THE IDENTIFIABILITY OF VERTEX SIGNALS ON GRAPHS UNDER QUADRATIC READOUTS

Anonymous authors
Paper under double-blind review

ABSTRACT

A readout layer R of a graph neural network is a permutation invariant map from vertex signals $X \in \mathbb{R}^{n\times d}$ to some fixed vector space agnostic to the size n of the graph. We measure how a readout R can fail to locally distinguish a signal X in $\mathbb{R}^{n\times d}$ using the dimension δ of the fibre $R^{-1}(R(X)) \subset \mathbb{R}^{n\times d}$. In this study, we consider a quadratic readout of the form $R(X) = (\mathbf{1}^T X, X^T X, X^T M X)$, where M is the adjacency or Laplacian matrix of the graph. We show that if the signal channel dimension d is sufficiently large relative to the size of the graph ($d \geq n$), and the spectrum of M is simple, we can achieve perfect local identifiability $\delta = 0$ for generic (full rank) vertex signals. In contrast, the commonly used sum readout $\mathbf{1}^T X$ or a graph agnostic $(\mathbf{1}^T X, X^T X)$ quadratic readout has fibre dimension δ that is non-decreasing with d. These observations inform the design of graph neural network architectures, as they demonstrate how information loss in the readout layers can be potentially mitigated with careful consideration of the nonlinearity of the readout, the number of vertex signal channels d, and how it relates to the size of the graph n.

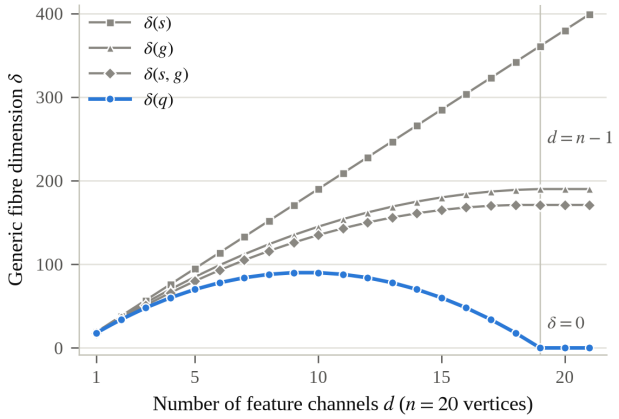

Figure 1: Generic fibre dimension δ of four readouts on a graph with $n = 20$ vertices whose graph matrix M has simple spectrum.

1 INTRODUCTION

For tasks such as graph regression or classification, a graph neural network (GNN) is often used to represent graphs with vertex attributes as vectors in a Euclidean space. A key component of such a graph neural network is the *readout* layer, which is used to aggregate vertex representations obtained from message passing layers of the GNN into one fixed-size vector. Crucially, a readout layer is an aggregation of vertex features that needs to be independent of how vertices are ordered. Concretely, for a *set* of vertex features represented as row vectors in a matrix $X \in \mathbb{R}^{n\times d}$, a readout R is a function that is invariant with respect to row permutations $P \in \Pi_n$:

$$R(X) = R(PX). \tag{1}$$

The requirement of permutation invariance inevitably introduces a *compression* of information. Ideally, a readout ought to be injective on the sets of features: two sets of vertex features $X, Y \in \mathbb{R}^{n\times d}$ should have the same readout, if and

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.

Table 1: The generic fibre dimension of the readouts when the vertex feature dimension $d \geq n$ exceeds the number of vertices. Note that the quadratic readouts plateaus. Here r_C is the rank of the matrix $C \in \mathbb{R}^{n\times c}$, and $m_1, \ldots, m_\ell$ are the multiplicities of the eigenvalues of the symmetric matrix M.

	Linear (Proposition 3.1)	Quadratic (Theorem 3.4)
Graph Agnostic	$\delta(\mathbf{s}) = d(n-1)$	$\delta(\mathbf{g}) = \binom{n}{2}$
Graph Informed	$\delta(\mathbf{s}_C) = d(n-r_C)$	$\delta(\mathbf{g}, \mathbf{g}_M) = \sum_{j=1}^{\ell} \binom{m_j}{2}$

only if they are related by some permutation matrix $Y = PX$. This ideal is often failed in practice. For the common *sum* readout, where

$$\mathbf{s} : X \in \mathbb{R}^{n\times d} \mapsto \mathbf{1}^T X \in \mathbb{R}^d, \tag{2}$$

the subspace of vertex features in $\mathbb{R}^{n\times d}$ that share the same sum readout is a linear subspace of dimension $n(d-1)$. This is far from the ideal injective case, where such a space would only consist of $n!$ points in $\mathbb{R}^{n\times d}$.

The compression introduced by readout layers can limit a graph neural network's performance in downstream tasks, as vertex feature sets with the same readout cannot be distinguished using the readout alone. Our work seeks to quantify the ambiguity a readout layer R introduces on d-dimensional vertex feature sets over graphs of a fixed size n. Regarding R as a permutation invariant function $R : \mathbb{R}^{n\times d} \to \mathbb{R}^h$, we study for a given vertex features $X \in \mathbb{R}^{n\times d}$, the set of other vertex features that are indistinguishable from X after applying the readout. We call this subset in $\mathbb{R}^{n\times d}$ the *fibre* through X:

$$R^{-1}(R(X)) := R^{-1}(R(X)) = \left\{Y \in \mathbb{R}^{n\times d} : R(Y) = R(X)\right\}. \tag{3}$$

We use the *dimension* of such fibres to quantify the local ambiguity on vertex feature sets introduced by the readout. Specifically, we study the *generic fibre dimension*, denoted by $\delta(R)$, which is the common dimension of the fibre manifolds on an open dense subset of vertex feature space (Definition 2.1). Positive fibre dimension describes continuous families of indistinguishable feature set. When the generic fibre dimension is zero, a generic feature set is *locally identifiable*: no other sufficiently nearby signal has the same readout. Furthermore, if the fibre manifold is compact, then $\delta(R) = 0$ also implies that the fibre consist of finitely many points.

For a given readout R, we aim to deduce how its generic fibre dimension $\delta(R)$ depends on the number of vertices of the graph, as well as the number of feature channels d. We focus on a family of readouts that are *polynomials* in entries of the matrices X, such as the sum readout $\mathbf{s}$ above. This has the advantage that for proper polynomial maps, the generic fibre dimension $\delta(R)$ is well-defined, and the fibres $R^{-1}(R(X))$ are generically embedded $\delta(R)$-dimensional submanifolds (Proposition 2.1). We restrict ourselves to deriving the fibre dimension of polynomial readouts up to second order, belonging to two families. The first family are purely graph agnostic on the readout level: this include the sum readout above, as well as a quadratic second order readout using the Gram matrix (Wang & Ji, 2023b):

$$\mathbf{s}(X) = X^\top \mathbf{1} \in \mathbb{R}^d, \tag{4}$$

$$\mathbf{g}(X) = X^\top X \in \mathrm{Sym}_d. \tag{5}$$

Here Sym_d denotes the vector space of real symmetric $d \times d$ matrices, which we can regard as a $d(d+1)/2$ dimensional vector space having coordinates in the upper triangular half of a $\mathbb{R}^{n\times n}$ matrix. The second family consists of generalisations of the readouts above to incorporate information from the underlying graph. These can be a fixed set of c-dimensional vertex features $C \in \mathbb{R}^{n\times c}$ (such as positional encodings, or one-hot encodings of categorial labels), or an $n \times n$ symmetric matrix $M \in \mathrm{Sym}_n$, such as the graph's Laplacian or adjacency matrix. We call the readouts below that encode such information as *graph-informed readouts*

$$\mathbf{s}_C(X) = X^T C \in \mathbb{R}^{d\times c} \tag{6}$$

$$\mathbf{g}_M(X) = X^\top M X \in \mathrm{Sym}_d. \tag{7}$$

These readouts are invariant under simultaneous relabelling of the graph and its vertex signal. Fixing the graph-level information C and M, the generic fibre dimension δ for such readouts measures how vertex features *on the same graph* can fail to be locally identifiable.

Our main findings consists of explicit computations of the generic fibre dimensions of such readout components. We give an illustration of the generic fibre dimensions of readouts $\mathbf{s}, \mathbf{g}, (\mathbf{s}, \mathbf{g})$, and $\mathbf{q} = (\mathbf{g}, \mathbf{g}_M)$ as d varies for fixed n in Figure 1. First, we prove that the Gram readout strictly reduces generic fibre dimensionality over their linear counterparts:

$$\delta(\mathbf{g}) < \delta(\mathbf{s}), \tag{8}$$

$$\delta(\mathbf{s}_C, \mathbf{g}) < \delta(\mathbf{s}_C). \tag{9}$$

Second, including graph-informed readouts also strictly reduces the fibre dimension:

$$\delta(\mathbf{s}_C, \mathbf{g}) < \delta(\mathbf{g}), \tag{10}$$

$$\delta(\mathbf{g}, \mathbf{g}_M) < \delta(\mathbf{g}). \tag{11}$$

Our third and most significant finding is that graph-informed readouts can achieve identifiability $\delta = 0$ under two settings. For the quadratic readouts, the combination $\mathbf{q} = (\mathbf{g}, \mathbf{g}_M)$ of the Gram map $\mathbf{g}$ and the graph-informed Gram map $\mathbf{g}_M$ can achieve local identifiability for generic vertex features X, if M has simple spectrum, and $d \geq n - 1$:

$$\text{For} \quad \mathbf{q} = (\mathbf{g}, \mathbf{g}_M), \quad \delta(\mathbf{q}) = \max\big\{d(n - d - 1),\, 0\big\}. \tag{12}$$

In other words, incorporating a graph-dependent quadratic term makes generic vertex signals locally identifiable when sufficiently many feature channels are available. We furthermore show that for $d \geq n$, the generic fibre of $\mathbf{q}$ consists of 2^n points. For the linear readout $\mathbf{s}_C$, local identifiability $\delta = 0$ corresponds to the trivial case where all vertices can be distinguished from one another by their positional encoding C and permutation symmetry between the vertices is broken. We give more details of our contributions in the list below.

Contributions.

- We characterize the fibres of the Gram readout $\mathbf{g}$ as Stiefel manifolds, which allow us to deduce the generic dimension of $\mathbf{g}$ (Proposition 3.3).
- We show that the fibre of $(\mathbf{s}_C, \mathbf{g})$ is also a Stiefel manifold with lower dimension (Proposition 3.4).
- We establish the generic fibre dimension of the graph-informed Gram readout $\mathbf{q} = (\mathbf{g}, \mathbf{g}_M)$ under simple spectrum, obtain a lower bound for its generic fibre dimension for arbitrary spectral multiplicities of M (Theorem 3.4).
- We establish that including additional readout components $\mathbf{g}_{p(M)}$ on top of $\mathbf{q} = (\mathbf{g}, \mathbf{g}_M)$, where $p(M)$ is a polynomial of M, does not reduce the fibre in the $d \geq n$ regime (Proposition 3.6).

