Proceedings of Machine Learning Research vol 237:1–54, 2024 35th International Conference on Algorithmic Learning Theory

Universal Representation of Permutation-Invariant Functions on Vectors and Tensors

Puoya Tabaghi PTABAGHI@UCSD.EDU
Yusu Wang YUSUWANG@UCSD.EDU
Halıcıoğlu Data Science Institute, University of California San Diego

Editors: Claire Vernade and Daniel Hsu

Abstract

A main object of our study is multiset functions — that is, permutation-invariant functions over inputs of varying sizes. Deep Sets, proposed by Zaheer et al. (2017), provides a *universal representation* for continuous multiset functions on scalars via a sum-decomposable model. Restricting the domain of the functions to finite multisets of D-dimensional vectors, Deep Sets also provides a *universal approximation* that requires a latent space dimension of $O(N^D)$ — where N is an upper bound on the size of input multisets. In this paper, we strengthen this result by proving that universal representation is guaranteed for continuous and discontinuous multiset functions through a latent space dimension of $O(N^D)$ (which we will further improve upon). We then introduce *identifiable* multisets for which we can uniquely label their elements using an identifier function, namely, finite-precision vectors are identifiable. Based on our analysis of identifiable multisets, we prove that a sum-decomposable model, for general continuous multiset functions requires only a latent dimension of $2DN$, as opposed to $O(N^D)$. We further show that both encoder and decoder functions of the model are continuous — our main contribution to the existing work which lacks such a guarantee. Additionally, this provides a significant improvement over the aforementioned $O(N^D)$ bound, derived for the universal representation of both continuous and discontinuous multiset functions. We then extend our results and provide special sum-decomposition structures to universally represent permutation-invariant tensor functions on identifiable tensors. These families of sum-decomposition models enable us to design deep network architectures and deploy them on a variety of learning tasks on sequences, images, and graphs.

Keywords: Deep learning, permutation-invariance, multiset functions, identifiable tensor functions, universal representation

1. Introduction

There is a wide gamut of machine learning problems that aim to identify an optimal function for unordered collections of entities, such as sets and multisets. Tasks such as set or audience expansion in image tagging, computational advertisement, and astrophysics (Ntampaka et al., 2016; Ravanbakhsh et al., 2016a), parsing objects in a scene (Eslami et al., 2016; Kosiorek et al., 2018), population statistics (Póczos et al., 2013), inference on point clouds (Qi et al., 2017a,b), min-cut and routing on a graph, reinforcement learning (Sunehag et al., 2017), and modeling interactions between objects in a set (Lee et al., 2019) exemplify these problems. Popular machine learning models are designed for ordered algebraic objects, such as vectors, matrices, and tensors. To adapt these standard models to operate on multisets, we must enforce various permutation invariance properties (Oliva et al., 2013; Szabó et al., 2016; Muandet et al., 2013, 2012; Shawe-Taylor, 1993).

To characterize a general class of multiset (or permutation-invariant) functions, several authors have proposed sum-decomposition models (Ravanbakhsh et al., 2016b; Zaheer et al., 2017). Notably, Deep Sets provides a universal representation for continuous multiset functions on *scalars*. This model is a form of Janossy pooling, which is easy to implement and parallelize (Murphy et al., 2018). At its core, it maps elements of the input multiset X individually via ϕ and then aggregates them to *uniquely* encode the input multiset. That is, $\Phi(X) = \sum_{x \in X} \phi(x) \in \mathbb{R}^M$ provides a unique encoding for X, indicating that Φ is an *injective* map. Injectivity is the most important property of the encoder Φ as it operates as an intermediate feature extraction step by uniquely mapping multisets to vectors. Then, to represent a multiset function $f(X)$, we map the resulting feature $\Phi(X)$ to $f(X)$, i.e., $f(X) = \rho \circ \Phi(X)$, where ρ belongs to a rich class of unconstrained functions. The existence of a continuous sum-decomposable model — a continuous encoder Φ and decoder ρ — is guaranteed only if the dimension of the model's intermediate features (M) is sufficiently large. If we reduce this dimension, Wagstaff et al. (2022) proves that no continuous decoder ρ exists such that $\rho \circ \Phi$ can approximate some multiset functions better than a naive constant baseline. Regarding multiset functions on *vectors*, the best available result is given by Zaheer et al. (2017), which only provides a *universal approximation* for continuous multiset functions by analyzing their finite-order Taylor approximation. As our first contribution, we provide a universal representation, through the sum-decomposable model, for continuous and discontinuous multiset functions on vectors, which generalizes the existing universal approximation results. It is important to note that all universal representation results are stronger than their universal approximation counterparts, as the former results imply the latter ones.

Beyond permutation-invariant functions on scalars and vectors, SignNet and BasisNet (Lim et al., 2022), along with other works, are neural network architectures (Dwivedi and Bresson, 2020; Dwivedi et al., 2020, 2021; Beaini et al., 2021; Kreuzer et al., 2021; Mialon et al., 2021; Kim et al., 2022), that provide sign and orthonormal basis invariances as displayed by eigenspaces (Eastment and Krzanowski, 1982; Rustamov et al., 2007; Bro et al., 2008; Ovsjanikov et al., 2008). Laplacian eigenvectors capture connectivity, clusters, subgraph frequencies, and help derive graph positional encodings to generalize Transformers to graphs, improving the performance of Graph Neural Networks (GNNs) (Dwivedi et al., 2020, 2021) and other useful graph properties (Von Luxburg, 2007; Cvetkovic et al., 1997). Under certain conditions, these network structures can universally approximate any continuous function with the desired invariances. Both networks utilize Invariant Graph Networks (IGNs) (Maron et al., 2018) to build permutation invariance or equivariance properties for functions on matrices. IGN treats graphs (with nodes and edges) as *tensors*. Its architecture involves permutation-invariant and equivariant *linear layers* for tensor input and output data. As the tensor order reaches $O(N^4)$, it achieves universality for graphs of size N (Azizian and Lelarge, 2020; Maron et al., 2019b; Keriven and Peyré, 2019).

The type of injective multiset functions, as introduced earlier, is useful in studying the separation power of Message-Passing Neural Networks and its relation to the Weisfeiler-Leman (WL) graph isomorphism test (Xu et al., 2018). They are also used to show the equivalence of high-order GNNs to high-order WL tests (Morris et al., 2019; Maron et al., 2019a), and results related to geometric GNNs and WL tests (Hordan et al., 2023; Joshi et al., 2023; Pozdnyakov and Ceriotti, 2022). Amir et al. (2023) provide a theoretical analysis of the required latent dimension for nonpolynomial encoders — namely, sigmoid, hyperbolic tangent, sinusoid — to achieve an injective multiset function.

Contributions. In this paper, we primarily focus on the study of multivariate multiset functions, that is, functions on multisets containing at most N vectors of dimension D. When $D = 1$, this simplifies to multiset functions on scalars. Our main contributions are as follows:

1. We propose extended versions of the sum-decomposition models of multiset functions on vectors (Zaheer et al., 2017). Multiset functions encompass permutation-invariant functions as they are invariant to the specific ordering of the input elements. We use the term "multiset function" to emphasize that the number of input elements can vary, unlike permutation-invariant functions. Our initial contribution, in Section 3, presents the universal representation for continuous and discontinuous multiset functions over D-dimensional vectors through a sum-decomposable model; refer to Theorem 8. The latent dimension of this model is $\binom{N+D}{D} - 1$, where N is the upper bound on the size of input multisets. For universal representation of *continuous* multiset functions, we show that both encoder and decoder functions of the sum-decomposable model are also continuous; see Theorem 3. In the case of a scalar domain ($D = 1$), this latent dimension aligns with the one in (Wagstaff et al., 2019, 2022): $\binom{N+1}{1} - 1 = N$. Our Theorems 3 and 8 contribute as novel additions to the existing universal approximation results for continuous multiset functions (Zaheer et al., 2017; Maron et al., 2019a; Segol and Lipman, 2019) where the same latent dimension is only achieved for universal approximation for $D > 1$. Universal approximation results rely on finite-order Taylor approximation of continuous multiset functions. This technique does not work for (1) universal representation and (2) discontinuous multiset functions. What's more, as we will see next, we will further significantly reduce this latent dimension bound for representing continuous multiset functions via a novel technique through the use of identifiable multisets.

