{\otimes 3} + w_{22}^{(3)}\mathbf{a}_2^{\otimes 3} + w_{23}^{(3)}\mathbf{a}_3^{\otimes 3} \end{bmatrix} \tag{A.2}$$

B Related Work

Neuroalgebraic geometry. As a merging field, neuroalgebraic geometry exhibits strong potential because it guides new directions for various fields of machine learning from the novel perspective of the classic subject *Algebraic Geometry*. Albeit a full survey is beyond our current scope, we outline the most salient areas and their relations below. For a comprehensive treatment on these topics, we recommend interested readers consult the recent reviews in (Marchetti et al., 2025a;b). Classic algebraic geometry is connected to algebraic functions and the underlying geometry, which are used to study subtler structures (Hartshorne, 1977; Cutkosky, 2018). While neural architectures share a foundational, similar nature with algebraic functions, they frequently incorporate transcendental activation functions that extend beyond polynomial regimes. More importantly, the recursive, layer-wise transformation characteristic of neural networks introduces structural complexities that preclude the direct application of classical algebraic geometry. This necessitates a more specialized framework and sparks the study of neurovarieties—to navigate the evolving geometry of latent representations. Owing to their construction from polynomial building blocks, Polynomial Neural Networks (PNNs) provided the initial foothold for algebraic geometers to explore the landscapes of deep learning. As the first architectures to fall naturally within the purview of classical algebraic geometry, PNNs catalyzed the emergence of *Neuroalgebraic Geometry* as a burgeoning field (Marchetti et al., 2025b;a).

As an independent interest to the community of computer science, training deep neural networks aims to, through optimization methods, learn a representation that reflects the data-generating process from the input data and leverage the obtained representation for downstream tasks. PNN, a powerful, highly expressive counterpart, reduces the problem by finding the best possible representation over algebraic or semi-algebraic sets defined on certain hypersurfaces. The identifiability of such a network determines the expressivity of neural networks. It hinges on the expected dimension of the vanishing locus of corresponding hypersurfaces (Kileel et al., 2019b; Kubjas et al., 2024), all of which can be examined in algebro-geometric languages. One representative area is the study of learnability and identifiability of neural models. The earliest work to date discussing the algebraic structure of polynomial neural networks is (Kileel et al., 2019b), which introduced the neurovariety of polynomial neural networks and their dimensions as one important geometric invariant. Subsequent work focused on evaluating the plausible pair of dimensions and degrees, and threshold at which expressivity and identifiability can be attained (Kubjas et al., 2024; Shahverdi, 2024; Kohn et al., 2021; 2023; Usevich et al., 2025; Massarenti & Mella, 2025). Another popular direction is the study of the algebraic structure and geometry of the loss landscape of optimization for neural networks. For example, studying the geometry of critical points of the function space helps researchers understand the training dynamics of deep learning models.

There are also related works that integrate cellular sheaf theory into machine learning, for instance, Sheaf Laplacian (SL) (Barbero et al.; Caralt & Bazaga, 2026) and Sheaf Neural Network (SNN) (Hansen & Gebhart, 2020). Those developments aim to address over-smoothing in training neural networks, to present a generalization of convolutional neural networks, and thus remain largely orthogonal to other lines of research in geometry.

Deep generative modeling. Deep generative models learn a map from latent variables to observed data and have been instantiated through variational objectives, adversarial training, invertible transformations, and diffusion-based denoising (Kingma & Welling, 2013; Goodfellow et al., 2014; Rezende & Mohamed, 2015; Ho et al., 2020). While these models are often evaluated by likelihood, reconstruction fidelity, or sample quality, the uniqueness of the learned latent mechanism is substantially more delicate: different decoders, priors, or latent coordinate systems can induce the same observational distribution, making the recovered representation non-identifiable without additional structure. This issue has motivated a growing line of work on identifiable deep latent-variable models, including impossibility results for fully unsupervised disentanglement (Locatello et al., 2019), auxiliary-variable and nonlinear-ICA conditions for identifiable variational autoencoders (Khemakhem et al., 2020), and recent identifiability guarantees for broad classes of deep generative models under architectural or prior restrictions (Kivva et al., 2022). *Our work departs from these statistical or architectural sufficient conditions by treating the neural generator itself as a geometric object.* Building on the algebraic view of polynomial neural networks as neurovarieties (Kileel et al., 2019a), we characterize when the generative map admits a unique parameterization through double-point schemes and cohomology vanishing. This perspective identifies non-identifiability not merely as a failure of supervision, optimization, or tensor decomposition, but as geometric defectiveness of the associated neurovariety. Consequently, our results provide a principled bridge from deep generative modeling to identifiable representation learning: they give identifiability guarantees in non-generic regimes and under non-injective transformations, where standard latent-variable arguments do not directly apply.

Causal world models and AI for Science Scientific discovery via artificial intelligence and machine learning requires informative priors, inductive biases, or human supervision to ensure that the knowledge obtained remains verifiable (Fu et al., 2026; Musslick et al., 2025). Current paradigms in Representation Learning and world models largely either hinge on restrictive assumptions—such as injectivity via a diffeomorphism—to guarantee identifiability with presumed existing or viable interventions and variability (Zhang et al., 2024; Ahuja et al., 2023), or rely on specific external, predefined physical laws such as spatial smoothness and stability to avoid learning nonsensical representations (Vafa et al., 2025; Qiu et al., 2026). However, fundamental scientific laws governing phenomena from thermodynamic collapse to symmetry breaking in physics and biology are inherently non-injective, and operate on a more generic hypothetical space beyond human experience and definition (250). Furthermore, the requisite inductive biases for such laws remain elusive and often demand independent theoretical development, rendering existing frameworks inapplicable to a vast range of natural phenomena.

C LEADING EXAMPLE

Example A (Linear Causal Disentanglement via Cumulant Tensors)**.** Consider the problem of recovering a mixing matrix $A \in \mathbb{R}^{n\times k}$ from the observed variables $X = Az$, where $z = (z_1, \ldots, z_k)^\top$ is an independent latent vector.