RELATED WORK

Readouts and graph pooling. In message passing networks (Gilmer et al., 2017), standard readouts take coordinate-wise sums, means or maxima of vertex features (Duvenaud et al., 2015; Xu et al., 2019; Buterez et al., 2022); alternatives use attention (Vinyals et al., 2016; Lee et al., 2019; Baek et al., 2021), sorting (Zhang et al., 2018) or averaging over vertex orderings (Murphy et al., 2019). Closest to our setting are second-order readouts. Second-order pooling uses the Gram matrix $X^\top X$ (Wang & Ji, 2023a), and DKEPool uses the product of the covariance matrix and mean of the vertex features (Chen et al., 2022). Recently, Talhi et al. (2026) explored a different expressivity bottleneck where the readout hampers 1-WL expressivity.

Injective multiset functions. Another line of work bounds the size of permutation-invariant embeddings that separate all multisets of n vectors. Representing every continuous permutation-invariant function of n reals by a continuous sum decomposition $\rho(\sum_i \phi(x_i))$ (Zaheer et al., 2017) requires a latent dimension of at least n (Wagstaff et al., 2019; 2022), and distinguishing all multisets of n reals requires at least n continuous aggregators (Corso et al., 2020). For vectors in $\mathbb{R}^d$, $2nd + 1$ random invariants from a continuous semialgebraic family of separating invariants suffice (Dym & Gortler, 2025), as do $2nd + 1$ neural moments with an analytic non-polynomial activation, for almost all parameters (Amir et al., 2023). All of these bounds grow with n. We instead fix low-degree polynomial readouts and determine how their generic fibre dimension depends on n and d.

Invariants and identifiability. In neuroalgebraic geometry, the generic fibres of the map from network parameters to functions formalise identifiability, and the dimension of the function space is the number of parameters minus the generic fibre dimension (Marchetti et al., 2025), computable from Jacobian ranks (Kileel et al., 2019). Henry et al. (2025a) describe the generic fibres of deep lightning self-attention networks and deduce the dimension of their function space. Deep polynomial networks with non-increasing layer widths are generically identifiable under mild conditions (Usevich et al., 2025), and for a sufficiently generic polynomial activation, almost every function of an MLP arises from only finitely many parameter choices (Shahverdi et al., 2026). Identifiability up to symmetries has also been studied for tanh and ReLU networks (Sussmann, 1992; Stock & Gribonval, 2023; Bona-Pellissier et al., 2022).

Relative to these lines of work, we study the map from vertex signals to readout values: we compute exact generic fibre dimensions of fixed low-degree readouts, including the Gram readout of second-order pooling, and show that adding the graph-dependent term $X^\top M X$ to it makes generic signals locally identifiable once $d \geq n - 1$, provided that M has simple spectrum (Theorem 3.4). Section 4 discusses readout design, the number of channels and the limits of local identifiability.

2 Generic Fibre Dimension of Polynomial Maps

Before we analyse the readout maps, we first formally define the generic fibre dimension.

Definition 2.1. *Given a map $f : \mathbb{R}^a \to \mathbb{R}^b$, if there is a generic (i.e. open and dense) subset $\mathcal{G}$ where the fibres of the map $f^{-1}(f(x))$ are all δ-dimensional manifolds for all $x \in \mathcal{G}$, then δ is known as the* generic fibre dimension *of f.*

Here a generic subset means an *open and dense* subset. While the generic fibre dimension may not be defined for any map f, they are defined for linear maps (where $\mathcal{G} = \mathbb{R}^a$), and for proper polynomial maps. This is useful to us, as the property of a map being proper holds for readout maps involving the Gram matrix $X^T X$ (Lemma A.1).

Proposition 2.1. *Let $f : \mathbb{R}^a \to \mathbb{R}^b$ be a* proper *polynomial map. Then the generic fibre dimension δ is defined for f. That is, there is a generic subset $\mathcal{G} \subset \mathbb{R}^a$, on which $f^{-1}(f(x))$ are δ-dimensional manifolds. Furthermore, for $x \in \mathcal{G}$, the Jacobian of the polynomial $\mathcal{D}f$ attains its maximal rank*

$$x \in \mathcal{G} \implies \operatorname{rank} \mathcal{D}f|_x = \max_{y \in \mathbb{R}^a} \operatorname{rank} \mathcal{D}f|_y =: \rho, \tag{13}$$

and the generic fibre dimension is given by $\delta = a - \rho$.

We give a self-contained proof of Proposition 2.1 in Section A.1. The dimension and Jacobian of real polynomial maps is a well-studied subject in real algebraic geometry; we refer the reader to standard texts such as Bochnak et al. (1998); Basu et al. (2006) for full treatment. In a machine learning setting, the generic dimension has also been used a tool to measure the identifiability of parameters in nuerla networks (Henry et al., 2025b).

Here we give some examples of generic subsets of Euclidean spaces that would feature in this study.

Example 2.1 ((Okamoto, 1973),(Mityagin, 2020, Prop. 1))**.** *Let $p : \mathbb{R}^n \to \mathbb{R}$ be a polynomial. Then $p^{-1}(0)$ has Lebesgue measure zero and nowhere dense, and $\mathbb{R}^n \setminus p^{-1}(0)$ is open and dense in $\mathbb{R}^n$,*

As rank deficient matrices in $\mathbb{R}^{n \times d}$ are those whose $r \times r$-matrix minors all vanish for $r = \min(n, d) + 1$, the rank deficient matrices are a finite union of zeros of polynomials. Hence their complement — full rank matrices— are generic.

Example 2.2. *Full row/column rank matrices $\{X \in \mathbb{R}^{n \times d} \mid \operatorname{rank}(X) = \min(n, d)\}$ are generic in $\mathbb{R}^{n \times d}$, and its complement has Lebesgue measure zero.*

3 The Fibre Dimension of Readouts

3.1 The Sum and Linear Readout

Recall $\mathbf{s}_C : \mathbb{R}^{n \times d} \to \mathbb{R}^{d \times c}$ is the readout given by $\mathbf{s}_C(X) = X^T C$. If C is the vector of ones $\mathbb{1} \in \mathbb{R}^{d \times 1}$, then, then $\mathbf{s}_{\mathbb{1}} = \mathbf{s}$ is just the standard sum readout. We can regard the rows of C as c-dimensional positional encodings of vertices, derived from the graph. Since $\mathbf{s}_C$ is a linear map, the following description of its fibre follows from applying the rank-nullity theorem.

Proposition 3.1. *Suppose $C \in \mathbb{R}^{n \times c}$ has rank $r_c \leq \min(n, c)$. The fibre of $\mathbf{s}_C$ is isomorphic to $\mathbb{R}^{(n - r_c) \times d}$.*

In particular, if $r_c = n$ (necessitating $c \geq n$) then $\mathbf{s}_C$ is injective. As this is equivalent to the n row vectors of C being linearly independent, this corresponds to the trivial case where the positional encodings of vertices (the rows of C) embed the vertices in a linear subspace.

3.2 The Gram Map

Recall the Gram map $\mathbf{g} : \mathbb{R}^{n \times d} \to \mathbb{R}^{d \times d}$ maps a matrix to its Gram matrix $X \mapsto X^T X$. The fibres of the Gram map which are known to be $O(n)$ orbits.

Theorem 3.1 ((Horn & Johnson, 2013, Thm. 7.3.11))**.** *The fibre of the Gram map $\mathbf{g}(\mathbf{g}^{-1}(X))$ through X is the orbit $O(n) \cdot X$.*

As such we can equivalently consider the action of $O(n)$ on $\mathbb{R}^{n \times d}$ to characterise the fibre of the Gram map through X. The following classical result for compact Lie groups states that such orbits are manifolds, and are defined by the stabiliser subgroup. For completeness we reproduce a proof in Section A.2.

Theorem 3.2. *Consider a compact Lie group G acting smoothly on a manifold M. Then for every $x \in M$, the space of cosets G/G_x of the stabiliser subgroup G_x is a compact, smooth manifold, and there is an embedding of $F : G/G_x \hookrightarrow M$ given by $gG_x \mapsto g \cdot x$ onto the orbit $G \cdot x$.*

We show below that the stabiliser subgroup $O(n)_X = \{U \in O(n) \mid UX = X\}$ of X, and its coset $O(n)/O(n)_X$, are determined by the rank of X up to isomorphism: the condition that $UX = X$ means U fixes any point in the r-dimensional subspace of $\mathrm{col}(X) \subseteq \mathbb{R}^n$ spanned by columns of X. In other words, $O(n)_X = O(n)_{\mathrm{col}(X)}$. The cosets $O(n)/O(n)_X$ are in fact a Stiefel manifold $\mathcal{V}_r(\mathbb{R}^n)$, which is the space of orthonormal k-frames in $\mathbb{R}^n$ (the set of matrices $A \in \mathbb{R}^{n\times r}$ that satisfy $A^TA = \mathrm{id}_k$). We defer the proof to Section A.2.

Proposition 3.2. *Suppose $X \in \mathbb{R}^{n\times d}$ has matrix rank $r \leq n$. Then $\mathbf{g}^{-1}(\mathbf{g}(X)) = O(n) \cdot X$ is an embedding of the compact Stiefel manifold $\mathcal{V}_r(\mathbb{R}^n) = O(n)/O(n-r)$, with dimension*

$$\dim \mathbf{g}^{-1}(\mathbf{g}(X)) = \dim O(n)/O(n-r) = \binom{n}{2} - \binom{n-r}{2} = nr - \binom{r+1}{2}. \tag{14}$$

Proof. As $O(n)$ is an compact Lie group acting smoothly on $\mathbb{R}^{n\times d}$, Theorem 3.2 implies the $O(n)$ orbit of a matrix X is an embedding of the coset space $O(n)/O(n)_X$ which is a compact manifold. As $O(n)_X = O(n)_{\mathrm{col}(X)}$, and $\dim \mathrm{col}(X) = \mathrm{rank}\, X = r$, we can apply Lemma A.7 to show that $O(n)/O(n)_X$ is diffeomorphic to $O(n)/O(n-r)$; thus

$$\dim \mathbf{g}^{-1}(\mathbf{g}(X)) = \dim O(n)/O(n)_X = \dim O(n)/O(n-r) = \binom{n}{2} - \binom{n-r}{2} = nr - \binom{r+1}{2}. \tag{15}$$

□

In particular, if $X \in \mathbb{R}^{n\times d}$ has full rank $r = \min(n, d)$, then $\mathbf{g}^{-1}(\mathbf{g}(X))$ is diffeomorphic to $O(n)$ for $d \geq n$, and $O(n)/O(n-d)$ for $d < n$. As full rank matrices are generic, the dimension $\mathbf{g}^{-1}(\mathbf{g}(X))$ is the generic fibre dimension of the Gram map.