2. In Section 4, we introduce the concept of *identifiable* multisets. These are multisets whose distinct elements can be uniquely labeled via a continuous functional; for instance, multisets containing finite-precision vectors are identifiable through a linear functional. We show that on identifiable D-dimensional vector multisets, the latent dimension of the sum-decomposable representations can be reduced to $2DN$ from the original $\binom{N+D}{D} - 1$. More importantly, by analyzing identifiable multisets, we establish that universal representation of continuous multiset functions, where both encoder and decoder functions are continuous, is possible with a latent dimension of $2DN$; refer to Theorem 6. The techniques used to derive these results focus on the concept of an identifier function, which distinguishes our approach from prior works using polynomial and nonpolynomial-based encoders (Zaheer et al., 2017; Dym and Gortler, 2022). While our result in Theorem 3 is suboptimal compared to this new result (Theorem 6), we still include Section 3 as it yields a better result compared to the existing work based on polynomial-based encoders (common in approximation approaches), making it of independent interest. In summary, the main contribution of our results to existing literature are (1) the lowest latent dimension bound, and (2) the continuity guarantee for the decoder function.

3. Finally, we offer universal representation for continuous and discontinuous permutation-invariant tensor functions of arbitrary order. We derive a nested sum-decomposable representation specifically for what we term *identifiable tensors*, akin to identifiable multisets. Depending on the chosen identifier function, we provide distinct bounds on the latent dimensions for this representation. This mirrors an existing decomposition result on permutation-equivariant functions on matrices (tensors of order two) (Fereydounian et al., 2022). However, our proposition introduces a modified encoder function that (1) yields a reduced latent dimension of $2DN$ compared to $\binom{D}{2}N$, (2) extends the sum-decomposition representation to tensors of arbitrary order, and (3) guarantees injectivity.

	Input	Latent	Representation	Approximation	Continuity
Zaheer et al. (2017)	$D = 1$	$N + 1$	✓	✓	✓
Zaheer et al. (2017)	$D > 1$	—	✗	✓	✓
Wagstaff et al. (2019)	$D = 1$	N	✓	✓	✓
Segol and Lipman (2019)	$D > 1$	$\binom{N+D}{D} - 1$	✗	✓	✓
Wang et al. (2023)	$D > 1$	$\text{poly}(N, D)$	✓	✓	✓
Dym and Gortler (2022)	$D > 1$	$2DN + 1$	✓*	✗	✗
Amir et al. (2023)	$D > 1$	$2N(D + 1) + 1$	✓*	✗	✗
Ours	$D > 1$	$2ND$	✓	✓	✓

Table 1: Summary of recent results in sum-decomposable representation of permutation-invariant functions. We list the input dimension of each element in the (multi)set (Input), the latent dimension (Latent), continuity of the decoder function ρ (Continuity). We denote the universal approximation and representation in a compact domain by ✓ and universal representation on a subset subdomain by ✓*.

More on related work. The most notable work on universal representation of nonlinear multiset functions concerns scalar-valued domains (Wagstaff et al., 2019, 2022). Much of the existing literature focuses on universal approximations for permutation-invariant and -equivariant functions. The sum-decomposition of multiset functions on multidimensional entities has been primarily approached through the universal approximation power of polynomial functions (Zaheer et al., 2017; Segol and Lipman, 2019). Wagstaff et al. (2022) thoroughly investigate the theoretical distinction between universal representation and approximation of multiset functions on *scalars*. However, this remains an open question for multiset functions on multivariate elements. Invariant and equivariant *linear* functions have been thoroughly studied in the literature (Maron et al., 2018; Ravanbakhsh, 2020). In comparison, our nonlinear model generalizes the permutation-invariant linear layers utilized in IGNs (Maron et al., 2018). For universal approximation on N points, the IGNs require $\Omega(N^N)$-sized intermediate tensors (Cai and Wang, 2022). An important class of permutation-compatible (invariant or equivariant) *nonlinear* functions is GNN — the primary iterative-based models for learning information over graphs. A substantial body of work aims to understand the expressive power of GNNs Maron et al. (2019b,a); Keriven and Peyré (2019); Garg et al. (2020); Azizian and Lelarge (2020); Bevilacqua et al. (2021). To provide insight into the capability of GNNs in representing graph functions, Fereydounian et al. (2022) introduces an algebraic formulation — akin to the sum-decomposition model for multiset functions — to represent permutation-equivariant *nonlinear* functions on matrices in terms of the composition of simple encoder and decoder functions. One can connect the notion of permutation-compatible functions (on 2-tensors) to our proposed algebraic form of permutation-invariant functions on k-tensors. However, by focusing on identifiable tensors, we lower the latent dimension required for representing 2-tensors to $O(DN)$ — compared to $O(D^2N)$ in (Fereydounian et al., 2022) — and guarantee the *injectivity* of the encoding function. In Table 1, we compare some of the most recent results in the literature.

Organization. In Section 2, we review the existing sum-decomposition results for multiset functions on scalars. Then, in Section 3, we present our universal representation results for multivariate multiset functions. In Section 4, we introduce identifiable multisets and show how they can be used to derive a lowered latent dimension bound for the continuous sum-decomposition of continuous multiset functions. Finally, focusing on permutation invariance, we propose a nested sum-decomposition model to represent invariant functions over k-tensors in Section 5. Identifiability for tensors is the main concept necessary to establish the aforementioned decomposition models. We relegate all proofs, supplementary results, and discussions to the Appendix.

Notations. We denote the nonnegative reals by $\mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$. For any $N \in \mathbb{N}$, we let $[N] = \{1, \ldots, N\}$. The function f maps elements from its domain to elements in its codomain, meaning $f : \mathrm{dom}(f) \to \mathrm{codom}(f)$ where $\mathrm{codom}(f) = \{f(x) : x \in \mathrm{dom}(f)\}$. Examples of domains include $\mathbb{R}$, $\mathbb{R}^D$, $\mathbb{N}$, and $\mathbb{Q}$. We denote the collection of subsets of a domain $\mathbb{D}$ as $2^{\mathbb{D}}$. Let $\mathbb{D}$ be a domain and $f : \mathbb{D} \to \mathrm{codom}(f)$. We then let $f(\mathbb{D}_1) \stackrel{\text{def}}{=} \{f(x) : x \in \mathbb{D}_1\} \subseteq \mathrm{codom}(f)$ where $\mathbb{D}_1 \subseteq \mathbb{D}$. A multiset is a pair (X, m) where X is a set of objects and m is a map from X to cardinals (representing the multiplicity of each element in X). We simplify this notation by identifying multisets by "multiset X" or using double curly brackets, namely, $X = \{\{1, 1, 2\}\}$ has three elements but $X = \{1, 1, 2\} = \{1, 2\}$ has two elements. For any domain $\mathbb{D}$ and multiset X, $X \subseteq \mathbb{D}$ means that the underlying set for X (repetitive elements removed) is a subset of $\mathbb{D}$, and $|X|$ is the size of the multiset (repetitive elements included). For $N \in \mathbb{N}$ and domain $\mathbb{D}$, we let $\mathbb{X}_{\mathbb{D},N} = \{\text{multiset } X \subseteq \mathbb{D} : |X| = N\}$, $\mathbb{X}_{\mathbb{D},S} = \{\text{multiset } X \subseteq \mathbb{D} : |X| \in S\}$ where $|\cdot|$ returns the cardinality of its input set (or multiset) and $S \subseteq \mathbb{N}$, namely, $\mathbb{X}_{\mathbb{D},[N]} = \{\text{multiset } X \subseteq \mathbb{D} : 1 \leq |X| \leq N\}$. We denote multisets (and sets) with X and tensors (and matrices) with T.