In statistics, the d-th order cumulant of X is a symmetric tensor $\mathcal{C}_d(X) \in (\mathbb{R}^n)^{\otimes d}$. Under the assumption of independence, this tensor decomposes as:

$$\mathcal{C}_d(X) = \sum_{j=1}^{k} \kappa_d(z_j) \cdot a_j \otimes a_j \otimes \cdots \otimes a_j,$$

where a_j are the columns of A and $\kappa_d(z_j)$ are the scalar cumulants of the latent factors.

Algebraic Mapping: From our perspective, this cumulant tensor $\mathcal{C}_d(X)$ is an element of the **symmetric power** $S^d(\mathbb{R}^n)$. By the duality between symmetric tensors and polynomials, $\mathcal{C}_d(X)$ corresponds to a homogeneous polynomial P_X of degree d in the dual space $(\mathbb{R}^n)^*$. Specifically, if we define variables $y = (y_1, \ldots, y_n)$, the tensor maps to the polynomial:

$$\begin{aligned}
\langle a_j^{\otimes d}, y^{\otimes d}\rangle &= \underbrace{\mathcal{C}(a_j \otimes a_j \otimes \cdots \otimes a_j,\, y \otimes y \otimes \cdots \otimes y)}_{\text{Full contraction operator } \mathcal{C}} \\
&= \sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\cdots\sum_{i_d=1}^{n}\left(\prod_{m=1}^{d}(a_j)_{i_m}\right)\cdot\left(\prod_{m=1}^{d} y_{i_m}\right) \\
&= \sum_{i_1,\ldots,i_d=1}^{n}(a_{i_1 j}y_{i_1})(a_{i_2 j}y_{i_2})\ldots(a_{i_d j}y_{i_d}) \\
&= \left(\sum_{i=1}^{n} a_{ij}y_i\right)^d = (a_j^\top y)^d.
\end{aligned}$$

The task of identifying A (identifiability) is thus equivalent to determining whether the polynomial P_X has a *unique* decomposition into a sum of d-th powers of linear forms. The contraction $\langle a_j^{\otimes d}, y^{\otimes d}\rangle$ is the evaluation of the symmetric tensor $a_j^{\otimes d}$ at y ^{otimesd} when viewed as a multilinear form on the dual space. This operator $\mathcal{C}$ effectively collapses the d-dimensional grid of tensor entries into a single scalar value.

Example C.1. Take an architecture $(\mathbf{d} = (2, 3, 3, 2), \mathbf{r} = (2, 3))$, we generate the final homogeneous polynomial functions via layer-wise composition:

- $F^{(1)} = W_1[x_0, x_1]$
  $\longrightarrow F^{(1)} = \begin{bmatrix} w_{11}^{(1)}x_0 + w_{12}^{(1)}x_1 \\ w_{21}^{(1)}x_0 + w_{22}^{(1)}x_1 \\ w_{31}^{(1)}x_0 + w_{32}^{(1)}x_1 \end{bmatrix}$

- $F^{(2)} = W_2[(F_j^{(1)})^2], j = 0, 1, 2$
  $\longrightarrow F^{(2)} = \begin{bmatrix} w_{11}^{(2)}(w_{11}^{(1)}x_0 + w_{12}^{(1)}x_1)^2 + w_{12}^{(2)}(w_{21}^{(1)}x_0 + w_{22}^{(1)}x_1)^2 + w_{13}^{(2)}(w_{31}^{(1)}x_0 + w_{32}^{(1)}x_1)^2 \\ w_{21}^{(2)}(w_{11}^{(1)}x_0 + w_{12}^{(1)}x_1)^2 + w_{22}^{(2)}(w_{21}^{(1)}x_0 + w_{22}^{(1)}x_1)^2 + w_{23}^{(2)}(w_{31}^{(1)}x_0 + w_{32}^{(1)}x_1)^2 \\ w_{31}^{(2)}(w_{11}^{(1)}x_0 + w_{12}^{(1)}x_1)^2 + w_{32}^{(2)}(w_{21}^{(1)}x_0 + w_{22}^{(1)}x_1)^2 + w_{33}^{(2)}(w_{31}^{(1)}x_0 + w_{32}^{(1)}x_1)^2 \end{bmatrix}$

- $F^{(3)} = W_3[(F_j^{(2)})^3], j = 0, 1, 2$

$$\longrightarrow F^{(3)} = \begin{bmatrix} w_{11}^{(3)}\left[\sum_{k=1}^3 w_{1k}^{(2)}(w_{k1}^{(1)}x_0 + w_{k2}^{(1)}x_1)^2\right]^3 + w_{12}^{(3)}\left[\sum_{k=1}^3 w_{2k}^{(2)}(w_{k1}^{(1)}x_0 + w_{k2}^{(1)}x_1)^2\right]^3 \\ +w_{13}^{(3)}\left[\sum_{k=1}^3 w_{3k}^{(2)}(w_{k1}^{(1)}x_0 + w_{k2}^{(1)}x_1)^2\right]^3 \\ w_{21}^{(3)}\left[\sum_{k=1}^3 w_{1k}^{(2)}(w_{k1}^{(1)}x_0 + w_{k2}^{(1)}x_1)^2\right]^3 + w_{22}^{(3)}\left[\sum_{k=1}^3 w_{2k}^{(2)}(w_{k1}^{(1)}x_0 + w_{k2}^{(1)}x_1)^2\right]^3 \\ +w_{23}^{(3)}\left[\sum_{k=1}^3 w_{3k}^{(2)}(w_{k1}^{(1)}x_0 + w_{k2}^{(1)}x_1)^2\right]^3 \end{bmatrix}$$

By this construction, we can show the effects on the Veronese levels by fixing the level-wise dual-basis monomials. Since the linear transformation at level $l = 1$ is trivial, we start with level $l = 2$. Viewing each entry of S_l as a dual form, the de-dualization corresponds to the extraction of the base vectors $(W_{k:}^{(1)})$ from the intersection of the polar hyperspaces. Geometrically, this is equivalent to recovering the base points of the linear system defined by H^0.