Proposition 3.3. *The generic fibre dimension of the Gram map $\mathbf{g} : \mathbb{R}^{n\times d} \to \mathrm{Sym}_d$ is given by*

$$\delta(\mathbf{g}) = \begin{cases} nd - \binom{d+1}{2} & d \leq n \\ \binom{n}{2} & d \geq n \end{cases}. \tag{16}$$

Proof. Recall full rank matrices X are a generic subset, and the fibre through such matrices $\mathbf{g}^{-1}(\mathbf{g}(X))$ have the same dimension (Equation (14) and proposition 3.2). As $\mathbf{g}$ is a proper map (Lemma A.1), Proposition 2.1 then implies the dimension given in eq. (14) is the generic fibre dimension of the Gram map $\mathbf{g}$. Substituting the full rank condition $r = d$ for $d \leq n$ and $r = n$ for $d \geq n$ into eq. (14) yields $\delta(\mathbf{g})$. □

We now consider how combining a matrix-informed linear readout $\mathbf{s}_C$ and the quadratic readout $\mathbf{g}$ can reduce the fibre dimension. In order to characterise the fibre of the joint map $f = (\mathbf{s}_C, \mathbf{g})$, let us consider an equivalent condition.

Lemma 3.1. *Consider the map $f(X) = (X^TC, X^TX)$ where $C \in \mathbb{R}^{n\times c}$. Let $\omega : \mathbb{R}^{n\times d} \hookrightarrow \mathbb{R}^{n\times(d+c)}$, be the map $w(X) = [X \mid C]$. Let $\tilde{\mathbf{g}} : \mathbb{R}^{n\times(d+c)} \to \mathbb{R}^{(d+c)\times(d+c)}$ be the Gram map on $\mathbb{R}^{n\times(d+c)}$. Consider then the augmented Gram map given by $\tilde{f} = \tilde{\mathbf{g}} \circ \omega$. Then the two maps have identical fibres:*

$$\tilde{f}^{-1}(\tilde{f}(X)) = f^{-1}(f(X)). \tag{17}$$

Proof. We can explicitly compute $\tilde{f}(X)$: in block form,

$$\tilde{f}(X) = \tilde{\mathbf{g}}(\omega(X)) = \begin{pmatrix} X^TX & X^TC \\ (X^TC)^T & C^TC \end{pmatrix}. \tag{18}$$

Since C is fixed, we thus have $\tilde{f}(X) = \tilde{f}(Y)$ iff $X^TC = Y^TC$ and $X^TX = Y^TY$ i.e. $f(X) = f(Y)$. □

By framing the map f in terms of a gram map $\tilde{g}$, we can apply the results for the Gram map from before to characterise the fibres of f.

Proposition 3.4. *Let $C \in \mathbb{R}^{n\times c}$ be some fixed, full rank matrix ($c \leq n$). Let $f : \mathbb{R}^{n\times d} \to \mathbb{R}^{d\times c} \times \mathrm{Sym}_d$ be the map*

$$f(X) = (X^T C, X^T X). \tag{19}$$

Let $k = \mathrm{rank}[X \mid C] \in \mathbb{R}^{n\times(d+c)}$. Then $f^{-1}(f(X))$ is a closed embedded submanifold of $\mathbb{R}^{n\times d}$, diffeomorphic to $O(n-c)/O(n-k)$, with dimension

$$\dim f^{-1}(f(X)) = \binom{n-c}{2} - \binom{n-k}{2}. \tag{20}$$

Proof. Due to Lemma 3.1 we can equivalently consider the fibre of $\tilde{f} = \tilde{\mathbf{g}} \circ \omega$. As a shorthand, let $Y = [X \mid C] = \omega(X)$. By Theorem 3.1, $\tilde{\mathbf{g}}^{-1}(\tilde{\mathbf{g}}(Y))$ is the $O(n)$ orbit on Y:

$$U[X \mid C] = [UX \mid UC], \qquad U \in O(n). \tag{21}$$

The orbit intersects the image of ω where $UC = C$; thus, the fibre through is X of $\tilde{f}$ is

$$\tilde{f}^{-1}(\tilde{f}(X)) = \{UX \mid U \in O(n),\ UC = C\} = O(n)_C \cdot X \tag{22}$$

where $O(n)_C$ is the stabiliser subgroup of C. Let $W = \mathrm{col}(C)$ be the $\dim c$-column space of C. By Lemma A.7, $O(n)_C = O(n)_W \cong O(n-c)$. Consider then the action of $O(n)_W$ on $\mathbb{R}^{n\times d}$. As a compact Lie group acting smoothly on $\mathbb{R}^{n\times d}$, Theorem 3.2 implies the orbit $O(n)_W \cdot X$ is an embedding of $O(n)_W/(O(n)_W)_X$. Now

$$U \in (O(n)_w)_X \quad \iff \quad U \in O(n),\ UX = X,\ UC = C \quad \iff \quad U \in O(n)_V,\ V = \mathrm{col}(X) + W. \tag{23}$$

Then $O(n)_W/(O(n))_X = O(n)_W/O(n)_V$. Let $k = \dim V = \mathrm{rank}[X \mid C]$. Then Lemma A.7 implies $O(n)_W/O(n)_V = O(n-c)/O(n-k)$, a manifold with dimension

$$\dim O(n-c)/O(n-k) = \binom{n-c}{2} - \binom{n-k}{2}. \tag{24}$$

□

We note that unlike the pure Gram map, combining the Gram map with a component $X^T C$ forces not the rank of X to be constant on fibres of f, but rather the rank of $[X \mid C]$. For the minimal fibre dimension, we wish to maximise the rank of $[X \mid C] = \min(n, d+c)$. The subset of matrices which maximises the combined rank is also generic.

Lemma 3.2. *Let $C \in \mathbb{R}^{n\times c}$ be some fixed, full rank matrix ($c \leq n$). There is a generic subset $\mathcal{G} \subseteq \mathbb{R}^{n\times d}$ such that for $X \in \mathcal{G}$, the matrix $[X \mid C]$ attains full rank $\min(n, d+c)$.*

We defer the proof to Section A.1.

Theorem 3.3. *Let $C \in \mathbb{R}^{n\times c}$ be some fixed, full rank matrix ($c \leq n$). The generic fibre dimension of $(\mathbf{s}_C, \mathbf{g})$ is given by*

$$\delta(\mathbf{s}_C, \mathbf{g}) = \begin{cases} d(n-c) - \binom{d+1}{2} & d \leq n-c \\ \binom{n-c}{2} & d \geq n-c \end{cases}. \tag{25}$$

Proof. We have shown that for X in a generic subset $\mathcal{G} \in \mathbb{R}^{n\times d}$, the matrix $[X \mid C]$ attains full rank lemma 3.2. From Proposition 3.4, the fibres of the map f through X are closed embedded submanifolds $O(n-c)/O(n-k)$ where k is the rank of $[X \mid C]$. In particular, for X in $\mathcal{G}$ the fibres $f^{-1}(f(X))$ are submanifolds $O(n-c)/O(n-k)$ where $k = \min(n, d+c)$, with dimensions given in eq. (25). By Proposition 2.1, the dimension of such manifolds is the generic fibre dimension δ of the map f, as f is proper (Lemma A.1). □

Note that the fibre dimension in eq. (25) is monotonically increasing in d up to $d = n - c$, where it plateaus.

3.3 Matrix Informed Gram Maps

For $M \in \mathrm{Sym}_n$, the M-weighted Gram map is the map $\mathbf{g}_M : \mathbb{R}^{n\times d} \to \mathrm{Sym}_d$.

$$\mathbf{g}_M(X) = X^T M X \tag{26}$$

As an example, if we take $M = L$ to be the Graph Laplacian, then the diagonal of $\mathbf{g}_L(X) = (X^T L X)$ would measure the Dirichlet energy of each column of X. As such, $\mathbf{g}_L$ can be thought of as measuring the spread of the vertex signals

while taking the graph structure into account. In contrast, the pure Gram matrix with $M = \mathrm{id}$ captures information about the covariance of the signal between channels. If we take $M = A$ to be the adjacency matrix, then the Gram matrix corresponds to measuring the inner product between signal X before and after a linear message passing aggregation AX.

Let us consider how adding an M-weighted Gram map can reduce the generic fibre dimension of the pure Gram map $\mathbf{g}$. We consider the map $\mathbf{q} = (\mathbf{g}, \mathbf{g}_M) : \mathbb{R}^{n\times d} \to \mathrm{Sym}_d \times \mathrm{Sym}_d$, and analyse the dimension generic fibre of $\mathbf{q}$.

Theorem 3.4. *Consider the map $\mathbf{q} : \mathbb{R}^{n\times d} \to \mathrm{Sym}_d \times \mathrm{Sym}_d$, given by $\mathbf{q}(X) = (X^TX, X^TMX)$ for some $M \in \mathrm{Sym}_n$. Then the generic fibre dimension of $\mathbf{q}$ is lower bounded by*

$$\delta \geq \max(d(n-d-1), 0). \tag{27}$$

Equality is attained when M has a simple spectrum.

We split the derivation of the generic fibre dimension into two cases; the first is the 'wide' regime $n \leq d$ where the number feature channels is no less than the number of vertices; the second is the 'tall' regime $n > d$ where the number of vertices is greater than the number of feature channels. These two cases are accounted for respectively in Propositions 3.5 and 3.7, which Theorem 3.4 is a summary of.