2. Review of the Sum-decomposable Model for Multiset Functions on Scalars

Standard machine learning algorithms operate on data arranged in canonical ways — namely, vectors, matrices, and tensors. However, in statistical estimation, set expansion, outlier detection (Zaheer et al., 2017), and problems involving point clouds or groups of atoms forming a molecule (Wagstaff et al., 2022), we often aim to learn maps defined on an unordered collection of entities, that is, a set or a multiset. Throughout this paper, we treat functions defined on sets and multisets differently. Here, we use a (multi)set function over $\mathbb{D}$ to refer to a function whose domain consists of sub(multi)sets of a domain $\mathbb{D}$. In other words, a multiset function assigns a value for every possible submultiset of the domain $\mathbb{D}$. A multiset function f must satisfy two key criteria: (1) be invariant to the ordering of its input elements (permutation invariance), and (2) be well-defined on multisets of different sizes.

Generally, when aiming to model a multiset function, it is not immediately evident how to ensure that the function satisfies condition (1), i.e., permutation invariance concerning the ordering of elements in input multisets. One powerful approach to address this challenge is to initially establish a complete representation of multiset functions through a specific composition of *unconstrained functions* — referred to here as *encoder and decoder functions*. This decomposition not only characterizes multiset functions but also holds crucial significance in the learning context. For instance, these unconstrained functions can be modeled (and learned) using neural networks, as seen in the popular Deep Sets architecture (Zaheer et al., 2017). A particular form of this composition is known as a *sum-decomposable* representation. The following provides such a result for set functions defined on a countable domain.

Theorem 1 (Zaheer et al. 2017) *Let $f : 2^{\mathbb{D}} \to \mathrm{codom}(f)$ where $\mathbb{D}$ is a countable domain. Then,*

$$\forall X \subseteq \mathbb{D} : f(X) = \rho \circ \Phi(X), \;\; \Phi(X) = \sum_{x \in X} \phi(x), \tag{1}$$

where $\phi : \mathbb{D} \to \mathrm{codom}(\phi) \subset \mathbb{R}$, $\rho : \mathrm{codom}(\Phi) \to \mathrm{codom}(f)$, *and* $\mathrm{codom}(\Phi) = [0, 1] \subset \mathbb{R}$.

Theorem 1 provides an algebraic construct for universal representation for set functions on countable sets. We use the term universal representation to distinguish it from the weaker universal approximation results in the literature. This universal representation is obtained via the so-called sum-decomposable representation formally defined as follows:

Definition 1 *A (multi)set function f over $\mathbb{D}$ is sum-decomposable, or it has a sum-decomposable representation, if it can be written as $f(X) = \rho \circ \Phi(X)$ for any (multi)set $X \subseteq \mathbb{D}$, where $\Phi(X) = \sum_{x \in X} \phi(x)$. We refer to ϕ, Φ and ρ as the* element-encoder, (multi)set-encoder, *and* decoder *functions, respectively. We also may call ϕ and Φ sometimes simply as encoder functions. Furthermore, suppose $\phi : \mathbb{D} \to \mathrm{codom}(\phi) \subseteq \mathbb{R}^M$, then we refer to $\mathbb{R}^M$ (which is the ambient space of $\mathrm{codom}(\phi)$) as the decomposition model's latent space, and say that f is sum-decomposable via $\mathbb{R}^M$. The latent dimension of this sum-decomposition is M. A continuous multiset function f is continuously sum-decomposable if has a sum-decomposable representation where both the encoder and decoder functions, that is, ϕ (and thus Φ) and ρ are continuous in the entire ambient space of their respective domains.*

Theorem 1 states that a set function over a countable set is essentially sum-decomposable via $\mathbb{R}$, and the latent dimension is one. Interestingly, (Wagstaff et al., 2022) shows that set functions on an *uncountable* domain $\mathbb{D}$ do not admit this sum-decomposable representation. Nevertheless, there is an extension of Theorem 1 to finite-sized multisets (Wagstaff et al., 2022).

Theorem 2 (Wagstaff et al. 2019) *Let $N \in \mathbb{N}$, and $f : \mathbb{X}_{\mathbb{D},[N]} \to \mathbb{R}$ be a continuous multiset function where $\mathbb{D} = [0, 1]$. Then, it is continuously sum-decomposable (see Definition 1) via $\mathbb{R}^N$ — that is, the latent space is a subset of $\mathbb{R}^N$ — and vice versa.*

Recall that $\mathbb{X}_{\mathbb{D},[N]}$ represents the collection of all multisets over $\mathbb{D}$ with a cardinality of at most N. Given that $\mathbb{D}$ in the above theorem is $[0, 1] \subset \mathbb{R}$, the result asserts that a continuous multiset function over scalars is continuously sum-decomposable via $\mathbb{R}^N$, where N is the maximum cardinality of input multisets. The continuity of the decoder ρ comes at the cost of an increased latent space dimension — compare universal representation in Theorems 1 and 2. This latent dimension is tight in the worst case, that is, there does not exist a sum-decomposition via a latent space with dimension less than N (Wagstaff et al., 2022). However, in practical scenarios, a specific multiset function may allow for a sum-decomposition with a significantly lower latent dimension. While one might anticipate a reduction in the latent dimension for universal approximation, it is interesting that, at least for multiset functions over scalars, contrary to intuitive expectations, universal approximation is not achievable (for all multiset functions) by reducing the latent dimension from N(Wagstaff et al., 2022).

3. Warmup: Sum-decomposable Model for Multiset Functions on Vectors

Theorem 2 pertains to multiset functions operating on scalar-valued elements (that is, the input is a multiset with elements from $\mathbb{R}$). In practice, we often encounter applications involving vector-valued multisets. For instance, a multiset of $\leq N$ points in $\mathbb{R}^D$ can be represented as a multiset of cardinality $\leq N$ over $\mathbb{R}^D$. Similarly, in graph learning settings, we may have a set of N nodes in a graph with D-dimensional node features. In what follows, we focus on multiset functions over vectors in $\mathbb{R}^D$, that is, functions of the form $f : \mathbb{X}_{\mathbb{D},[N]} \to \mathbb{R}$ where $\mathbb{D} \subset \mathbb{R}^D$. For simplicity, we initially consider functions over multisets of precisely cardinality N, that is, $f : \mathbb{X}_{\mathbb{D},N} \to \mathbb{R}$. Our primary result in this section is the following theorem:

Theorem 3 *A continuous multivariate multiset function $f : \mathbb{X}_{\mathbb{D},N} \to \mathrm{codom}(f)(\subseteq \mathbb{R}^n)$, over a multisets of N elements in a compact set $\mathbb{D} \subseteq \mathbb{R}^D$, is continuously sum-decomposable via $\mathbb{R}^{\binom{N+D}{D}-1}$. That is, encoder ϕ is continuous over $\mathbb{D}$, and decoder ρ is continuous over $\mathbb{R}^{\binom{N+D}{D}-1}$.*

The above theorem asserts that a continuous multiset function over multisets of N vectors from $\mathbb{D} \subset \mathbb{R}^D$ is continuously sum-decomposable via a latent dimension of $\binom{N+D}{D} - 1$. In the special case of $D = 1$, this aligns with the previous result for multiset functions over scalars in Theorem 2. In Section 4, we present a stronger result with a significantly lower latent dimension. Nevertheless, we include this result for two reasons: (1) it is obtained using a similar proof technique to Theorem 2 by employing polynomial-based encoders, and (2) this novel finding arrives at the same latent dimension reported in (Zaheer et al., 2017) for the universal approximation of continuous multiset functions. Detailed proofs are provided in Appendices A and B. A high-level description is outlined here.