$$S_2 = \begin{bmatrix} w_{11}^{(2)}(W_{1:}^{(1)})^{\otimes 2} + w_{12}^{(2)}(W_{2:}^{(1)})^{\otimes 2} + w_{13}^{(2)}(W_{3:}^{(1)})^{\otimes 2} \\ w_{21}^{(2)}(W_{1:}^{(1)})^{\otimes 2} + w_{22}^{(2)}(W_{2:}^{(1)})^{\otimes 2} + w_{23}^{(2)}(W_{3:}^{(1)})^{\otimes 2} \\ w_{31}^{(2)}(W_{1:}^{(1)})^{\otimes 2} + w_{32}^{(2)}(W_{2:}^{(1)})^{\otimes 2} + w_{33}^{(2)}(W_{3:}^{(1)})^{\otimes 2} \end{bmatrix} \tag{C.1}$$

Level 3 becomes more complex because the dual basis S_l of fixed level l is not a global Veronese embedding for a single vector. As a consequence, the symmetric tensor algebra embeds the preceding polynomials into the Veronese variety $v_2(\mathbb{P}^1) \subset \mathbb{P}^2$. Let $\mathbf{y} = [x_0^2, x_0x_1, x_1^2]^T$ be the dual basis. The coefficients vector $\mathbf{a}_k$ represents the coordinates of $F_k^{(2)}$ in this lifted projective space.

For $w_{jk}, j = 1$, the coefficients corresponding to each dual basis are

$$\begin{aligned} x_0^2 &: w_{11}^{(2)}(w_{11}^{(1)})^2 + w_{12}^{(2)}(w_{21}^{(1)})^2 + w_{13}^{(2)}(w_{31}^{(1)})^2 \Rightarrow a_0 \\ x_0x_1 &: 2(w_{11}^{(2)}w_{11}^{(1)}w_{12}^{(1)} + w_{12}^{(2)}w_{21}^{(1)}w_{22}^{(1)} + w_{13}^{(2)}w_{31}^{(1)}w_{32}^{(1)}) \Rightarrow a_1 \\ x_1^2 &: w_{11}^{(2)}(w_{12}^{(1)})^2 + w_{12}^{(2)}(w_{22}^{(1)})^2 + w_{13}^{(2)}(w_{32}^{(1)})^2 \Rightarrow a_2 \end{aligned}$$

Collecting all coefficients in a vector $\mathbf{a}_k := [a_0, a_1, a_2]_k^T$, and then taking the dual to the polynomial space yields

$$F^{(3)} = \begin{bmatrix} w_{11}^{(3)}\mathbf{a}_1^{\otimes 3} + w_{12}^{(3)}\mathbf{a}_2^{\otimes 3} + w_{13}^{(3)}\mathbf{a}_3^{\otimes 3} \\ w_{21}^{(3)}\mathbf{a}_1^{\otimes 3} + w_{22}^{(3)}\mathbf{a}_2^{\otimes 3} + w_{23}^{(3)}\mathbf{a}_3^{\otimes 3} \end{bmatrix} \tag{C.2}$$

D Proof of Identifiability for Neurovarieties

D.1 Proof of Theorem 3.3

Our primary objective is to identify a non-trivial homogeneous polynomial $F(\mathbf{x}) = \sum_{i=1}^k L_i^d$ within the relevant functional space $S_d = S^dV^\vee$. Under the *Veronese embedding* $\nu_d : \mathbb{P}^n \hookrightarrow \mathbb{P}^N$, each linear form L_i (Ha & Mantero, 2021, Lemma A.12) corresponds to a point on the Veronese variety through its dual basis representation $[y_0 : y_1 : \cdots : y_N]$, where $N = \binom{n+d}{d} - 1$. Consequently, the task of identifying the target polynomial reduces to determining the coefficients of each linear form. For example, a homogeneous polynomial in two unknown x, y of degree 2 , we prescribe a set of general points $\mathcal{P}_{generic} = \{P_1, \ldots, P_k\}$ and employ the *evaluation operator* ev_P to constrain the polynomial:

$$\sum_{j=0}^{2} a_j y_j(\mathcal{P}_{generic}) = 0 \tag{D.1}$$

where $\{y_j\} := [x^2 : xy : y^2]$ denotes the projective basis of dual polynomial functions, and evaluates any general point by preserving its j-th component via $\langle \mathcal{P}_{generic}, y_j \rangle$. As the selected points are generic, the resulting linear system provides the global section space $\Gamma(\mathbb{P}^n, \mathcal{O}(d))$ restricted by the generic point $\mathcal{P}_{generic}$. In this case, F can be uniquely determined by imposing independent conditions on N coefficients of the dual polynomial basis. However, for a rather arbitrary polynomial in n unknowns of degree d, specifying such a polynomial is intrinsically difficult since the

dual basis is on the twisted Veronese curves, and manually imposing k independent conditions is not computationally efficient. This is the famous problem in algebraic geometry termed as *polynomial extrapolation*.

The space of homogeneous polynomials of degree r is denoted by $\mathcal{O}_{\mathbb{P}^n}(r)$ and thus the tensor product $\mathcal{I}_X^2 \otimes \mathcal{O}_{\mathbb{P}^n}(r) := \mathcal{I}_X^2(r)$ represents all polynomials singular at generic points X. The canonical functor

$$H^i(X, \mathcal{F}) = H^i(\Gamma(X, \mathcal{F}^\bullet)) \tag{D.2}$$

is the i-th cohomology group of a abelian sheaf $\mathcal{F}$ and it induces family of universal δ-functor $Ab(X) \to Ab$. Geometrically, it measures the extent to which one can extend local properties to global ones. There exists a set of *restrictions* between two module homomorphisms, and it induces an exact sequence as shown in Lemma D.1.