Notation 3.1. *For a symmetric matrix $M \in \mathrm{Sym}_d$, we let $\lambda_1 < \cdots < \lambda_\ell$ denote the unique eigenvalues of M, with multiplicity $m_j \geq 1$. Let $E_j \subseteq \mathbb{R}^n$ be the m_j-dimensional eigenspace corresponding to eigenvalue λ_j. As M is symmetric we have an orthogonal decomposition into eigenspaces $\mathbb{R}^n = \bigoplus_{j=1}^{\ell} E_j$. We say the spectrum of M is simple if $m_j = 1$ and $\ell = n$. Let $\Pi_j \in \mathbb{R}^{n\times n}$ denote the projection matrix onto the j-th eigenspace; for any matrix $P_j \in \mathbb{R}^{n\times m_j}$ whose columns are an orthonormal basis for $E_j \subset \mathbb{R}^n$, we can write $\Pi_j = P_jP_j^T$. Writing $P = [P_1 \mid \cdots \mid P_\ell]$, we have an eigendecomposition $M = P\Lambda P^T$, where $\Lambda = \mathrm{diag}(\lambda_1 \mathrm{id}_{m_1}, \ldots, \lambda_\ell \mathrm{id}_{m_l})$.*

3.3.1 Wide Regime

In the case where $n \leq d$, let us restrict ourselves to the generic subset of $\mathbb{R}^{n\times d}$ where matrices have full rank. Consider the fibre $\mathbf{q}^{-1}(\mathbf{q}(X))$ for X full rank. As $\mathbf{g}$ is a component of $\mathbf{q}$, we have

$$\mathbf{q}^{-1}(\mathbf{q}(X)) \subset \mathbf{g}^{-1}(\mathbf{g}(X)) = O(n) \cdot X. \tag{28}$$

The fibre $\mathbf{q}^{-1}(\mathbf{q}(X))$ can then be characterised by matrices UX for $U \in O(n)$, such that $\mathbf{g}_M(X) = \mathbf{g}_M(UX)$. These are matrices such that

$$(UX)^TMUX = X^TMX \iff X^T(U^TMU - M)X = 0. \tag{29}$$

Assuming X is full rank, this condition is equivalent to $U^TMU - M = 0$. That is, U are elements of the stabiliser subgroup of the $O(n)$ action on $\mathbb{R}^{n\times n}$ matrices $M \mapsto U^TMU$. Denoting this stabiliser subgroup as $O(n)_{:M}$ to distinguish it from the stabiliser subgroups of left $O(n)$ actions on matrices $X \in \mathbb{R}^{n\times d}$,

$$\begin{aligned} &\text{if} \quad \operatorname{rank} X = \min(n,d) = n, \quad \text{then} \\ \mathbf{q}(X) = \mathbf{q}(Y) \qquad &\iff \qquad Y \in O(n)_{:M} \cdot X, \quad O(n)_{:M} = \{U \in O(n),\ U^TMU = M\}. \end{aligned} \tag{30}$$

If X has rank $\min(n,d)$, the stabiliser of X under the left $O(n)$ is trivial; thus the orbit is an embedding of $O(n)_{:M}$. We characterise this subgroup with the following lemma in terms of the multiplicity of the spectrum of M.

Lemma 3.3. *Suppose the spectrum of $M \in \mathrm{Sym}_n$ has multiplicity $(m_1, \ldots, m_\ell)$. Then there is a Lie group isomorphism*

$$O(n)_{:M} \cong O(m_1) \times \cdots \times O(m_\ell). \tag{31}$$

We defer the proof to Section A.2. Having described the fibre of $\mathbf{q}$ as the orbit of a *free* group action, we can derive the following description of the fibre $\mathbf{q}^{-1}(\mathbf{q}(X))$ for generic wide matrices $\mathbb{R}^{n\times d}$.

Proposition 3.5. *Suppose $\operatorname{rank} X = \min(n,d) = n$, and the spectrum of $M \in \mathrm{Sym}_n$ has multiplicity $(m_1, \ldots, m_\ell)$. Then $\mathbf{q}^{-1}(\mathbf{q}(X))$ is an embedding of $\prod_{j=1}^{\ell} O(m_j)$, a compact manifold with 2^ℓ connected components, and dimension*

$$\dim \mathbf{q}^{-1}(\mathbf{q}(X)) = \sum_{j=1}^{\ell} \binom{m_j}{2}. \tag{32}$$

In particular, if the spectrum is simple ($m_j = 1$), then $\mathbf{q}^{-1}(\mathbf{q}(X))$ is a set of 2^n points.

Proof. Since $\operatorname{rank} X = \min(n, d) = n$, the left $O(n)$ action on X is free. As $f^{-1}(f(X))$ is the orbit of free the $O(n)$-action restricted to a closed subgroup isomorphic to $\prod_{j=1}^{\ell} O(m_j)$ (Lemma 3.3), Theorem 3.2 implies $f^{-1}(f(X))$ is diffeomorphic to $\prod_{j=1}^{\ell} O(m_j)$ as the stabiliser subgroup is trivial. Since the each $O(m_j)$ component is a manifold with dimension $\binom{m_j}{2}$, and two connected components, the dimension of the product manifold is the sum over the individual dimensions, and the connected components is the product. □

With a view towards guiding architectural design, we now show that for the $n \leq d$ regime, we cannot improve identifiability by appending more matrices in the form

$$f = (\mathbf{g}, \mathbf{g}_M, \mathbf{g}_{p_1(M)}, \ldots, \mathbf{g}_{p_B(M)}), \tag{33}$$

where $p_i(M)$ are polynomials in M. All other polynomials contain no extra information about X than the basic readout $\mathbf{q} = (\mathbf{g}, \mathbf{g}_M)$. For the choice of $M = L$ being the Laplacian, this implies incorporating or optimising Chebyshev filters of the Laplacian in the spirit of ChebNet Defferrard et al. (2016) would not improve the readout beyond L. For the choice of $M = A$ being the Laplacian, this implies having comparing 'deeper' convolutions $X^T A^k X$ is redundant; all information is retained in the graph-agnostic signal spread $X^T X$, and the one step message propagation $X^T AX$.

Proposition 3.6. *Consider $M \in \operatorname{Sym}_n$ and a matrix polynomial $\tilde{M} = p(M)$. Let $\mathbf{q} = (X^T X, X^T MX)$ and $\tilde{\mathbf{q}} = (X^T X, X^T \tilde{M} X)$. Suppose $d \leq n$ and let $\mathcal{G} \subset \mathbb{R}^{n \times d}$ be the generic subset where $\operatorname{rank} X = \min(n, d) = n$. Then for $X \in \mathcal{G}$,*

$$\mathbf{q}^{-1}(\mathbf{q}(X)) \subseteq \tilde{\mathbf{q}}^{-1}(\tilde{\mathbf{q}}(X)) \tag{34}$$

Consequently, for any set of polynomials $p_1, \ldots, p_B$, and a map $f : \mathbb{R}^{n \times d} \to \mathbb{R}^m$,

$$f(X) = (X^T X, X^T MX, X^T p_1(M)X, X^T p_2(M)X, \ldots, X^T p_B(M)X), \tag{35}$$

the fibres of f are the same as those of $\mathbf{q}$ for $X \in \mathcal{G}$:

$$f^{-1}(f(X)) = \mathbf{q}^{-1}(\mathbf{q}(X)). \tag{36}$$

Proof. If X has full rank $\min(n, d) = n$, by eq. (30),

$$\tilde{\mathbf{q}}^{-1}(\tilde{\mathbf{q}}(X)) = O(n)_{:\tilde{M}} \cdot X, \quad O(n)_{:\tilde{M}} = \{U \in O(n),\ U^T \tilde{M} U = \tilde{M}\} \tag{37}$$

Since $\tilde{M}$ is a matrix polynomial in M, $UMU^T = U$ implies $U^T \tilde{M} U = U$, and thus $O(n)_{:M} \leq O(n)_{:\tilde{M}}$. Consequently

$$\mathbf{q}^{-1}(\mathbf{q}(X)) = O(n)_{:M} \cdot X \subseteq O(n)_{:\tilde{M}} \cdot X = \tilde{\mathbf{q}}^{-1}(\tilde{\mathbf{q}}(X)). \tag{38}$$

For the second statement in the proposition, if we write $\mathbf{q}_j(X) = (X^T X, X^T p_j(M) X)$, then for X having full rank $\min(n, d) = n$, the fact that we have $\mathbf{q}^{-1}(\mathbf{q}(X)) \subseteq \mathbf{q}_j^{-1}(\mathbf{q}_j(X))$ implies

$$f^{-1}(f(X)) = \mathbf{q}^{-1}(\mathbf{q}(X)) \cap \mathbf{q}_1^{-1}(\mathbf{q}_1(X)) \cap \cdots \cap \mathbf{q}_B^{-1}(\mathbf{q}_B(X)) = \mathbf{q}^{-1}(\mathbf{q}(X)). \tag{39}$$

□

3.3.2 Tall Regime

In the regime where $d < n$, the full rank assumption that guarantees the generic fibres of $\mathbf{q}$ are orbits of the subgroup $O(n)_{:M}$ (eq. (30)) in the $n \geq d$ case fails. As such, we adopt a more elementary approach to deriving the generic fibre dimension δ in this regime. Recall Proposition 2.1: since $\mathbf{q}$ is a proper polynomial map, there is a generic subset $\mathcal{G}$, through which fibres of $\mathbf{q}$ are δ-dimensional submanifolds. Furthermore, for *any* $X \in \mathcal{G}$, the Jacobian $\mathcal{D}\mathbf{q}|_X$ attains its maximal possible rank ρ eq. (13), and $\delta = nd - \rho$.

The proof strategy thus centers on solving for the maximal possible rank to derive the generic fibre dimension

$$\delta = nd - \max_{X \in \mathbb{R}^{n \times d}} \mathcal{D}\mathbf{q}|_X. \tag{40}$$

We show that in fact in the regime $d < n$, and assuming M has a simple spectrum, we can find X such that $\mathcal{D}\mathbf{q}|_X$ is a surjection, and thus

$$\max_{X \in \mathbb{R}^{n \times d}} \mathcal{D}\mathbf{q}|_X = \dim(\operatorname{Sym}_d^2) = d(d+1). \tag{41}$$

This then implies the generic fibre dimension of $\mathbf{q}$ is $\delta = nd - \rho = d(n - d - 1)$. Furthermore, this is the *minimal possible* fibre dimension δ_f for any polynomial map $f : \mathbb{R}^{n \times d} \to \operatorname{Sym}_d^2$, as for any

$$\delta_f = nd - \max_{X \in \mathbb{R}^{n \times d}} \mathcal{D}\mathbf{q}|_X \geq nd - \dim(\operatorname{Sym}_d^2) = d(n - d - 1). \tag{42}$$

The proof of the proposition then involves finding an *ansatz* for the maximisation problem eq. (40).