In the remainder of this section, we designate $\mathbb{D} \subset \mathbb{R}^D$ as a compact subset of $\mathbb{R}^D$. Following the proof technique outlined in (Zaheer et al., 2017), to show the existence of a sum-decomposition of $f = \rho \circ \Phi$, our aim is to construct a multiset encoder Φ that is *injective* over $\mathbb{X}_{\mathbb{D},N}$. Once we establish an injective encoder Φ, we can then define $\rho = f \circ \Phi^{-1}$ for all admissible inputs, that is, within $\mathrm{codom}(\Phi)$. The encoder Φ is inherently continuous by construction. The primary challenge lies in proving that $\rho = f \circ \Phi^{-1}$ is not only well-defined but also continuous across the latent space $\mathrm{codom}(\Phi)$.

To construct an injective multiset function $\Phi(X) = \sum_{x \in X} \phi(x)$, we use permutation-invariant polynomials as in (Maron et al., 2019a; Segol and Lipman, 2019). We express these polynomials as follows:

$$\forall X \in \mathbb{X}_{\mathbb{R}^D,N} : \; p(X) = \mathrm{poly}(e_1(X), \cdots, e_K(X)), \tag{2}$$

where $e_k(X) = \sum_{x \in X} \prod_{d=1}^{D} x_d^{k_d}$ is a power-sum multi-symmetric polynomial, $k_1 \dots k_D$ is the D-digit representation of $k \in [K]$ in base $N+1$, $K = \binom{N+D}{D} - 1$, and poly is a polynomial function (Rydh, 2007).

Remark 1 *It is known that one can universally approximate continuous multivariate multiset functions over a compact set with a multiset polynomial in equation (2). Given $K = \binom{N+D}{D} - 1$ power-sum multi-symmetric polynomial bases $\big(e_k(X)\big)_{k \in [K]}$, we can design an encoder ϕ to provide a universally approximate sum-decomposable model for multivariate continuous multiset functions via $\mathbb{R}^{\binom{N+D}{D}-1}$; refer to Theorem 9 in (Zaheer et al., 2017).*

In Appendix A, we first state Theorem 8, which ensures a universal representation (not universal approximation, as mentioned in the preceding remark) of *any* multivariate $(D > 1)$ multiset functions — whether continuous or discontinuous — through the sum-decomposable model using $\mathbb{R}^{\binom{N+D}{D}-1}$. The resulting decoder ρ constructed this way may not be continuous; however, this contributes significantly to the existing literature. From a technical standpoint, Theorem 8 is valuable as it does not rely on approximating the multiset function f using finite-order polynomials. Instead, it aims to show the injectivity of Φ through an analysis of the parameterized roots of a class of multivariate polynomials. Building upon Theorem 8, in Appendix B, we establish that *if f is a continuous multiset* *function*, then its decoder $\rho = f \circ \Phi^{-1}$ maintains continuity in the ambient space of $\mathrm{codom}(\Phi)$, i.e., $\mathbb{R}^{\binom{N+D}{D}-1}$. The key approach involves proving that (1) Φ^{-1} is a continuous function on $\mathrm{codom}(\Phi)$ and (2) $\mathrm{codom}(\Phi)$ is a compact subset of $\mathbb{R}^{\binom{N+D}{D}-1}$. This concludes the proof of Theorem 3.

We can further generalize the results in Theorems 3 and 8 to multisets of varying sizes.

Theorem 4 *Theorems 3 and 8 are valid for multivariate multiset functions of at most N elements from a compact subset $\mathbb{D} \subset \mathbb{R}^D$, that is, $\mathbb{X}_{\mathbb{D},[N]}$.*

As a direct consequence of the proof technique in Theorem 4 — particularly noting that the construction of the injective multiset encoder Φ is independent of the multiset function f being represented — in Proposition 1, we show that when dealing with functions on the product of *distinct* multisets of D-dimensional vectors, we can use *the same* encoder within its sum-decomposable model.

Proposition 1 *A (continuous) multiset function $f : \mathbb{X}_{\mathbb{D},[N_1]} \times \mathbb{X}_{\mathbb{D},[N_2]} \to \mathrm{codom}(f)$, where $\mathbb{D}$ is compact subset of $\mathbb{R}^D$, is (continuously) sum-decomposable via $\mathbb{R}^{\binom{N+D}{D}-1} \times \mathbb{R}^{\binom{N+D}{D}-1}$, that is,*

$$\forall X \in \mathbb{X}_{\mathbb{D},[N_1]}, X' \in \mathbb{X}_{\mathbb{D},[N_2]} :\ f(X, X') = \rho\Big(\sum_{x \in X} \phi(x), \sum_{x' \in X'} \phi(x') \Big),$$

where continuous $\phi : \mathbb{R}^D \to \mathbb{R}^{\binom{N+D}{D}-1}$, $N = \max\{N_1, N_2\}$ and (continuous) $\rho : \mathbb{R}^{\binom{N+D}{D}-1} \times \mathbb{R}^{\binom{N+D}{D}-1} \to \mathrm{codom}(\rho)$, and $\mathrm{codom}(f) \subset \mathrm{codom}(\rho)$.

Remark 2 *Proposition 1 simply follows from the injectivity of the multiset encoding function Φ. This is completely different than the universal approximation approach where we rely on approximating the function using finite order permutation-invariant polynomials and finding the complete set of "basis" for them. Using the polynomial approximation approach, it is rather challenging to arrive at the sum-decomposition in Proposition 1. This issue is compounded when we consider deriving a decomposition for functions defined on multisets of multisets. This is yet another reason why we focus on universal representations as opposed to universal approximations.*

Relation to the results of Fereydounian et al. (2022). We note that Fereydounian et al. (2022) propose an encoder Φ that is injective over particular (multi)sets $\mathbb{X}^s_{N,D}$ (not all multisets) of D-dimensional vectors. The function Φ provides unique encodings for these (multi)sets in $\mathrm{codom}(\Phi) \subset \mathbb{R}^{\binom{D}{2}N}$ — where $\mathrm{codom}(\Phi) = \{\Phi(X) : X \in \mathbb{X}^s_{N,D}\}$ and N is the size of the input (multi)sets. This leads to a sum-decomposition for functions over (multi)sets in $\mathbb{X}^s_{N,D}$. More importantly, for a continuous multiset function, a continuous sum-decomposition $f = \rho \circ \Phi$ is not guaranteed over all multisets; in particular, the continuity of $\rho = f \circ \Phi^{-1}$ is only guaranteed over $\mathrm{codom}(\Phi)$ — an *open* subset of $\mathbb{R}^{\binom{D}{2}N}$. Therefore, it does not guarantee the existence of a continuous extension for ρ to $\mathbb{R}^{\binom{D}{2}N}$; see Appendix L for a detailed discussion.

4. Sum-decomposable Models on Identifiable Multisets

Inspired by the theoretical difference in latent space dimensions for sum-decomposition representations between set and multiset functions (refer to the result in (Fereydounian et al., 2022)), our objective is to reduce the dimension of the latent space. In this section, we present an approach to achieve this by first restricting the domain to what we term "identifiable multisets," introduced below. Subsequently, we will show the extension of these results to scenarios where this restriction is lifted.

Definition 2 *Let $l : \mathbb{D} \to \mathbb{R}$ be a continuous function and $\mathbb{D}$ be a domain. We denote $\mathbb{X}^l_{\mathbb{D},N} = \{X \in \mathbb{X}_{\mathbb{D},N} : \forall x, x' \in X, l(x) = l(x') \to x = x'\}$, as the set of multisets of size N that are identifiable via l, that is, l-identifiable.*

According to Definition 2, the continuous identifier function l uniquely labels distinct elements of multisets in $\mathbb{X}^l_{\mathbb{D},N}$. In Theorem 5 and Proposition 2 we provide improved bounds on latent dimensions given in Theorems 4 and 8 — by restricting the domain of multiset functions to l-identifiable multisets.