Lemma D.1. *Given the structure sheaf $\mathcal{O}_{\mathbb{P}^n}(r)$ and the ideal sheaf $\mathcal{I}_X$ of a general collection of points $X \subset \mathbb{P}^n$, the following sequence is exact:*

$$0 \longrightarrow \mathcal{I}_X^2 \otimes \mathcal{O}_{\mathbb{P}^n}(r) \longrightarrow \mathcal{O}_{\mathbb{P}^n}(r) \longrightarrow \mathcal{O}_X^{(2)} \otimes \mathcal{O}_{\mathbb{P}^n}(r) \longrightarrow 0$$

Proof. Consider the defining short exact sequence of the ideal sheaf $\mathcal{I}_X^2$:

$$0 \longrightarrow \mathcal{I}_X^2 \longrightarrow \mathcal{O}_{\mathbb{P}^n} \longrightarrow \mathcal{O}_X^{(2)} \longrightarrow 0$$

Since $\mathcal{O}_{\mathbb{P}^n}(r)$ is an invertible sheaf (a line bundle), it is locally free and thus flat as an $\mathcal{O}_{\mathbb{P}^n}$-module. Since the tensor product with a flat module is an exact functor, the exactness of the sequence remains preserved under the twist by $\mathcal{O}_{\mathbb{P}^n}(r)$. □

For $n+1$ unknowns with degree r, the projective dimension of dual basis is $N = \binom{n+r}{r} - 1$. Thus, the restriction exact sequence indicates that the double points in X impose $k(n+1)$ conditions on the structure sheaf given by the polynomial function in N coordinates. At present, we construct the local behavior of candidate polynomials that are singular at these points. By gluing those sheaves, we get a long sequence

$$\begin{tikzcd} 0 \arrow[r] & H^0(\mathcal{I}_X^2 \otimes \mathcal{O}_{\mathbb{P}^n}(r)) \arrow[r] & H^0(\mathcal{O}_{\mathbb{P}^n}(r)) \arrow[r] & H^0(\mathcal{O}_X^{(2)} \otimes \mathcal{O}_{\mathbb{P}^n}(r)) \arrow[dll, "\delta"] \\ & H^1(\mathcal{I}_X^2 \otimes \mathcal{O}_{\mathbb{P}^n}(r)) \arrow[r] & \dots & \end{tikzcd} \tag{D.3}$$

derived by the connecting functor δ. Therefore, we denote the sequence by M, then Lemma D.2 gives the following dimension equality.

Lemma D.2. *A long sequence M satisfies the alternating sum equation*

$$\dim H^0(\mathcal{O}_{\mathbb{P}^n}(r)) - \dim H^0(\mathcal{I}_X^2(r)) = \dim H^0(\mathcal{O}_X^{(2)}(r)) - \dim H^1(\mathcal{I}_X^2(r))$$

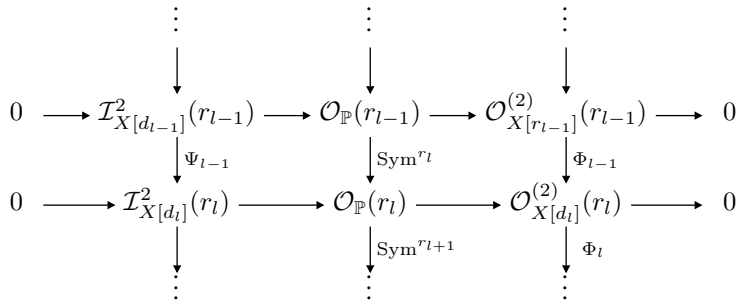
The first cohomology group $H^1(\mathcal{I}_X^2(r))$ measures the failure of the global evaluation map to be surjective. Specifically, it captures the constraints in $H^0(\mathcal{O}_X^{(2)}(r))$ that cannot be satisfied by any global polynomial in $H^0(\mathcal{O}_{\mathbb{P}^n}(r))$. Algebraically, it is identified with the cokernel of the morphism ϕ:

$$\begin{tikzcd} H^0(\mathcal{O}_{\mathbb{P}^n}(r)) \arrow[r, "\phi"] & H^0(\mathcal{O}_X^{(2)} \otimes \mathcal{O}_{\mathbb{P}^n}(r)) \arrow[r, two heads] & H^1(\mathcal{I}_X^2 \otimes \mathcal{O}_{\mathbb{P}^n}(r)) \arrow[d, "\cong"] \\ & & \mathrm{Coker}(\phi) \end{tikzcd}$$

Consequently, a non-vanishing H^1 indicates that the variety defined by the vector space of the global sections is superabundant, meaning the constraints imposed by the double points X are linearly dependent.

The final step is to leverage Lemma D.2 and Lemma D.1 to construct a diagram that, under our conditions, is commutative and consistent over square modules. We present the formal version of Proposition 3.2 with structured diagrams that represent modules.

Corollary D.3 (Morphisms between Layer-wise Ideal Sheaf)**.** *The layer-wise propagation defines a sequence of morphisms between ideal sheaves $\mathcal{I}^2_{X[d_l]}$, given by a generic collection of d_{l-1} double points. The diagram is commutative:*



Proof. To establish the commutativity of the diagram, we must verify that the composition of morphisms in each rectangular block is consistent. Let r_l denote the homogeneous degree of the l-th layer, and let $\mathbb{P}$ be the corresponding projective space. The vertical morphisms are induced by the layer-wise polynomial mapping $\mathcal{F}_l$, which corresponds to a Veronese embedding of degree r_l.