Proposition 3.7. *For $d \leq n-1$, and $M \in \mathrm{Sym}_n$ having a simple spectrum, there is a generic subset $\mathcal{G} \subset \mathbb{R}^{n\times d}$ where $\mathbf{q}|_{\mathcal{G}} : \mathcal{G} \to \mathrm{Sym}_d^2$ is a submersion; thus $\rho = d(d+1)$ and the generic fibre dimension of $\mathbf{q}$ is $\delta = d(n-d-1)$. If M is not simple, then its generic fibre dimension is is at least $d(n-d-1)$.*

We defer the full proof to Section A.1.

4 Discussion

In this work, we contrast how linear and quadratic readouts have different properties in terms of identifiability. While linear readouts $\mathbf{s}_C$ have generic fibre dimension δ that increases linearly and unbounded with d, the Gram readouts $\mathbf{g}$ and $\mathbf{q} = (\mathbf{g}, \mathbf{g}_M)$ plateaus when d exceeds n (Table 1, see also Figure 1). In particular, the fibre of $\mathbf{q} = (\mathbf{g}, \mathbf{g}_M)$ consists of simply 2^n points when $n \geq d$ and the spectrum of M is simple.

As expected, the difficulty of identifiability increases with the number of vertices n. This is reflected in the linear dependency of δ on n for linear readouts $\mathbf{s}$ and $\mathbf{s}_C$. For the quadratic readout, the constraint imposed by the number of vertices is reflected by the threshold at which the generic fibre dimension plateaus. In the $n \leq d$ regime, for $\mathbf{q} = (\mathbf{g}, \mathbf{g}_M)$ under simple spectrum, its generic fibre dimension $\delta(\mathbf{q}) = d(n-d-1)$ first grows with d, peaks around $d \approx n/2$ and vanishes from $d = n-1$ on. Note that it never decreases for $\mathbf{s}_C$, $\mathbf{g}$, or even the combination of $(\mathbf{s}_C, \mathbf{g})$.

Implications for GNN Design Our theoretical results can inform how a practitioner can incorporate a Gram-type readout in a graph neural network architecture. Our results imply that in order to potentially achieve local identifiability of generic vertex features with a quadratic readout involving $\mathbf{q}$, we would need at least the number of vertex features to exceed the number of vertices. Incorporating this number of vertex channels in a GNN may be an unreasonable one to pay if the graphs involved are large.

Assumptions and limitations. The perfect local identifiability of $\mathbf{q}$ in the $d \geq n$ limit rests on the matrix M having simple spectrum. This holds with high probability for adjacency matrices of Erdős–Rényi graphs (Tao & Vu, 2017; Luh & Vu, 2020) and Laplacians of dense ones (Christoffersen et al., 2025), but a graph with an automorphism of order greater than two, such as a cycle, never has a simple adjacency spectrum (Mowshowitz, 1969; Cvetković et al., 1980). Repeated eigenvalues are reportedly rare on some benchmarks (Wang & Zhang, 2022) but common in real-world graphs (Lu et al., 2024). With repeated eigenvalues, the generic fibre dimension $\delta(\mathbf{q})$ grows with the multiplicities for $d \geq n$ (Proposition 3.5); for $d < n$ we only have a lower bound.

Our results also concern generic (full rank) signals, which trained GNNs need not produce: oversmoothing, e.g., makes vertex features nearly identical (Cai & Wang, 2020; Rusch et al., 2023), so signals approach rank one. It may also be impossible to learn full rank $= n$ vertex features if the input vertex features on datasets are fundamentally low rank, such as atom types of molecular graphs.

AI USE STATEMENT

(This section is **required** and does not count toward the page limit.)

In this work, we used generative AI tools to help develop theoretical models or conceptual frameworks, assist in the writing of proofs, formulate mathematical claims, provide critical ingredients for proving mathematical claims, propose or refine hypotheses, design or provide feedback on research methodology or experiments, and implement methods. We have not used generative AI tools to generate synthetic data sets, clean and reformat dataset, support qualitative and thematic data analysis, and interpret results. Using AI to to assist with translation is not applicable to this work. Additionally, we used generative AI tools to create or edit software code, draft parts of the research paper, summarize or analyse existing literature, and identify relevant literature. We have reviewed all AI-assisted work, including all mathematical proofs, which are verified for correctness by the authors. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Tal Amir, Steven J. Gortler, Ilai Avni, Ravina Ravina, and Nadav Dym. Neural injective functions for multisets, measures and graphs via a finite witness theorem. In *Advances in Neural Information Processing Systems*, volume 36, pp. 42516–42551, 2023.

Jinheon Baek, Minki Kang, and Sung Ju Hwang. Accurate learning of graph representations with graph multiset pooling. In *International Conference on Learning Representations*, 2021.

Saugata Basu, Richard Pollack, and Marie-Francoise Roy. *Algorithms in real algebraic geometry*. Algorithms and Computation in Mathematics. Springer, Berlin, Germany, 2 edition, July 2006. ISBN 978-3-540-33099-8.

Jacek Bochnak, Michel Coste, and Marie-Francoise Roy. *Real Algebraic Geometry*. Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics. Springer, Berlin, Germany, 1998 edition, August 1998. ISBN 978-3-540-64663-1.

Joachim Bona-Pellissier, François Malgouyres, and François Bachoc. Local identifiability of deep ReLU neural networks: the theory. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

David Buterez, Jon Paul Janet, Steven J. Kiddle, Dino Oglic, and Pietro Liò. Graph neural networks with adaptive readouts. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

Chen Cai and Yusu Wang. A note on over-smoothing for graph neural networks. *arXiv preprint arXiv:2006.13318*, 2020. ICML 2020 Workshop on Graph Representation Learning and Beyond.

Kaixuan Chen, Jie Song, Shunyu Liu, Na Yu, Zunlei Feng, Gengshi Han, and Mingli Song. Distribution knowledge embedding for graph pooling. *IEEE Transactions on Knowledge and Data Engineering*, 2022.

Nicholas J. Christoffersen, Kyle Luh, Hoi H. Nguyen, and Jingheng Wang. Eigenvalue gaps of the Laplacian of random graphs. *arXiv preprint arXiv:2501.00234*, 2025.

Gabriele Corso, Luca Cavalleri, Dominique Beaini, Pietro Liò, and Petar Veličković. Principal neighbourhood aggregation for graph nets. In *Advances in Neural Information Processing Systems*, volume 33, pp. 13260–13271, 2020.

Dragoš M. Cvetković, Michael Doob, and Horst Sachs. *Spectra of Graphs: Theory and Application*. Academic Press, New York, 1980.

Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in Neural Information Processing Systems*, volume 29, 2016.

David K. Duvenaud, Dougal Maclaurin, Jorge Iparraguirre, Rafael Bombarell, Timothy Hirzel, Alán Aspuru-Guzik, and Ryan P. Adams. Convolutional networks on graphs for learning molecular fingerprints. In *Advances in Neural Information Processing Systems*, volume 28, 2015.

Nadav Dym and Steven J. Gortler. Low-dimensional invariant embeddings for universal geometric learning. *Foundations of Computational Mathematics*, 25:375–415, 2025. Published online 2024.

Alan Edelman, Tomás A Arias, and Steven T Smith. The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.*, 20(2):303–353, June 1998.

Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1263–1272, 2017.

Allen Hatcher. *Algebraic Topology*. Cambridge University Press, Cambridge, England, December 2001.

Nathan W. Henry, Giovanni Luca Marchetti, and Kathlén Kohn. Geometry of lightning self-attention: Identifiability and dimension. In *International Conference on Learning Representations*, 2025a.

Nathan W. Henry, Giovanni Luca Marchetti, and Kathlén Kohn. Geometry of lightning self-attention: Identifiability and dimension. In *The Thirteenth International Conference on Learning Representations*, 2025b. URL `https://openreview.net/forum?id=XtY3xYQWcW`.

Roger A. Horn and Charles R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, 2nd ed. edition, 2013. ISBN 1-139-79342-X.

Joe Kileel, Matthew Trager, and Joan Bruna. On the expressive power of deep polynomial neural networks. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

John M Lee. *Introduction to Smooth Manifolds*. Graduate Texts in Mathematics. Springer, New York, NY, 2 edition, December 2014.

Juho Lee, Yoonho Lee, Jungtaek Kim, Adam Kosiorek, Seungjin Choi, and Yee Whye Teh. Set transformer: A framework for attention-based permutation-invariant neural networks. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, 2019.

Kangkang Lu, Yanhua Yu, Hao Fei, Xuan Li, Zixuan Yang, Zirui Guo, Meiyu Liang, Mengran Yin, and Tat-Seng Chua. Improving expressive power of spectral graph neural networks with eigenvalue correction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 14158–14166, 2024. doi: 10.1609/aaai.v38i13.29326.

Kyle Luh and Van Vu. Sparse random matrices have simple spectrum. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 56(4):2307–2328, 2020.

Giovanni Luca Marchetti, Vahid Shahverdi, Stefano Mereta, Matthew Trager, and Kathlén Kohn. Position: Algebra unveils deep learning – an invitation to neuroalgebraic geometry. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 81759–81773, 2025.

B S Mityagin. The zero set of a real analytic function. *Mathematical Notes*, 107(3-4):529–530, March 2020. URL `https://link.springer.com/article/10.1134/S0001434620030189`.

Abbe Mowshowitz. The group of a graph whose adjacency matrix has all distinct eigenvalues. In Frank Harary (ed.), *Proof Techniques in Graph Theory*. Academic Press, 1969.

Ryan L. Murphy, Balasubramaniam Srinivasan, Vinayak Rao, and Bruno Ribeiro. Janossy pooling: Learning deep permutation-invariant functions for variable-size inputs. In *International Conference on Learning Representations*, 2019.

Masashi Okamoto. Distinctness of the eigenvalues of a quadratic form in a multivariate sample. *Ann. Stat.*, 1(4): 763–765, July 1973.

T. Konstantin Rusch, Michael M. Bronstein, and Siddhartha Mishra. A survey on oversmoothing in graph neural networks. *arXiv preprint arXiv:2303.10993*, 2023.

Vahid Shahverdi, Giovanni Luca Marchetti, and Kathlén Kohn. Learning on a razor's edge: Identifiability and singularity of polynomial neural networks. In *International Conference on Learning Representations*, 2026.

Pierre Stock and Rémi Gribonval. An embedding of ReLU networks and an analysis of their identifiability. *Constructive Approximation*, 57(2):853–899, 2023. doi: 10.1007/s00365-022-09578-1.

Héctor J. Sussmann. Uniqueness of the weights for minimal feedforward nets with a given input-output map. *Neural Networks*, 5(4):589–593, 1992. doi: 10.1016/S0893-6080(05)80037-1.