Theorem 5 *Let $f : \mathbb{X}_{\mathbb{R}^D,N} \to \operatorname{codom}(f)$ be a multiset function and $\ell : \mathbb{R}^D \to \operatorname{codom}(l) \subseteq \mathbb{R}$ be continuous. Then, there is a continuous function $\phi : \mathbb{R}^D \to \operatorname{codom}(\phi) \subset \mathbb{C}^{D\times N}$ such that*

$$\forall X \in \mathbb{X}^l_{\mathbb{R}^D,N} : f(X) = \rho\Big(\sum_{x\in X} \phi(x)\Big) = \rho \circ \Phi(X),$$

where $\rho : \Phi(\mathbb{X}^l_{\mathbb{R}^D,N}) \to \operatorname{codom}(f)$ and $\Phi(\mathbb{X}^l_{\mathbb{R}^D,N}) \stackrel{\text{def}}{=} \{\Phi(X) : X \in \mathbb{X}^l_{\mathbb{R}^D,N}\}$.

Proposition 2 *Theorem 5 is valid for multivariate multiset functions of at most N elements from a compact subset of $\mathbb{R}^D$.*

Remark 3 *Theorem 5 asserts the feasibility of sum-decomposition for* arbitrary *(continuous or discontinuous) multiset functions via a latent dimension of $O(ND)$ on inputs that are identifiable through a continuous identifier $l : \mathbb{R}^D \to \operatorname{codom}(l) \subseteq \mathbb{R}$. In comparison, the universal representation results in Theorems 3 and 4 require a latent space dimension of $O(N^D)$, which becomes impractical even for a small number of features. Additionally, the bound in Theorem 5 improves from the $O(ND^2)$ bound proposed in (Fereydounian et al., 2022). Furthermore, we propose a concrete characterization of the input domain in Definition 2, adaptable for* any *continuous function l designed for specific applications. However, as the set of identifiable multisets $\mathbb{X}^l_{\mathbb{D},N}$ (where $\mathbb{D}$ is a compact subset of $\mathbb{R}^D$) does not form a compact set, there is no guarantee that $\rho : \operatorname{codom}(\mathbb{X}^l_{\mathbb{D},N}) \to \operatorname{codom}(f)$ has a continuous extension to $\mathbb{C}^{D\times N}$. In other words, using the multiset encoding function Φ (introduced in the proofs), some multiset functions f, there may not exits a continuous $\rho : \mathbb{C}^{D\times N} \to \operatorname{codom}(\rho)$ enabling the sum-decomposition. However, we address this issue in Section 4.1. It is worth noting that our specific multiset encoder Φ maps multisets to complex-valued matrices in $\mathbb{C}^{D\times N}$, which, without causing technical issues, can be viewed as $\mathbb{R}^{2D\times N}$.*

Remark 4 *The multiset encoding function Φ in Proposition 2 resembles the notion of separating invariants introduced in (Dym and Gortler, 2022). In this context, the quantity $\Phi(X)$ remains invariant under permutations, treated as group actions. However, a subtle distinction exists; multiset functions are permutation-invariant, but the converse is not necessarily true, as multiset functions may accommodate varying-sized inputs. In the work by Dym and Gortler (2022), they claim that for randomized invariants of dimension $2DN + 1$ (compared to ours, which is $2DN$), almost all matrices in $\mathbb{R}^{D\times N}$ can be identified up to the permutation of their columns using separating invariants. They achieve this by applying linear projections on multidimensional elements to obtain* *scalars and then using a continuous separating (injective) map on them. They prove that the measure of matrices that can not be identified via the permutation-invariant encoding is zero. Consequently, the sum-decomposition does not apply to all matrices (similar to multisets in our paper), and there is no guarantee for the existence of a continuous decoder ρ (over the ambient space) to represent a continuous permutation-invariant function. On another note, Amir et al. (2023) propose utilizing a nonpolynomial element-encoder, denoted as ϕ in our notation, to construct an injective multiset function Φ. They arrive at a latent dimension of $2N(D+1)+1$. However, their construction of ϕ requires random parameter selection, and the injectivity only holds in the almost surely sense. Therefore, it may not work for certain parameters. Zweig and Bruna (2022) study a theoretical bottleneck for the latent dimension of sum-decomposable models. Very recently, we were made aware of a concurrent work (Wang et al., 2023): They provide polynomial bounds for the latent dimension of sum-decomposable models with feature maps from specific function classes, in particular, multidimensional features $\phi(w^\top x)$ — where ϕ is the exponential function — require latent dimension in the range $[N(D+1), N^5D^2]$; and if ϕ is a power mapping, the range of latent dimension becomes $[ND, N^4D^2]$. This is a great contribution to the previous work as it ensures the continuity of the decoder function ρ while offering reduced latent dimensions. In comparison, we provide a latent dimension that is linear in both N and D and guarantee the continuity of the decoder function.*

4.1. Towards a Continuous Decoder

In Theorem 5, we prove how our concept of ℓ-identifiable multisets enables a reduced latent dimension for the sum-decomposition representation of multiset functions. Current leading approaches that allow such reduced-dimensional representations rely on probabilistic arguments, specifically excluding multisets of measure zero from all valid multisets; refer to Remark 4. However, these approaches do not yet yield a continuous sum-decomposition, particularly a continuous decoder function ρ. In what follows, we focus on utilizing ℓ-identifiable multisets, aiming to facilitate the representation on a dense *subset of multisets*, as indicated by Proposition 3 and Lemma 1 below. Our goal is to ultimately achieve a continuous sum-decomposition, akin to Theorem 6. You can find the proofs of all these results in Appendices G to I.

Proposition 3 *Let $\mathbb{X}_{\mathbb{Q}^D,N}$ be the set of all multisets of N vectors from $\mathbb{Q}^D$ where $\mathbb{Q}$ denotes the set of rational numbers. Then, $\mathbb{X}_{\mathbb{Q}^D,N}$ is an l-identifiable subset of $\mathbb{X}_{\mathbb{R}^D,N}$.*

Lemma 1 *Let $\mathbb{D} \subseteq \mathbb{R}^D$ be a compact set with continuous nonempty interior, $Q(\mathbb{D}) = \mathbb{D} \cap \mathbb{Q}^D$ be the set of all vectors with rational elements in $\mathbb{D}$. Then, $\mathbb{X}_{Q(\mathbb{D}),N}$ is a dense subset of $\mathbb{X}_{\mathbb{D},N}$. Similarly, $\Phi(\mathbb{X}_{Q(\mathbb{D}),N})$ is a dense subset of $\Phi(\mathbb{X}_{\mathbb{D},N})$ where Φ is the multiset encoder in Theorem 5.*

Remark 5 *To illustrate the utility of Theorem 5, consider its application in circuit design tasks. These tasks encompass various electronic design aspects, such as routed wire length prediction (Xie et al., 2021), circuit partitioning (Lu et al., 2020), logic synthesis (Zhu et al., 2020), and placement optimization (Li et al., 2020). Circuits can be represented as geometric graphs, where nodes are positioned on integer-valued vector coordinates, each node possessing multidimensional features that characterize circuit elements. As supported by Proposition 3, we are able to uniquely identify each node using a continuous identifier. Given their significance as a subset of ℓ-identifiable multisets, Corollary 1 specializes Theorem 5 to rational-valued multisets.*

Corollary 1 *Let $f : \mathbb{X}_{\mathbb{R}^D,N} \to \mathrm{codom}(f)$ be a multiset function. Then, there is a continuous function $\phi : \mathbb{R}^D \to \mathrm{codom}(\phi) \subset \mathbb{C}^{D\times N}$ such that*

$$\forall X \in \mathbb{X}_{\mathbb{Q}^D,N} : \; f(X) = \rho\big(\sum_{x\in X} \phi(x)\big) = \rho \circ \Phi(X),$$

and $\rho : \Phi(\mathbb{X}_{\mathbb{Q}^D,N}) \to \mathrm{codom}(f)$.