1. Commutativity of the Left Square (Ψ_{l-1} and Sym^{r_l}):

By definition, the ideal sheaf $\mathcal{I}^2_{X[d_{l-1}]}(r_{l-1})$ is a subsheaf of the structure sheaf $\mathcal{O}_{\mathbb{P}}(r_{l-1})$. The vertical map $\mathrm{Sym}^{r_l} : \mathcal{O}_{\mathbb{P}}(r_{l-1}) \to \mathcal{O}_{\mathbb{P}}(r_l)$ is a ring homomorphism induced by the l-th layer's activation and weight functions. Since the subsequent ideal sheaf $\mathcal{I}^2_{X[d_l]}$ is defined as the sheaf of functions vanishing (to second order) on the image of the previous layer under $\mathcal{F}_l$, it follows that for any local section $s \in \mathcal{I}^2_{X[d_{l-1}]}$, its image under Sym^{r_l} must vanish on the scheme $X[d_l]$. Thus, $\mathrm{Sym}^{r_l}(s) \in \mathcal{I}^2_{X[d_l]}$. This inclusion ensures the existence of the morphism Ψ_{l-1} such that:

$$\iota_l \circ \Psi_{l-1} = \mathrm{Sym}^{r_l} \circ \iota_{l-1}$$

where ι is the natural inclusion $\mathcal{I}^2 \hookrightarrow \mathcal{O}_{\mathbb{P}}$.

2. Commutativity of the Right Square (Φ_{l-1} and Sym^{r_l}):

Consider the quotient sheaves $\mathcal{O}^{(2)}_X$ representing the structure of the singular scheme. The horizontal morphisms are the natural projections $p : \mathcal{O}_{\mathbb{P}} \to \mathcal{O}_{\mathbb{P}}/\mathcal{I}^2$. Since Sym^{r_l} is an algebraic homomorphism, it induces a unique morphism between the quotient spaces Φ_{l-1} provided that $\mathrm{Sym}^{r_l}(\mathcal{I}^2_{X[d_{l-1}]}) \subseteq \mathcal{I}^2_{X[d_l]}$. Having proved this inclusion in the previous step, the universal property of quotient modules guarantees the existence of Φ_{l-1} such that:

$$\Phi_{l-1} \circ p_{l-1} = p_l \circ \mathrm{Sym}^{r_l}$$

Specifically, for any $f \in \mathcal{O}_{\mathbb{P}}(r_{l-1})$, the mapping follows $\Phi_{l-1}(f + \mathcal{I}^2_{l-1}) = \mathrm{Sym}^{r_l}(f) + \mathcal{I}^2_l$. □

E Proof of Vanishing Theorem for Causality

E.1 Proof of Theorem 4.4

Module extensions lead to a system

$$\mathcal{S}_{\mathrm{Ext}} : \sum_{i=1}^{q} \kappa_i A_{i:}^{\otimes d} = \sum_{i=1}^{q} \kappa_i \left(\sum_{k=1}^{p} f_{ik} (I - \Lambda)^{-1}_{k:} \right)^{\otimes d}$$

Choosing a dual basis $y^{\otimes d}$ and applying contraction yields

$$\mathcal{S}^{\vee}_{\mathrm{Ext}} : \sum_{i=1}^{q} \kappa_i \langle A_{i:}^{\otimes d}, y^{\otimes d} \rangle = \sum_{i=1}^{q} \kappa_i \langle \left(\sum_{k=1}^{p} f_{ik} (I - \Lambda)^{-1}_{k:} \right)^{\otimes d}, y^{\otimes d} \rangle \tag{E.1}$$

where the defining ideal of ground-truth parameters is

$$J = \left\langle a_{ij} - \sum_{k=1}^{p} a_{ik}\lambda_{kj} - f_{ij} \right\rangle_{1\le i\le p, 1\le j\le q}.$$

The dual system $\mathcal{S}_{\text{Ext}}^{\vee}$ constitutes a structure sheaf $\mathcal{O}_{\mathcal{S}_{\text{Ext}}}$, allowing us to impose double points $\mathcal{I}_X^2$ onto the hypersurfaces defined by the sheaf. By viewing the tensor contraction as a single coefficient, the vanishing theorem guarantees the imposition of $q(p+1)$ independent conditions per double point. Given that a single module extension $\mathcal{S}_{\text{Ext}}^{\vee}$ possesses $(p \times q) + (q \times q)$ generic coordinates, identifying the full system requires additional constraints. To verify that the subscheme $\mathcal{I}_X^2$ consistently imposes independent conditions across an augmented system, we leverage the distributivity of the cohomology functor over the direct sum of module extensions. For any two sheaves of $\mathcal{O}$-modules $\mathcal{F}$ and $\mathcal{G}$, we have the natural isomorphism:

$$H^q(\mathcal{F} \oplus \mathcal{G}) \cong H^q(\mathcal{F}) \oplus H^q(\mathcal{G})$$

Furthermore, the tensor product distributes over the direct sum:

$$V \otimes (W \oplus U) \cong (V \otimes W) \oplus (V \otimes U)$$

Consequently, by accumulating double points across multiple extended modules (representing varied interventions or higher-order moments), the global section space satisfies:

$$H^0\left(\mathcal{I}^2 \otimes \bigoplus_k \mathcal{S}_{\text{Ext}}^{(k)\vee}\right) \cong \bigoplus_k H^0\left(\mathcal{I}^2 \otimes \mathcal{S}_{\text{Ext}}^{(k)\vee}\right) \tag{E.2}$$

Consequently, we can finish the proof for the major evaluation morphism by constructing a long sequence:

Lemma E.1. *Given a linear causal disentanglement problem $X = FZ$, where the mixing matrix $F \in \mathbb{R}^{p\times q}$ is generic but unnecessarily injective, and the variable Z follows a causal graph with parameters $Z = (I-\Lambda)^{-1}x$ with $\Lambda \in \mathbb{R}^{q\times q}$ and $x \in \mathbb{R}^q$ a zero-mean noise variable. If the $\operatorname{Spec}(Z)$ and $\operatorname{Spec}(x)$ are not Artinian, then the LSD identifiability coincides with the identifiability of its modules of the related algebraic extension system $\mathcal{S}_{\text{Ext}}$ over the ring K_n:*

$$\begin{array}{l} 0 \longrightarrow H^0\left(\mathcal{I}^2 \otimes \bigoplus_k \mathcal{S}_{\text{Ext}}^{(k)\vee}\right) \longrightarrow H^0(\bigoplus_k \mathcal{O}_{\mathcal{S}_{\text{Ext}}}) \xrightarrow{\bigoplus_k \phi} H^0(\bigoplus_k \mathcal{O}_X^{(2)} \otimes \mathcal{O}_{\mathcal{S}_{\text{Ext}}}) \\ \qquad\qquad \xrightarrow{\bigoplus_k \delta} \\ \qquad H^1\left(\mathcal{I}^2 \otimes \bigoplus_k \mathcal{S}_{\text{Ext}}^{(k)\vee}\right) \longrightarrow \cdots \end{array} \tag{E.3}$$

where $\mathcal{I}_X^2$ is a new *zero-dimensional subscheme of k double points and $\mathcal{O}_{\mathcal{S}_{\text{Ext}}}$ is a structure sheaf on the dual function space to $\mathcal{S}_{\text{Ext}}$.*

The aggregate evaluation map $\bigoplus_k \phi^{(k)} = \phi_1 \oplus \phi_2 \oplus \cdots \oplus \phi_k$ provides independent constraints from each double-point subscheme. As the number of independent conditions increases, the rank of the joint morphism $\bigoplus_k \phi^{(k)}$ rises until the first cohomology group $H^1\left(\mathcal{I}_X^2 \otimes \bigoplus_k \mathcal{S}_{\text{Ext}}^{(k)\vee}\right)$ vanishes. By the exactness of the sequence, the vanishing of H^1 implies that the evaluation map is surjective, effectively collapsing the kernel of the identification problem and ensuring that the connecting homomorphism δ maps to a trivial space.

E.2 Proof of Proposition 4.5

Let G be a DAG on $[q] := \{1, \dots, q\}$, and write

$$\text{ch}_G(j) := \{\, i \in [q] \mid j \to i \text{ in } G \,\}.$$

For each context $k \in [m]$, let $S_k \subseteq [q]$ be the set of intervened nodes. We say that a matrix $\Lambda^{(k)}$ is *S_k-compatible* with Λ if

$$\lambda_{ij}^{(k)} = 0 \iff \lambda_{ij} = 0 \qquad \text{and} \qquad \lambda_{ij}^{(k)} = \lambda_{ij} \text{ for every } i \notin S_k.$$

Thus only the rows indexed by S_k are allowed to vary in context k. For a subset $T \subseteq [q]$, we write $A_T^{(k)}$ for the submatrix of $A^{(k)}$ formed by the columns indexed by T, and $A_{\bullet,i}^{(k)}$ for the i-th column of $A^{(k)}$.

Hereafter, we restate Proposition 4.5 in linear algebra language and explicitly introduce necessary extended modules as intervention design matrices.

Theorem E.2 (Partial identifiability under multi-target interventions)**.** *Assume that there exist matrices*

$$F \in \mathbb{C}^{p\times q}, \qquad \Lambda \in \mathbb{C}^{q\times q}, \qquad \Lambda^{(1)}, \dots, \Lambda^{(m)} \in \mathbb{C}^{q\times q},$$

such that Λ is supported on G, each $\Lambda^{(k)}$ is S_k-compatible with Λ, and

$$A^{(0)} = F(I-\Lambda)^{-1}, \qquad A^{(k)} = F(I-\Lambda^{(k)})^{-1} \quad (k = 1, \dots, m).$$

For each $j \in [q]$, set $E_j := \mathrm{ch}_G(j)$ and define

$$x_j := \big(\lambda_{ij}\big)_{i\in E_j} \in \mathbb{C}^{|E_j|}.$$

For each pair (k, j) let

$$T_{k,j} := E_j \cap S_k, \qquad R_{k,j} := E_j \setminus S_k.$$

Choose a matrix $P_{k,j}$ whose rows form a basis of the left kernel

$$\ker\big((A_{T_{k,j}}^{(k)})^\top\big) \subseteq \mathbb{C}^p,$$

with the convention $P_{k,j} = I_p$ when $T_{k,j} = \emptyset$. Let $M_{k,j}$ be the matrix whose columns are indexed by E_j and whose column indexed by $i \in E_j$ is

$$\big(M_{k,j}\big)_{\bullet,i} := \begin{cases} P_{k,j} A_{\bullet,i}^{(0)}, & i \in T_{k,j}, \\ P_{k,j}\big(A_{\bullet,i}^{(0)} - A_{\bullet,i}^{(k)}\big), & i \in R_{k,j}, \end{cases}$$

and set

$$b_{k,j} := P_{k,j}\big(A_{\bullet,j}^{(0)} - A_{\bullet,j}^{(k)}\big).$$

Finally, let $M[j]$ be the matrix obtained by stacking the blocks $M_{k,j}$ over $k = 1, \dots, m$, and let $b[j]$ be the vector obtained by stacking the $b_{k,j}$ in the same order.

Then the set of baseline parameters

$$\mathcal{F}_G\big(A^{(0)}, \dots, A^{(m)}\big) := \Big\{(F,\Lambda) \;\Big|\; \exists\, \Lambda^{(1)}, \dots, \Lambda^{(m)} \text{ as above with } A^{(k)} = F(I-\Lambda^{(k)})^{-1}\ \forall k\Big\}$$

is an affine linear subspace of the parameter space. More precisely, the map

$$(F,\Lambda) \longmapsto (x_1, \dots, x_q)$$

identifies $\mathcal{F}_G\big(A^{(0)}, \dots, A^{(m)}\big)$ with the direct product

$$\prod_{j=1}^{q} \{\, x_j \in \mathbb{C}^{|E_j|} \mid M[j]x_j = b[j] \,\}.$$

Consequently,

$$\dim \mathcal{F}_G\big(A^{(0)}, \dots, A^{(m)}\big) = \sum_{j=1}^{q} \big(|E_j| - \operatorname{rank} M[j]\big).$$

In particular, the baseline pair (F,Λ) is uniquely determined by the family $\{A^{(k)}\}_{k=0}^m$ if and only if

$$\operatorname{rank} M[j] = |E_j| \qquad \textit{for every } j \in [q].$$

E.3 Proof of Theorem E.2

Proof. We divide the proof into three steps.