Mouad Talhi, Arne Wolf, and Anthea Monod. Breaking symmetry bottlenecks in GNN readouts. *arXiv preprint arXiv:2602.05950*, 2026.

Terence Tao and Van Vu. Random matrices have simple spectrum. *Combinatorica*, 37(3):539–553, 2017.

Konstantin Usevich, Ricardo Augusto Borsoi, Clara Dérand, and Marianne Clausel. Identifiability of deep polynomial neural networks. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Oriol Vinyals, Samy Bengio, and Manjunath Kudlur. Order matters: Sequence to sequence for sets. In *International Conference on Learning Representations*, 2016.

Edward Wagstaff, Fabian Fuchs, Martin Engelcke, Ingmar Posner, and Michael A. Osborne. On the limitations of representing functions on sets. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 6487–6494, 2019.

Edward Wagstaff, Fabian B. Fuchs, Martin Engelcke, Michael A. Osborne, and Ingmar Posner. Universal approximation of functions on sets. *Journal of Machine Learning Research*, 23:1–56, 2022.

Xiyuan Wang and Muhan Zhang. How powerful are spectral graph neural networks. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 23341–23362, 2022.

Zhengyang Wang and Shuiwang Ji. Second-order pooling for graph neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(6):6870–6880, 2023a. doi: 10.1109/TPAMI.2020.2999032. Published online 2020.

Zhengyang Wang and Shuiwang Ji. Second-order pooling for graph neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(6):6870–6880, 2023b. doi: 10.1109/TPAMI.2020.2999032. URL https://doi.org/10.1109/TPAMI.2020.2999032.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *International Conference on Learning Representations*, 2019.

Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabás Póczos, Ruslan Salakhutdinov, and Alexander J. Smola. Deep sets. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Muhan Zhang, Zhicheng Cui, Marion Neumann, and Yixin Chen. An end-to-end deep learning architecture for graph classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

A Mathematical Proofs

A.1 Polynomials and Genericity

Lemma A.1. *Any polynomial map $f = (\mathbf{g}, p) : \mathbb{R}^{n\times d} \to \mathbb{R}^m$ containing the gram map $\mathbf{g}$ as a component is a proper map.*

Proof. Consider a compact subset A in the range of f. Since as the trace of the Gram matrix $X^T X$ is the Frobenius norm of X, A being bounded implies the Gram matrix and by extension X being bounded as well. Hence the the preimage of the compact subset is not only closed by continuity, but also bounded. Thus f is proper. □

Proposition 2.1. *Let $f : \mathbb{R}^a \to \mathbb{R}^b$ be a* proper *polynomial map. Then the generic fibre dimension δ is defined for f. That is, there is a generic subset $\mathcal{G} \subset \mathbb{R}^a$, on which $f^{-1}(f(x))$ are δ-dimensional manifolds. Furthermore, for $x \in \mathcal{G}$, the Jacobian of the polynomial $\mathcal{D}f$ attains its maximal rank*

$$x \in \mathcal{G} \implies \operatorname{rank} \mathcal{D}f|_x = \max_{y\in\mathbb{R}^a} \operatorname{rank} \mathcal{D}f|_y =: \rho, \tag{13}$$

and the generic fibre dimension is given by $\delta = a - \rho$.

Proof. As $\mathcal{G}$ is open, Lemma A.6 fixes $\delta = a - \rho$. The existence of such an open dense subset $\mathcal{G}$ is guaranteed by Φ if f is proper in Lemmas A.3 and A.4. □

We build up the proof below. Consider a polynomial map $f : \mathbb{R}^a \to \mathbb{R}^b$ and the equivalence relation

$$x \sim_f y \iff f(x) = f(y), \qquad [x] = f^{-1}(f(x)). \tag{43}$$

We call $[x] \subset \mathbb{R}^a$ the fibre of f through x. While $[x]$ may not be a constant dimension manifold for all x, there is an open dense subset of $\Phi \subset \mathbb{R}^a$ that is closed under $\sim_f$, and $[x]$ is a δ-dimensional smooth manifold for any $x \in \Phi$. We call such fibres $[x] \subset \Phi$ a generic fibre of the map, and δ the generic fibre dimension of the map f. We detail this construction here.

Let us denote $\mathcal{D}f : T\mathbb{R}^a \to T\mathbb{R}^b$ as the Jacobian of the polynomial map f. At each point $x \in \mathbb{R}^a$, we have a linear transformation $\mathcal{D}f|_x : T_x\mathbb{R}^a \to T_{f(x)}\mathbb{R}^b$, and we can obtain a function $\mathrm{rk} : \mathbb{R}^a \to \{0, \dots, \min(a, b)\}$ which assigns to x the rank of the Jacobian $\mathcal{D}f|_x$:

$$\mathrm{rk}(x) := \operatorname{rank} \mathcal{D}f|_x, \quad \rho := \max_{x\in\mathbb{R}^a} \mathrm{rk}(x) \le \min(a, b). \tag{44}$$

Because the rank of a matrix is a *lower semi-continuous* map on the entries of the matrix, and $\mathcal{D}f|_\bullet : \mathbb{R}^a \to \mathbb{R}^{b\times a}$ is continuous, rk is also lower semi-continuous and sublevel sets are closed subsets of $\mathbb{R}^a$. Let $F_r = \{\mathrm{rk} \le r\}$; then we have a filtration of $\mathbb{R}^a$ by closed sublevel sets of the Jacobian rank:

$$\emptyset = F_{-1} \subseteq F_0 \subseteq \cdots \subseteq F_{\rho-1} \subseteq F_\rho = \mathbb{R}^a, \qquad \Sigma := F_\rho \setminus F_{\rho-1}. \tag{45}$$

As the rank of a matrix is the largest r such that there exists a non-zero order-r minor (the determinant of a $r \times r$ submatrix), $F_r = \{\mathrm{rk} \le r\}$ is the subset whose order $(r + 1)$ minors of the Jacobian all vanish. Thus any F_r is the zero set of a polynomial $\mu_r(x)$, which is the sum of squares of all order $(r + 1)$ minors of $\mathcal{D}f|_x$. We focus on Σ, the open subset of points where the Jacobian attains maximal rank ρ.

Lemma A.2. *$F_{\rho-1}$ is nowhere dense; thus $\Sigma = F_\rho \setminus F_{\rho-1}$ is open and dense (generic).*

Proof. If $\rho = 0$ (i.e. $\mathcal{D}f \equiv 0$ and f is just a constant) then the statement is trivial. Assume $\rho > 0$. As $\mathcal{D}f|_x$ attains rank ρ at some x, not all order ρ minors of the Jacobian are identically zero polynomials. Thus $\mu_{\rho-1} \not\equiv 0$, and $F_{\rho-1} = \mu_{\rho-1}^{-1}(0)$ is closed and has non-empty interior (Example 2.1). Consequently $F_{\rho-1}$ is nowhere dense and its complement $F_\rho \setminus F_{\rho-1}$ is open and dense. □

The fibre $[x]$ through a point $x \in \Sigma$ may not be contained in Σ: the fibre over $f(x)$ may also contain points in $F_{\rho-1}$ whose Jacobian is rank deficient. Thus, we restrict our consideration to points $x \in \Sigma$ where the Jacobian $\mathcal{D}F$ on the whole fibre $[x]$ has constant rank ρ:

$$\Phi = F_\rho \setminus f^{-1}(f(F_{\rho-1})) \subseteq \Sigma. \tag{46}$$

The following two lemmas establish that we can restrict to Φ from Σ without losing genericity. Furthermore, we gain the benefit of fibres of f on Φ being actual manifolds of the same dimension.

Lemma A.3. *If f is proper, then Φ is open.*

Proof. If f is proper, then f is a closed map as it is also a continuous map between Euclidean spaces. Since $F_{\rho-1}$ is closed, $f(F_{\rho-1})$ is also closed, and thus $\Phi = \mathbb{R}^a \setminus f^{-1}(f(F_{\rho-1}))$ is open. □

Lemma A.4. *Φ is dense in $\mathbb{R}^a$.*

Proof. To show that Φ is dense in $\mathbb{R}^a$, we need to show that for any non-empty open set U_0 in $\mathbb{R}^a$, we can always find an element $x \in \Phi \cap U_0$.

As Σ is open and dense, $\Sigma \cap U_0$ is open and non-empty. We show that while not all $x_0 \in \Sigma \cap U_0$ may be in Φ, we may find $\tilde{x}$ in some neighbourhood $U \subset \Sigma \cap U_0$ of x_0, such that $\tilde{x} \in \Phi$. As $\Sigma \cap U_0 \subset \Sigma$, the Jacobian of f on $\Sigma \cap U_0$ has constant rank ρ, and thus $f|_{\Sigma \cap U_0} : \Sigma \cap U_0 \to \mathbb{R}^b$ is a smooth map with constant rank. We can apply the rank theorem (Lee, 2014, Theorem 4.12) to obtain a local coordinate representation: for any $x_0 \in \Sigma \cap U_0$, we can find an open neighbourhood $U \subset \Sigma \cap U_0$, an open neighbourhood V of $f(x_0)$ such that $f(U) \subset V$, as well as a pair of smooth charts $\varphi : (U, x_0) \to (\mathbb{R}^a, 0)$ and $\varphi : (V, f(x_0)) \to (\mathbb{R}^b, 0)$, such that we obtain a local coordinate representation

$$\psi f \varphi^{-1} = \tilde{f}, \qquad \tilde{f}(x^1, \ldots, x^a) = (x^1, \ldots, x^\rho, 0, \ldots, 0). \tag{47}$$

Let us focus on the image of $\psi f = \tilde{f}\varphi$, which is $\mathbb{R}^\rho \subseteq \mathbb{R}^b$. Let $\jmath : \mathbb{R}^b \twoheadrightarrow \mathbb{R}^\rho$ and $\imath : \mathbb{R}^\rho \hookrightarrow \mathbb{R}^b$ be the coordinate projection and inclusion maps of $\mathbb{R}^\rho = \psi f(U)$ as a subspace of $\mathbb{R}^\rho$. As ψ is a diffeomorphism, the map $\bar{\psi} := \jmath \circ \psi = V \twoheadrightarrow \mathbb{R}^\rho$ is a submersion and surjection of V into $\mathbb{R}^\rho$.