Corollary 1 states that the sum-decomposable model is valid — via a latent dimension of $2DN$ — on a dense subset of multisets in $\mathbb{X}_{\mathbb{R}^D,N}$; see Lemma 1. However, this representation has some drawbacks: (1) the measure of valid multisets $\mathbb{X}_{\mathbb{Q}^D,N}$ is zero and (2) there is no guarantee on the existence of a continuous extension of ρ to $\mathbb{C}^{D\times N}$. It is important to note that we choose to focus on $\mathbb{X}_{\mathbb{Q}^D,N}$ despite the fact that it has a measure zero. We argue that one should not focus on the measure of valid multisets $\mathbb{X}_{\mathbb{Q}^D,N}$; but rather take advantage of the fact that valid multisets form a dense subset of all multisets, that is, $\mathbb{X}_{\mathbb{R}^D,N}$. In Theorem 6, we leverage this fact and resolve both aforementioned issues by focusing on the sum-decomposable representation of *continuous* multiset function.

Theorem 6 *Consider a compact set $\mathbb{D} \subset \mathbb{R}^D$ with nonempty interior. Let $f : \mathbb{X}_{\mathbb{D},N} \to \mathrm{codom}(f)$ be a* continuous *multiset function and $\Phi : \mathbb{X}_{\mathbb{D},N} \to \mathrm{codom}(\Phi)$ be the function in Theorem 5. Then, there exists a* continuous *function $\rho : \mathbb{C}^{D\times N} \to \mathrm{codom}(\rho) \subseteq f(\mathbb{X}_{\mathbb{D},N})$ such that*

$$\forall X \in \mathbb{X}_{\mathbb{D},N} : f(X) = \rho \circ \Phi(X).$$

The major contribution of Theorem 6 is the continuity of ρ over the whole latent space. The detailed proof of this key theorem is in Appendix I. At the high level, we begin with the result in Corollary 1. There, we claim that there exits decoding function $\rho : \Phi(\mathbb{X}_{Q(\mathbb{D}),N}) \to \mathrm{codom}(\rho)$ such that the stated decomposition remains valid on rational-valued vectors in $\mathbb{D} \subset \mathbb{C}^{D\times N}$. This result does not guarantee the continuity of ρ in $\mathbb{C}^{D\times N}$. However, we leverage the facts that (1) f is a continuous multiset function and (2) $\Phi(\mathbb{X}_{Q(\mathbb{D}),N})$ is a *dense* (noncompact) subset of $\Phi(\mathbb{X}_{\mathbb{D},N})$ and prove that ρ has a *continuous extension* to $\Phi(\mathbb{X}_{\mathbb{D},N})$ —- a compact subset of $\mathbb{C}^{D\times N}$ — and therefore has a continuous extension to $\mathbb{C}^{D\times N}$. The continuity guarantee of the decoder function ρ is the major contribution of Theorem 6 over existing results in (Dym and Gortler, 2022) and (Fereydounian et al., 2022).

Remark 6 *With regard to machine learning applications, we note that one could implement a network structure of the form $\rho \circ \sum_{x\in X} \phi(x)$ where ρ and ϕ are two parameterized neural nets. However, without the result of Theorem 6, it is not guaranteed that this composite network (of any depth) can even approximate an arbitrary target function f — even if f is a continuous permutation-invariant function. For instance, a target function f might only allow a discontinuous decoder ρ — which neural nets are not guaranteed to be able to approximate. However, Theorem 6 proves the continuity of both encoder and decoder functions when they meet the latent dimension requirements.*

5. Permutation-Invariant Tensor Functions

Data with underlying a hypergraph structure — that is, nodes connected with weighted (hyper)edges — are ubiquitous in many applications (Chen et al., 2019; Ma et al., 2018; Wang et al., 2019; Yang et al., 2019). Inspired by such data, we study functions defined on *tensors* and adopt graph-theoretic notions to describe relevant concepts. The tensor setting is also used for the higher order graph neural network called IGN (Invariant graph network) (Maron et al., 2018).

Definition 3 *Let $N, K \in \mathbb{N}$. We denote $\mathbb{T}_{N,K}$ as the set of K-th order D-dimensional tensors on N entities, that is, $\mathbb{T}_{N,K} = \mathbb{R}^{N^K \times D}$.*

We can use tensors to represent (1) node features, (2) graph adjacency matrix (second order tensor), and (3) hypergraph hyperedges with multidimensional features. In Definition 4, we introduce a tensor notation for permuting node entities.

Definition 4 *Let $N \in \mathbb{N}$, $\Pi(N)$ be the set of permutations over $[N]$, and $\pi \in \Pi(N)$. Then, we let*

$$T, T' \in \mathbb{T}_{N,K} : T' = \pi(T) \Longleftrightarrow T'_{n_1 \ldots n_K} = T_{\pi(n_1) \ldots \pi(n_K)} \;\; \text{for all} \;\; n_1, \ldots, n_K \in [N].$$

Tensors $T, T' \in \mathbb{T}_{N,K}$ are congruent, denoted by $T \equiv T'$, if there is $\pi \in \Pi(N)$ such that $T' = \pi(T)$.

Similar to multiset functions, tensor functions must also exhibit permutation invariance. Using the notation in Definition 4, a permutation-invariant tensor function $f : \mathbb{T}N, K \to \mathrm{codom}(f)$ satisfies $f(T) = f(\pi(T))$ for all $T \in \mathbb{T}N, K$ and the permutation operator $\pi : [N] \to [N]$. This represents a specific form of G-invariant functions (Maron et al., 2019b), where G is the permutation group. It extends the notion of permutation-compatible functions, originally established for 2-tensors denoting input graphs with node features and an adjacency second-order tensor (matrix) (Fereydounian et al., 2022).

Hereafter, we introduce a sum-decomposable model to universally represent permutation-invariant tensor functions of arbitrary order. Our algebraic approach hinges on assigning a unique label to each node. This becomes applicable when dealing with tensors accompanied by distinct node features or hypergraph structures that allow for this unique labeling. In Definition 5, we formalize the set of all tensors that permit this necessary unique labeling

Definition 5 *Let $l : \mathbb{T}_{N,K} \to \mathbb{R}^{N \times M}$ be an identifier — with M-dimensional labels — such that*

$$\forall T \in T \in \mathbb{T}_{N,M} : \; l(\pi(T)) = \pi l(T)$$

We denote the set of tensors that are identifiable via l, that is, l-identifiable, as $\mathbb{T}^l_{N,K} \subset \mathbb{T}_{N,K}$ such that $\forall T \in \mathbb{T}^l_{N,K}$ the multiset $\{\{e_n^\top l(T) \in \mathbb{R}^M : n \in [N]\}\}$ consists of distinct elements where e_n is the n-th standard bais of $\mathbb{R}^N$ and $n \in [N]$.

In the first step of our approach, given a tensor and an identifier, we first construct a *set* that remains invariant with respect to the permutation of the node entities.

Definition 6 *Let $K, N \in \mathbb{N}$. For any l-identifiable tensor $T \in \mathbb{T}^l_{N,K}$, let $\alpha^K_{n_1 \ldots n_K}(T) = T_{n_1 \ldots n_K} \in \mathbb{R}^D$ for all $n_1, \ldots, n_K \in [N]$. Then, we define recursively that:*

$$\forall k \in K \text{ down to } 1, n_1, \ldots, n_{k-1} \in [N] : \; \alpha^{k-1}_{n_1 \ldots n_{k-1}}(T) = \{\big(e_{n_k}^\top l(T), \alpha^k_{n_1 \ldots n_k}(T)\big) : n_k \in [N]\}.$$

We define the set $S(T) = \{\big(e_{n_1}^\top l(T), \alpha^1_{n_1}(T)\big) : n_1 \in [N]\}$.

Proposition 4 *Let $K, N \in \mathbb{N}$ and $T, T' \in \mathbb{T}^l_{N,K}$. Then, we have $S(T) = S(T')$ if and only if $T' = \pi(T)$ for a permutation $\pi \in \Pi(N)$, that is, $T \equiv T'$.*

Proposition 4 establishes a bijection between identifiable tensors $\mathbb{T}^l N, K$ — up to a permutation factor — and sets in $S(\mathbb{T}^l N, K)$. In Theorem 7, we provide an algebraic characterization of (nonlinear) permutation-invariant tensor functions with distinct node features, showing that the sum-decomposable model is applicable only on identifiable tensors.