Step 1: elimination of the mixing matrix. From the observational context, we have

$$A^{(0)} = F(I - \Lambda)^{-1},$$

hence necessarily

$$F = A^{(0)}(I - \Lambda).$$

Therefore the pair (F, Λ) is uniquely determined by Λ, and it suffices to characterize the possible baseline matrices Λ.

Step 2: columnwise reduction to a linear system. Fix $j \in [q]$. Since the support of Λ is G, only the entries λ_{ij} with $i \in E_j = \mathrm{ch}_G(j)$ may be nonzero. Thus the j-th column of Λ is completely encoded by the vector

$$x_j = (\lambda_{ij})_{i \in E_j}.$$

Now fix a context $k \in [m]$ and write

$$T := T_{k,j} = E_j \cap S_k, \qquad R := R_{k,j} = E_j \setminus S_k.$$

Because $\Lambda^{(k)}$ is S_k-compatible with Λ, the coefficients $\lambda_{ij}^{(k)}$ agree with λ_{ij} for every $i \in R$, whereas the coefficients with $i \in T$ are free variables. Let

$$y_{k,j} := (\lambda_{ij}^{(k)})_{i \in T} \in \mathbb{C}^{|T|}.$$

The identity

$$A^{(k)}(I - \Lambda^{(k)}) = A^{(0)}(I - \Lambda)$$

holds because both sides equal F. Reading the j-th column of this matrix identity yields

$$A^{(k)}_{\bullet,j} - \sum_{i \in R} A^{(k)}_{\bullet,i}\lambda_{ij} - \sum_{i \in T} A^{(k)}_{\bullet,i}\lambda^{(k)}_{ij} = A^{(0)}_{\bullet,j} - \sum_{i \in R} A^{(0)}_{\bullet,i}\lambda_{ij} - \sum_{i \in T} A^{(0)}_{\bullet,i}\lambda_{ij}.$$

Equivalently,

$$A^{(k)}_T\, y_{k,j} = A^{(k)}_{\bullet,j} - A^{(0)}_{\bullet,j} + \sum_{i \in R}\big(A^{(0)}_{\bullet,i} - A^{(k)}_{\bullet,i}\big)\lambda_{ij} + \sum_{i \in T} A^{(0)}_{\bullet,i}\lambda_{ij}. \tag{E.4}$$

We now eliminate the nuisance variables $y_{k,j}$. By construction, the rows of $P_{k,j}$ form a basis of

$$\ker\big((A^{(k)}_T)^\top\big).$$

Hence, by the fundamental theorem of linear algebra, a vector $v \in \mathbb{C}^p$ lies in $\mathrm{Im}(A^{(k)}_T)$ if and only if $P_{k,j}v = 0$. Therefore (E.4) admits a solution $y_{k,j}$ if and only if

$$P_{k,j}\left(A^{(k)}_{\bullet,j} - A^{(0)}_{\bullet,j} + \sum_{i \in R}\big(A^{(0)}_{\bullet,i} - A^{(k)}_{\bullet,i}\big)\lambda_{ij} + \sum_{i \in T} A^{(0)}_{\bullet,i}\lambda_{ij}\right) = 0.$$

By the definition of $M_{k,j}$ and $b_{k,j}$, this is exactly the linear system

$$M_{k,j}x_j = b_{k,j}.$$

Stacking these equations over all contexts $k = 1, \ldots, m$ shows that the j-th column of the baseline matrix Λ is admissible if and only if

$$M[j]x_j = b[j].$$

Step 3: product decomposition and dimension count. The preceding paragraph shows that the constraints on distinct columns j do not interact. Hence, the set of all admissible baseline matrices Λ is exactly the direct product

$$\prod_{j=1}^{q} \{\, x_j \in \mathbb{C}^{|E_j|} \mid M[j]x_j = b[j] \,\}.$$

Each factor is an affine linear subspace of dimension $|E_j| - \operatorname{rank} M[j]$. Since $F = A^{(0)}(I - \Lambda)$ depends linearly on Λ, the same product description and the same dimension formula hold for the baseline fiber $\mathcal{F}_G\big(A^{(0)}, \dots, A^{(m)}\big)$.

Finally, an affine linear space is a singleton if and only if its dimension is zero. Since the dimension is the sum of the nonnegative integers $|E_j| - \operatorname{rank} M[j]$, this happens if and only if every summand vanishes, namely if and only if

$$\operatorname{rank} M[j] = |E_j| \qquad \text{for all } j \in [q].$$

This proves the theorem. □

E.4 Proof of Corollary 4.6

Proof. The first statement is immediate, since the definition of a soft intervention is exactly the row-wise compatibility condition imposed in Theorem E.2. Hence all conclusions of that theorem apply directly.

The final claim follows from the fact that Theorem E.2 describes only the projection of the compatible parameter set onto the baseline coordinates (F, Λ). The context-specific matrices $\Lambda^{(k)}$ remain free on the intervened rows, and thus may still generate a positive-dimensional compatible family even when the baseline pair is uniquely determined. □

Corollary E.3 (Hard interventions: strengthened linear system)**.** *Assume that, for each context $k \in [m]$, the intervention is hard on the target set S_k, i.e.*

$$\lambda_{ij}^{(k)} = 0 \qquad \textit{for all } i \in S_k,\ j \in [q],$$

and

$$\lambda_{ij}^{(k)} = \lambda_{ij} \qquad \textit{for all } i \notin S_k,\ j \in [q].$$

Fix $j \in [q]$, and retain the notation

$$E_j = \mathrm{ch}_G(j), \qquad T_{k,j} = E_j \cap S_k, \qquad R_{k,j} = E_j \setminus S_k.$$