Consider $f^{-1}(f(U)) \subset \mathbb{R}^a$, and the smooth map $f^{-1}(f(U)) \xrightarrow{f} V \xrightarrow{\bar{\psi}} \mathbb{R}^\rho$. Let us compute the rank of $\mathcal{D}(\bar{\psi} f)$; for $x \in f^{-1}(f(U))$,

$$\operatorname{rank} \mathcal{D}(\bar{\psi} \circ f)|_x = \operatorname{rank} \mathcal{D}(\imath \circ \bar{\psi} \circ f)|_x = \operatorname{rank} \mathcal{D}(\psi \circ f)|_x = \operatorname{rank} \mathcal{D}f|_x. \tag{48}$$

where in the first equality we noted that $\imath : \mathbb{R}^\rho \hookrightarrow \mathbb{R}^b$ is an inclusion of a linear subspace that does not change rank; the second equality uses the fact that $\imath \circ \bar{\psi}|_{f(U)} = \imath \circ \jmath \circ \psi|_{f(U)} = \psi|_{f(U)}$; and the third equality is then due to ψ being a diffeomorphism.

Thus, points $x \in f^{-1}(f(U)) \cap \Sigma$ where $\operatorname{rank} \mathcal{D}f|_x = \rho$ are precisely the regular points of the map $\bar{\psi} \circ f$, while points $x \in f^{-1}(f(U)) \cap F_{\rho-1}$ where $\operatorname{rank} \mathcal{D}f|_x < \rho$ are the critical points of $\bar{\psi} f$. Since $\Sigma \sqcup F_{\rho-1} = \mathbb{R}^a$, these two cases account for all the points in $f^{-1}(f(U))$. By Sard's theorem (Lee, 2014, Theorem 6.10), the image of $f^{-1}(f(U)) \cap F_{\rho-1}$ under $\bar{\psi} \circ f$ is a measure zero subset of $\mathbb{R}^\rho$. So we can always pick some $y \in \mathbb{R}^\rho$ off this set. As $\mathbb{R}^\rho$ is the image of $\bar{\psi} \circ f$, there must be some $x \in U$ such that $\bar{\psi} \circ f(x) = y$. Since y is not in the image of the measure zero subset $f^{-1}(f(U)) \cap F_{\rho-1}$,the fibre $[x] = f^{-1}f(x) \subset f^{-1}(f(U))$ cannot intersect $F_{\rho-1}$. Thus, we have identified some $x \in U \subset U_0 \cap \Sigma$, such that $x \in \Phi$. In other words, we have found for any open non-empty subset $U_0 \in \mathbb{R}^a$, an element in Φ. Thus Φ is dense in $\mathbb{R}^a$. □

Lemma A.5. *For any $x \in \Phi$, the fibre $[x]$ is a codimension ρ submanifold of $\mathbb{R}^a$.*

Proof. Since Σ is open in $\mathbb{R}^a$ as it is a dimension-a smooth manifold. The restriction $f|_\Sigma : \Sigma \to \mathbb{R}^b$ then is a smooth map with constant Jacobian rank ρ by construction. By the constant-rank level set theorem (Lee, 2014, Theorem 5.12), the fibre $f|_\Sigma^{-1}(f(x)) = f^{-1}(f(x)) \cap \Sigma = [x] \cap \Sigma$ are embedded submanifolds of Σ of codimension ρ. If $x \in \Phi$, we have $[x] \subset \Sigma$, and the inclusion $[x] \hookrightarrow \Sigma \hookrightarrow \mathbb{R}^a$ is an embedding of $[x]$ as a codimension ρ submanifold of $\mathbb{R}^a$. □

Lemmas A.4 and A.5 together imply if any generic set $\mathcal{G}$ enjoys the property that all fibres of f through $\mathcal{G}$ are manifolds of constant dimension, then they must have codimension ρ.

Lemma A.6. *For f a polynomial map, and $U \subset \mathbb{R}^a$ an open subset, suppose for all $x \in U$, the fibre $[x]$ are submanifold with the same dimension δ. Then the dimension of such manifolds is fixed by the maximal rank eq.* (13) *$\delta = a - \rho$.*

Proof. Since Φ is dense (Lemma A.4), U would contain a point $\tilde{x} \in \Phi$. By Lemma A.5, the submanifold $[\tilde{x}]$ has codimension ρ. Since we have assumed all fibres $[x]$ have dimension δ, the codimension $a - \delta$ must be equal to ρ. □

Lemma 3.2. *Let $C \in \mathbb{R}^{n \times c}$ be some fixed, full rank matrix ($c \leq n$). There is a generic subset $\mathcal{G} \subseteq \mathbb{R}^{n \times d}$ such that for $X \in \mathcal{G}$, the matrix $[X \mid C]$ attains full rank $\min(n, d + c)$.*

Proof. Since we have assumed C is full rank and $c \leq n$, if $c = n$, then $[X \mid C]$ attains full rank n for any $X \in \mathbb{R}^{n\times d}$. For $c < n$, let $P \in \mathbb{R}^{n\times n}$ be a projection onto the orthogonal complement of the column space $\mathrm{col}(C)$. The matrix $[PX \mid C]$ amounts to performing column operations on X that subtracts suitable linear combinations of columns in C from X, and has the same rank as $[X \mid C]$. As PX and C are now orthogonal $\mathrm{rank}[PX \mid C] = \mathrm{rank}\, PX + c$. Writing $P = UU^T$ for $U \in \mathbb{R}^{n\times(n-c)}$ whose column vectors are an orthonormal basis for $\mathrm{col}(C)$, as U is an injective linear map, $\mathrm{rank}\, PX = \mathrm{rank}\, U^T X$. Note that as $\mathrm{rank}[X \mid C] = c + \mathrm{rank}\, U^T X$, our original matrix $[X \mid C]$ attains maximal (full) rank iff $\mathrm{rank}\, U^T X$ attains maximal (full) rank.

Now $U^T X \in \mathbb{R}^{(n-c)\times d}$ and by construction, the map $\Pi : X \mapsto U^T X \in \mathbb{R}^{(n-c)\times d}$ is a linear surjection. Let us consider the open dense subset S of full rank matrices in $\mathbb{R}^{(n-c)\times d}$. As Π is an open, continuous, and surjective map, the subset of matrices $X \in \Pi^{-1}(S) \subset \mathbb{R}^{n\times d}$ such that $U^T X \in S$ (i.e. full rank) is also open and dense. Thus we have identified a generic subset $\mathcal{G} = \Pi^{-1}(S)$ where the matrix $[X \mid C]$ attains full rank $\min(n, d + c)$. □

Proposition 3.7. *For $d \leq n - 1$, and $M \in \mathrm{Sym}_n$ having a simple spectrum, there is a generic subset $\mathcal{G} \subset \mathbb{R}^{n\times d}$ where $\mathbf{q}|_{\mathcal{G}} : \mathcal{G} \to \mathrm{Sym}^2_d$ is a submersion; thus $\rho = d(d+1)$ and the generic fibre dimension of $\mathbf{q}$ is $\delta = d(n - d - 1)$. If M is not simple, then its generic fibre dimension is is at least $d(n - d - 1)$.*

Proof. In this proof let J_X be a shorthand for the Jacobian $\mathcal{D}\mathbf{q}|_X : T_X\mathbb{R}^{n\times d} \to T_{\mathbf{q}(X)}\mathrm{Sym}^2_d$. As a shorthand also let $V = T_X\mathbb{R}^{n\times d} \cong \mathbb{R}^{n\times d}$, and W be the $d(d + 1)/2$-dimensional vector space Sym_d (thus we can write $T_{\mathbf{q}(X)}\mathrm{Sym}^2_d \cong \mathrm{Sym}^2_d = W^2$). With these shorthands, we write the Jacobian as a linear map $J_X : V \to W^2$.

Let us endow both V and W with the (Frobenius) inner product: for $H_1, H_2 \in V$, it is given by $\langle H_1, H_2\rangle_V = \mathrm{tr}(H_1^T H_2)$; for $(R_1, S_1), (R_2, S_2) \in W^2$, the Frobenius inner product is given by

$$2\langle(R_1, S_1), (R_2, S_2)\rangle_W = \mathrm{tr}(R_1^T R_2) + \mathrm{tr}(S_1^T S_2) = \mathrm{tr}(R_1 R_2) + \mathrm{tr}(S_1 S_2). \tag{49}$$

where we recall R_i, S_i are symmetric matrices in the last equality. Given inner products on V and W^2, recall the *adjoint* of a linear map, such as the Jacobian J, is a linear map $J^*_X : W^2 \to V$, such that for any $H \in V$, and $(R, S) \in W^2$,

$$\langle(R, S), J_X H\rangle_{W^2} = \langle J^*_X(R, S), H\rangle_V. \tag{50}$$

Also recall the fact that the orthogonal complement of the map J_X is related to the adjoint by $(\mathrm{im}J)^\perp = \ker J^*_X$. Thus the map J_X is surjective iff J^*_X is injective.

We show that there is some $X \in \mathbb{R}^{n\times d}$ such that its Jacobian is surjective via showing that its adjoint is injective. We compute the adjoint explicitly. Given $\mathbf{q}(X) = (X^T X, X^T M X)$ and M is symmetric, the Jacobian acting on $H \in V$ is given by

$$J_X H = (X^T H + H^T X, X^T M H + H^T M X). \tag{51}$$

Computing the inner product between an arbitrary element $(R, S) \in W^2$ with the $J_X H$ yields

$$2\langle(R, S), J_X H\rangle_{W^2} = \mathrm{tr}(R(X^T H + H^T X, X^T)) + \mathrm{tr}(S(MH + H^T M X)). \tag{52}$$

We can use the invariance of trace under transpose and cyclic permutations to write

$$\mathrm{tr}(R(X^T H + H^T X)) = \mathrm{tr}(RX^T H) + \mathrm{tr}(RH^T X) = \mathrm{tr}(RX^T H) + \mathrm{tr}(X^T H R^T) = 2\mathrm{tr}(RX^T H). \tag{53}$$

Performing an analogous simplification on the second term the yields

$$\langle J^*_X(R, S), H\rangle_V = \langle(R, S), J_X H\rangle_{W^2} = \mathrm{tr}(RX^T H) + \mathrm{tr}(SX^T M H) = \langle XR + MXS, H\rangle_V.$$

We can thus read off the adjoint as $J^*_X(R, S) = XR + MXS \in V = \mathbb{R}^{n\times d}$. The adjoint is injective iff

$$XR + MXS = 0 \implies R = S = 0. \tag{54}$$

To simplify the expression on the left hand side, let's write the symmetric matrix $M \in \mathrm{Sym}_n$ in terms of its eigendecomposition $M = P\Lambda P^T$, and write $P^T X = Y^T$. As $P \in O(n)$, the matrix Y fully specifies X, and

$$XR + MXS = 0 \iff P^T XR + P^T MXS = Y^T R + \Lambda Y^T S = 0 \iff RY + SY\Lambda = 0. \tag{55}$$