Theorem 7 *Let $K, N \in \mathbb{N}$. Let $f : \mathbb{T}_{N,K} \to \mathrm{codom}(f)$ be a permutation-invariant tensor function. Then we have*

$$\forall T \in \mathbb{T}^l_{N,K} : f(T) = \rho\Big(\sum_{n_1 \in [N]} \phi_1(e_{n_1}^\top l(T), \beta^1_{n_1}(T))\Big)$$

where $l : \mathbb{T}_{N,K} \to \mathrm{codom}(l)$ is an identifier function, $\beta^K_{n_1 \ldots n_K}(T) = T_{n_1 \ldots n_K} \in \mathbb{R}^D$ for all $n_1, \ldots, n_K \in [N]$, and

$$\forall k \in [K], n_1, \ldots, n_{k-1} \in [N] : \ \beta^{k-1}_{n_1 \ldots n_{k-1}}(T) = \sum_{n_k \in [N]} \phi_k(e_{n_k}^\top l(T), \beta^k_{n_1 \ldots n_k}(T)),$$

where ϕ_k is continuous over its compact domain and its codomain resides in $\mathbb{R}^{D_k}$ ($k \in [K]$), and

1. *$D_k = 2(M + D_{k+1})N$ if $\mathrm{codom}(l) \subset \mathbb{Q}^{N \times M}$*
2. *$D_k = \binom{N+D_{k+1}}{N} - 1$ if $\mathrm{codom}(l) \subset \mathbb{R}^{N \times M}$*

for all $k \in [K-1]$ and $D_K = D$. The function ρ is defined on $\mathbb{D} \subset \mathbb{R}^{D_1}$ where

$$\mathbb{D} = \{ \sum_{n_1 \in [N]} \phi_1(e_{n_1}^\top l(T), \beta^1_{n_1}(T)) : T \in \mathbb{T}^l_{N,K}\},$$

and it is not guaranteed to have a continuous extension to $\mathbb{R}^{D_1}$.

6. Conclusion

In this work, we present several contributions concerning the universal representation theory of multiset functions and permutation-invariant tensor functions. We show the existence of a universal sum-decomposition model for multivariate multiset functions and provide the most optimal dimension bound for encoded multiset features available to date. Our extensive analyses are built upon the novel notion of ℓ-identifiable multisets, enabling us to uniquely label distinct elements within multisets. Furthermore, our proposed decomposable model for permutation-invariant tensor functions extends existing models designed for linear permutation invariant tensor functions, which are commonly employed as layers in IGNs. It is worth noting that our universal representation via sum-decomposables is stronger than the concept of universal approximation, as the former imply the latter. All these findings lead to universal approximation results for multiset (or tensor) functions using sum-decomposables—a proposition that indicates natural architectures for neural networks, akin to Deep Sets.

Acknowledgments

This work is partially supported by National Science Foundation (NSF) under grants CCF-2112665 and CCF-2310411. We thank Mohammad Fereydounian and Zhengchao Wan, whose insightful discussions, comments, and inputs helped with the development of this paper.

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. *arXiv preprint arXiv:2306.06529*, 2023.

Mary P Attenborough. *Mathematics for electrical engineering and computing*. Elsevier, 2003.

Waiss Azizian and Marc Lelarge. Expressive power of invariant and equivariant graph neural networks. *arXiv preprint arXiv:2006.15646*, 2020.

Dominique Beaini, Saro Passaro, Vincent Létourneau, Will Hamilton, Gabriele Corso, and Pietro Liò. Directional graph networks. In *International Conference on Machine Learning*, pages 748–758. PMLR, 2021.

Beatrice Bevilacqua, Fabrizio Frasca, Derek Lim, Balasubramaniam Srinivasan, Chen Cai, Gopinath Balamurugan, Michael M Bronstein, and Haggai Maron. Equivariant subgraph aggregation networks. *arXiv preprint arXiv:2110.02910*, 2021.

Rasmus Bro, Evrim Acar, and Tamara G Kolda. Resolving the sign ambiguity in the singular value decomposition. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 22(2):135–140, 2008.

Chen Cai and Yusu Wang. Convergence of invariant graph networks. In *International Conference on Machine Learning*, pages 2457–2484. PMLR, 2022.

Yu Chen, Lingfei Wu, and Mohammed J Zaki. Reinforcement learning based graph-to-sequence model for natural question generation. *arXiv preprint arXiv:1908.04942*, 2019.

Branko Ćurgus and Vania Mascioni. Roots and polynomials as homeomorphic spaces. *Expositiones Mathematicae*, 24(1):81–95, 2006.

Dragos Cvetkovic, Dragoš M Cvetković, Peter Rowlinson, and Slobodan Simic. *Eigenspaces of graphs*. Number 66. Cambridge University Press, 1997.

Klaus Deimling. *Nonlinear functional analysis*. Courier Corporation, 2010.

Vijay Prakash Dwivedi and Xavier Bresson. A generalization of transformer networks to graphs. *arXiv preprint arXiv:2012.09699*, 2020.

Vijay Prakash Dwivedi, Chaitanya K Joshi, Thomas Laurent, Yoshua Bengio, and Xavier Bresson. Benchmarking graph neural networks. 2020.

Vijay Prakash Dwivedi, Anh Tuan Luu, Thomas Laurent, Yoshua Bengio, and Xavier Bresson. Graph neural networks with learnable structural and positional representations. *arXiv preprint arXiv:2110.07875*, 2021.

Nadav Dym and Steven J Gortler. Low dimensional invariant embeddings for universal geometric learning. *arXiv preprint arXiv:2205.02956*, 2022.

HT Eastment and WJ Krzanowski. Cross-validatory choice of the number of components from a principal component analysis. *Technometrics*, 24(1):73–77, 1982.

Ryszard Engelking. General topology. *Sigma series in pure mathematics*, 6, 1989.

SM Eslami, Nicolas Heess, Theophane Weber, Yuval Tassa, David Szepesvari, Geoffrey E Hinton, et al. Attend, infer, repeat: Fast scene understanding with generative models. *Advances in neural information processing systems*, 29, 2016.

William Feller. An introduction to probability theory and its applications. Technical report, Wiley series in probability and mathematical statistics, 3rd edn.(Wiley, New ..., 1967.

Mohammad Fereydounian, Hamed Hassani, Javid Dadashkarimi, and Amin Karbasi. The exact class of graph functions generated by graph neural networks. *arXiv preprint arXiv:2202.08833*, 2022.

Vikas Garg, Stefanie Jegelka, and Tommi Jaakkola. Generalization and representational limits of graph neural networks. In *International Conference on Machine Learning*, pages 3419–3430. PMLR, 2020.

Snir Hordan, Tal Amir, Steven J Gortler, and Nadav Dym. Complete neural networks for euclidean graphs. *arXiv preprint arXiv:2301.13821*, 2023.

Chaitanya K Joshi, Cristian Bodnar, Simon V Mathis, Taco Cohen, and Pietro Lio. On the expressive power of geometric graph neural networks. *arXiv preprint arXiv:2301.09308*, 2023.

Nicolas Keriven and Gabriel Peyré. Universal invariant and equivariant graph neural networks. *Advances in Neural Information Processing Systems*, 32, 2019.

Jinwoo Kim, Dat Nguyen, Seonwoo Min, Sungjun Cho, Moontae Lee, Honglak Lee, and Seunghoon Hong. Pure transformers are powerful graph learners. *Advances in Neural Information Processing Systems*, 35:14582–14595, 2022.

Adam Kosiorek, Hyunjik Kim, Yee Whye Teh, and Ingmar Posner. Sequential attend, infer, repeat: Generative modelling of moving objects. *Advances in Neural Information Processing Systems*, 31, 2018.