Then the admissible baseline column

$$x_j = (\lambda_{ij})_{i \in E_j} \in \mathbb{C}^{|E_j|}$$

is characterized by the explicit linear system

$$\widetilde{M}[j]\, x_j = \widetilde{b}[j],$$

where the block corresponding to context k is given by

$$\widetilde{M}_{k,j} = \Big[\, A^{(0)}_{\bullet,i} \,\Big]_{i \in T_{k,j}} \,\Big|\, \Big[\, A^{(0)}_{\bullet,i} - A^{(k)}_{\bullet,i} \,\Big]_{i \in R_{k,j}},$$

and

$$\widetilde{b}_{k,j} = A^{(0)}_{\bullet,j} - A^{(k)}_{\bullet,j}.$$

Consequently, the set of baseline pairs (F, Λ) compatible with the data is an affine linear space of dimension

$$\sum_{j=1}^{q} \big(|E_j| - \operatorname{rank} \widetilde{M}[j]\big),$$

and the baseline pair (F, Λ) is uniquely determined if and only if

$$\operatorname{rank} \widetilde{M}[j] = |E_j| \qquad \textit{for all } j \in [q].$$

F NON-IDENTIFIABILITY AND OPEN PROBLEMS

F.1 UNIVERSE OF NON-IDENTIFIABILITY

In this section, we discuss the exceptional cases where non-identifiability holds. As we have remarked at the very beginning, non-identifiability and identifiability are neither completely statistical assumptions nor modeling coincidences, but an insurmountable defectiveness in nature that, unfortunately, cannot be fully retained by any measures. In accordance with many previous sections, we will discuss when generic points become "less generic" due to invisible conditions imposed by geometric principles. While the fundamental geometric defectiveness—inherent to the network's algebraic structure — presents an unavoidable theoretical hurdle, we demonstrate that one can nonetheless extend the identifiable regime by pushing the trajectories to areas closer to non-defectiveness. Lastly, we provide the first formal proof that the widely used computational formulation indeed does a similar job.

Provided that the hypothetical models are simple, identifiability can always be treated as a general problem of point-dimension decision.

F.1.1 GENERIC POINT CAUSING DEPENDENCE

Our theorems establish that, barring a few pathological conditions, even transcendental varieties associated with neural networks attain their expected dimensions, thereby ensuring model identifiability. A key technical challenge arises when geometric defectiveness occurs independently of the parameter-level genericity. In such instances, the architecture forces points to lie on twisted curves or higher-dimensional singular loci, rendering them 'non-general' from the perspective of the ambient space. We introduce the term non-general generic points to characterize this phenomenon: parameters that are generic within the constrained learning model, yet lack the degrees of freedom required for general position in the embedding space. To make the demonstration accessible, we formulate the following fact.

Fact 1. Even when points are sampled generically in the parameter space, their images under the model map are not necessarily in General Position within the ambient projective space. Similar to the classical exceptions in the Alexander-Hirschowitz Theorem, these points are constrained to lie on specific algebraic structures—such as rational normal curves or hypersurfaces—which forces their associated tangent schemes to overlap excessively, thereby inhibiting identifiability.

F.1.2 NON-GENERIC POINT CAUSING DEFECTIVENESS

In a set of points, if it contains any non-generic points, for example, the pathological parameters with severe co-linearity, the defectiveness will be a must.

F.2 OPEN PROBLEMS

F.3 VANISHING SINGULARITY FOR TRANSCENDENTAL ACTIVATIONS

Algebraic neurovarieties, induced by polynomial neural networks, provide a foundation for the identifiability theory of neural models. Unfortunately, although the transparent identifiability of polynomial neural networks is well established, training such networks poses several challenges, necessitating the exploration of other, more efficient architectures. In this section, we focus on translating established results for neurovarieties to more complex architectures. We highlight that a few problems of transcendental neurovarieties remain open but are worth further exploration (Shahverdi et al., 2026b;a).

Question 1. Given an architecture of generic neural networks with its activation function transcendental, such as $\mathrm{erf}(x) := x \cdot \Phi(x)$, can we establish its vanishing theorem, and are exceptional cases measure zero?

Conjecture I *Let $\mathcal{V}_\sigma$ be a neurovariety generated by a transcendental activation $\sigma \in \mathbb{R}_{\text{an,exp}}$. The evaluation morphism $\bigoplus \phi$ on the sheaf of transcendental modules $\mathcal{O}_{\mathcal{V}_\sigma}$ attains maximal rank for a generic configuration of double points. The set of "non-identifiable" parameters $\Theta_{sing} \subset \Theta$ is a definable set of $dim(\Theta_{sing}) < dim(\Theta)$, and thus possesses measure zero.*

This conjecture is a non-trivial transition from fully algebraic functions to systems that can never be defined by a finite number of algebraic equations. We hypothesize that the vanishing theorem for transcendental neurovarieties holds provided the activation functions belong to a Pfaffian chain. In such cases, the system is confined to an analytic sub-manifold where the number of non-identifiable configurations is bounded by the Khovanskii index of the architecture (Khovanskii, 1991; Cassidy, 1972). Consequently, the independence of conditions imposed by $\mathcal{I}_X^2$ is preserved, and the set of exceptional parameters remains a σ-finite set of measure zero within the transcendental parameter space.

F.3.1 Intervention Lower Bounds

In Proposition 4.5 and Theorem E.2, we established that a set of general interventions (encompassing both hard and soft types) for each module extension can *always* impose independent linear constraints alongside the symbolic square $\mathcal{I}_X^{(2)}$. However, a fundamental question remains: what is the *minimal* number of interventions required to ensure global identifiability?

Conjecture II (Generic Surjectivity). For a given configuration of dimensions (p, q, k), there exists a threshold $N(p, q, k) \in \mathbb{N}$ such that for any number of independent intervention targets $n \geq N$ in module extension k, the multi-target evaluation map

$$\mathcal{E} := \bigoplus_{i=1}^{n} \phi_i$$

is generically of full rank.