Let us choose an ansatz for $Y \in \mathbb{R}^{d\times n}$, such that the expression above implies $R = S = 0$ (this is the same as choosing an ansatz for X). This ansatz is chosen to be

$$\hat{Y} = [\mathrm{id}_d \mid \mathbf{1} \mid 0] \tag{56}$$

where $\mathbf{1} \in \mathbb{R}^d$ is the vector of ones. Denoting $\Lambda_d \in \mathbb{R}^{d\times d}$ to be the truncation of Λ to the upper diagonal $d \times d$ block, and $\lambda_{d+1} = \Lambda_{dd}$,

$$R\hat{Y} + S\hat{Y}\Lambda = [R \mid R\mathbf{1} \mid 0] + S[\Lambda_d \mid \lambda_{d+1}\mathbf{1} \mid 0] = [R + S\Lambda_d \mid (R + \lambda_{d+1}S)\mathbf{1} \mid 0]. \tag{57}$$

Thus

$$R\hat{Y} + S\hat{Y}\Lambda = 0 \iff R + S\Lambda_d = 0 \quad \& \quad (R + \lambda_{d+1}S)\mathbb{1} = 0. \tag{58}$$

For the first term, since R, S are symmetric,

$$R + S\Lambda_d = 0 \iff R + \Lambda_d S = 0 \tag{59}$$

which implies $\Lambda_d S - S\Lambda_d = 0$. Per entry, as S is symmetric, this is equivalent to $(\lambda_i - \lambda_j)S_{ij} = 0$. For $i \neq j$, since the spectrum of M is simple $\lambda_i \neq \lambda_j$ and thus $S_{ij} = 0$. Thus S is diagonal. As $R = -S\Lambda_d$, R is also diagonal as well. Now consider the other condition $(R + \lambda_{d+1}S)\mathbb{1} = 0$. Since multiplying a diagonal matrix with $\mathbb{1}$ extracts the diagonal, this condition is equivalent to $R_{ii} + \lambda_{d+1}S_{ii} = 0$ for $i \leq d$. However, $R + \Lambda_d S = 0$ also implies $R_{ii} + \lambda_i S_{ii} = 0$ for all $i \leq d$. Substituting the former into the latter, we obtain $(\lambda_{d+1} - \lambda_i)S_{ii} = 0$; as the spectrum is simple then, this implies $S_{ii} = 0$. Consequently we also have $R_{ii} = -\lambda_i S_{ii} = 0$. Thus, we have shown that with the ansatz $\hat{Y}$

$$R\hat{Y} + S\hat{Y} = 0 \implies R = S = 0. \tag{60}$$

This is equivalent to taking the ansatz $\hat{X} = P\hat{Y}^T \in \mathbb{R}^{n\times d}$, and having

$$J^*_{\hat{X}}(R, S) = \hat{X}R + M\hat{X}S = 0 \implies R = S = 0. \tag{61}$$

Thus, we have found an element $\hat{X} \in \mathbb{R}^{n\times d}$ where the adjoint of the Jacobian is injective, or equivalently, the Jacobian itself is surjective, i.e. $\operatorname{rank} J_{\hat{X}} = d(d+1)$. This implies $\rho = \max_{X\in\mathbb{R}^{n\times d}} J_X = d(d+1)$, and $\delta = nd - \rho = d(n-d-1)$. □

A.2 Group actions

Theorem 3.2. *Consider a compact Lie group G acting smoothly on a manifold M. Then for every $x \in M$, the space of cosets G/G_x of the stabiliser subgroup G_x is a compact, smooth manifold, and there is an embedding of $F : G/G_x \hookrightarrow M$ given by $gG_x \mapsto g \cdot x$ onto the orbit $G \cdot x$.*

Proof. Since θ_x is continuous the stabiliser $G_x = \theta_x^{-1}(x)$ is a closed subgroup in G. Applying (Lee, 2014, Thm. 21.17), the coset space G/G_x is a smooth manifold, such that the quotient $\pi : G \to G/G_x$ is a smooth submersion. Because θ_x is also constant on fibres of π, there is a unique smooth map $F : G/G_x \to M$ such that $F \circ \pi = \theta_x$ (Lee, 2014, Theorem 4.30). In other words, $F(gG_x) = \theta_x(g) = g \cdot x$.

Since G/G_x is compact (being in the continuous image of compact G), we can show that F is an embedding by proving that it is an injective smooth immersion (Lee, 2014, Proposition 4.22), . We can verify that F is injective follows from its definition:

$$F(hG_x) = F(gG_x) \iff h \cdot x = \theta_x(h) = \theta_x(g) = g \cdot x \iff h^{-1}g \in G_x \iff hG_x = gG_x. \tag{62}$$

Given F is injective, to show that it is an immersion, it suffices to show that F is equivariant with respect to the G-action on G/G_x by the equivariant rank theorem (Lee, 2014, Theorem 7.25):

$$F(h \cdot gG_x) = F((hg)G_x) = (hg) \cdot x = h \cdot (g \cdot x) = h \cdot F(gG_x). \tag{63}$$

As a smooth injective immersion of a compact manifold G/G_x into M, the map F is an embedding. □

Lemma A.7. *Let $W \subseteq V \subseteq \mathbb{R}^n$ be subspaces of dimension $0 \leq s \leq r \leq n$, and consider pointwise stabilisers $O(n)_V \leq O(n)_W$. Then the coset space $O(n)_W/O(n)_V$ is diffeomorphic to $O(n-s)/O(n-r)$, the Stiefel manifold $\mathcal{V}_r(\mathbb{R}^{n-s})$ with dimension*

$$\dim O(n-s)/O(n-r) = \dim O(n-s) - \dim O(n-r) = \binom{n-s}{2} - \binom{n-r}{2}. \tag{64}$$

Proof. Let $E_j = \operatorname{span}\{e_1, \ldots, _j\}$ where e_i are the standard coordinate unit vectors. For any linear subspaces $W \subseteq V \subseteq \mathbb{R}^n$, there is some $Q \in O(n)$ such that $x \mapsto Qx$ sends W to E_s and V to E_r. This also defines an $O(n)$ Lie group inner automorphism $\alpha : U \mapsto QUQ^T$ which sends the pointwise stabilisers $O(n)_W$ to $O(n)_{E_s}$, and $O(n)_V$ to $O(n)_{E_r} \leq O(n)_{E_s}$, since

$$U \in O(n)_W \iff QUQ^T \in O(n)_{E_s}. \tag{65}$$

Recalling $s \leq r \leq n$, the pointwise stabilisers $O(n)_{E_s}$ are respectively

$$O(n)_{E_r} = \begin{pmatrix} \mathrm{id}_r & \\ & O(n-r) \end{pmatrix} \leq \begin{pmatrix} \mathrm{id}_s & \\ & O(n-s) \end{pmatrix} = O(n)_{E_s}. \tag{66}$$

As $\alpha^{-1} : O(n) \xrightarrow{\cong} O(n)$, such that $\alpha^{-1}(O(n)_{E_s}) = O(n)_W$ and $\alpha^{-1}(O(n)_{E_r}) = O(n)_V$, where $O(n)_{E_r} \leq O(n)_{E_s}$ is a closed subgroup, the map α^{-1} restricted to $O(n)_{E_s}$ is a Lie group diffeomorphism descends to an equivariant diffeomorphism on the quotient manifold $\beta : O(n-s)/O(n-r) := O(n)_{E_s}/O(n)_{E_r} \cong O(n)_W/O(n)_V$.

For general $p \geq q$, the coset space $O(p)/O(q)$ is diffeomorphic to the Stiefel manifold $\mathcal{V}_q(\mathbb{R}^p)$ (see (Edelman et al., 1998, §2.3.1) and (Hatcher, 2001, Ch.3, p.302)) As $O(p)$ is a closed subgroup of a Lie group $O(n)$, we can apply (Lee, 2014, Thm 21.17) to deduce that $O(p)/O(q)$ is a manifold of dimension $\dim O(p) - \dim O(q) = \binom{p}{2} - \binom{q}{2}$. □

Lemma 3.3. *Suppose the spectrum of $M \in \mathrm{Sym}_n$ has multiplicity $(m_1, \ldots, m_\ell)$. Then there is a Lie group isomorphism*

$$O(n)_{:M} \cong O(m_1) \times \cdots \times O(m_\ell). \tag{31}$$

Proof. Let $M = P\Lambda P^T$ be an eigendecomposition of M, where $P \in O(n)$, and

$$\Lambda = \begin{pmatrix} \lambda_1 \mathrm{id}_{m_1} & & \\ & \ddots & \\ & & \lambda_\ell \mathrm{id}_{m_l} \end{pmatrix}. \tag{67}$$

The stabiliser subgroup of M is then

$$U \in O(n) \text{ s.t. } U^T M U = M \iff Q^T \Lambda Q = \Lambda,\ Q = P^T U P,\ U \in O(n). \tag{68}$$

However $U \mapsto P^T U P$ is an inner automorphism and diffeomorphism of $O(n)$, so the stabiliser subgroup $O(n)_{:M}$ of M is isomorphic to that of Λ by diffeomorphism.

We an show that $O(n)_{:\Lambda} = \prod_{j=1}^{\ell} O(m_j)$. Let p_j be a polynomial that vanishes on $\lambda_k \neq \lambda_j$ and 1 on λ_j (this can be constructed by Lagrange interpolation). Then applying the polynomial to both sides of $Q^T \Lambda Q = \Lambda$, we have

$$Q^T E_j Q = E_j, \quad E_j = \begin{pmatrix} 0_{\sum_{k<j} m_k} & & \\ & \mathrm{id}_{m_j} & \\ & & 0_{\sum_{k>j} m_k} \end{pmatrix}, \quad \forall j. \tag{69}$$

Note that as $\sum_j \lambda_j E_j = \Lambda$, this is is equivalent to $Q^T \Lambda Q = \Lambda$. Now, writing Q in block form, we can check that Q must have the form

$$Q = \begin{pmatrix} Q_1 & & \\ & \ddots & \\ & & Q_\ell \end{pmatrix} \tag{70}$$

where $Q_j \in O(m_j)$, in order to satisfy $Q^T Q = \mathrm{id}$ and $Q^T E_j Q = E_j,\ \forall j$. Thus, the stabilisers subgroup of Λ is the direct product $\prod_{j=1}^{\ell} O(m_j)$. □