Devin Kreuzer, Dominique Beaini, Will Hamilton, Vincent Létourneau, and Prudencio Tossou. Rethinking graph transformers with spectral attention. *Advances in Neural Information Processing Systems*, 34:21618–21629, 2021.

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 *International conference on machine learning*, pages 3744–3753. PMLR, 2019.

Yaguang Li, Yishuang Lin, Meghna Madhusudan, Arvind Sharma, Wenbin Xu, Sachin S Sapatnekar, Ramesh Harjani, and Jiang Hu. A customized graph neural network model for guiding analog ic placement. In *Proceedings of the 39th International Conference on Computer-Aided Design*, pages 1–9, 2020.

Derek Lim, Joshua Robinson, Lingxiao Zhao, Tess Smidt, Suvrit Sra, Haggai Maron, and Stefanie Jegelka. Sign and basis invariant networks for spectral graph representation learning. *arXiv preprint arXiv:2202.13013*, 2022.

Yi-Chen Lu, Sai Surya Kiran Pentapati, Lingjun Zhu, Kambiz Samadi, and Sung Kyu Lim. Tp-gnn: A graph neural network framework for tier partitioning in monolithic 3d ics. In *2020 57th ACM/IEEE Design Automation Conference (DAC)*, pages 1–6. IEEE, 2020.

Tengfei Ma, Jie Chen, and Cao Xiao. Constrained generation of semantically valid graphs via regularizing variational autoencoders. *Advances in Neural Information Processing Systems*, 31, 2018.

Haggai Maron, Heli Ben-Hamu, Nadav Shamir, and Yaron Lipman. Invariant and equivariant graph networks. *arXiv preprint arXiv:1812.09902*, 2018.

Haggai Maron, Heli Ben-Hamu, Hadar Serviansky, and Yaron Lipman. Provably powerful graph networks. *Advances in neural information processing systems*, 32, 2019a.

Haggai Maron, Ethan Fetaya, Nimrod Segol, and Yaron Lipman. On the universality of invariant networks. In *International conference on machine learning*, pages 4363–4371. PMLR, 2019b.

Grégoire Mialon, Dexiong Chen, Margot Selosse, and Julien Mairal. Graphit: Encoding graph structure in transformers. *arXiv preprint arXiv:2106.05667*, 2021.

Christopher Morris, Martin Ritzert, Matthias Fey, William L Hamilton, Jan Eric Lenssen, Gaurav Rattan, and Martin Grohe. Weisfeiler and leman go neural: Higher-order graph neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 4602–4609, 2019.

Krikamol Muandet, Kenji Fukumizu, Francesco Dinuzzo, and Bernhard Schölkopf. Learning from distributions via support measure machines. *Advances in neural information processing systems*, 25, 2012.

Krikamol Muandet, David Balduzzi, and Bernhard Schölkopf. Domain generalization via invariant feature representation. In *International conference on machine learning*, pages 10–18. PMLR, 2013.

Ryan L Murphy, Balasubramaniam Srinivasan, Vinayak Rao, and Bruno Ribeiro. Janossy pooling: Learning deep permutation-invariant functions for variable-size inputs. *arXiv preprint arXiv:1811.01900*, 2018.

Michelle Ntampaka, Hy Trac, Dougal J Sutherland, Sebastian Fromenteau, Barnabás Póczos, and Jeff Schneider. Dynamical mass measurements of contaminated galaxy clusters using machine learning. *The Astrophysical Journal*, 831(2):135, 2016.

Junier Oliva, Barnabás Póczos, and Jeff Schneider. Distribution to distribution regression. In *International Conference on Machine Learning*, pages 1049–1057. PMLR, 2013.

Maks Ovsjanikov, Jian Sun, and Leonidas Guibas. Global intrinsic symmetries of shapes. In *Computer graphics forum*, volume 27, pages 1341–1348. Wiley Online Library, 2008.

Barnabás Póczos, Aarti Singh, Alessandro Rinaldo, and Larry Wasserman. Distribution-free distribution regression. In *Artificial Intelligence and Statistics*, pages 507–515. PMLR, 2013.

Sergey N Pozdnyakov and Michele Ceriotti. Incompleteness of graph neural networks for points clouds in three dimensions. *Machine Learning: Science and Technology*, 3(4):045020, 2022.

Charles Chapman Pugh and CC Pugh. *Real mathematical analysis*, volume 2011. Springer, 2002.

Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017a.

Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30, 2017b.

Siamak Ravanbakhsh. Universal equivariant multilayer perceptrons. In *International Conference on Machine Learning*, pages 7996–8006. PMLR, 2020.

Siamak Ravanbakhsh, Junier Oliva, Sebastian Fromenteau, Layne Price, Shirley Ho, Jeff Schneider, and Barnabás Póczos. Estimating cosmological parameters from the dark matter distribution. In *International Conference on Machine Learning*, pages 2407–2416. PMLR, 2016a.

Siamak Ravanbakhsh, Jeff Schneider, and Barnabas Poczos. Deep learning with sets and point clouds. *arXiv preprint arXiv:1611.04500*, 2016b.

Raif M Rustamov et al. Laplace-beltrami eigenfunctions for deformation invariant shape representation. In *Symposium on geometry processing*, volume 257, pages 225–233, 2007.

David Rydh. A minimal set of generators for the ring of multisymmetric functions. In *Annales de l'institut Fourier*, volume 57, pages 1741–1769, 2007.

Nimrod Segol and Yaron Lipman. On universal equivariant set networks. In *International Conference on Learning Representations*, 2019.

Raymond Séroul. *Programming for mathematicians*. Springer Science & Business Media, 2012.

John Shawe-Taylor. Symmetries and discriminability in feedforward network architectures. *IEEE Transactions on Neural Networks*, 4(5):816–826, 1993.

Elias M Stein and Rami Shakarchi. *Complex analysis*, volume 2. Princeton University Press, 2010.

Peter Sunehag, Guy Lever, Audrunas Gruslys, Wojciech Marian Czarnecki, Vinicius Zambaldi, Max Jaderberg, Marc Lanctot, Nicolas Sonnerat, Joel Z Leibo, Karl Tuyls, et al. Value-decomposition networks for cooperative multi-agent learning. *arXiv preprint arXiv:1706.05296*, 2017.

Wilson A Sutherland. *Introduction to metric and topological spaces*. Oxford University Press, 2009.

Zoltán Szabó, Bharath K Sriperumbudur, Barnabás Póczos, and Arthur Gretton. Learning theory for distribution regression. *The Journal of Machine Learning Research*, 17(1):5272–5311, 2016.

Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17:395–416, 2007.

Edward Wagstaff, Fabian Fuchs, Martin Engelcke, Ingmar Posner, and Michael A Osborne. On the limitations of representing functions on sets. In *International Conference on Machine Learning*, pages 6487–6494. PMLR, 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(151): 1–56, 2022.

Peihao Wang, Shenghao Yang, Shu Li, Zhangyang Wang, and Pan Li. Polynomial width is sufficient for set representation with high-dimensional features. *arXiv preprint arXiv:2307.04001*, 2023.

Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. Kgat: Knowledge graph attention network for recommendation. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 950–958, 2019.

Zhiyao Xie, Rongjian Liang, Xiaoqing Xu, Jiang Hu, Yixiao Duan, and Yiran Chen. Net2: A graph attention network method customized for pre-placement net length estimation. In *Proceedings of the 26th Asia and South Pacific Design Automation Conference*, pages 671–677, 2021.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*, 2018.

Xu Yang, Kaihua Tang, Hanwang Zhang, and Jianfei Cai. Auto-encoding scene graphs for image captioning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10685–10694, 2019.

Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R Salakhutdinov, and Alexander J Smola. Deep sets. *Advances in neural information processing systems*, 30, 2017.

Keren Zhu, Mingjie Liu, Hao Chen, Zheng Zhao, and David Z Pan. Exploring logic optimizations with reinforcement learning and graph convolutional network. In *Proceedings of the 2020 ACM/IEEE Workshop on Machine Learning for CAD*, pages 145–150, 2020.

Aaron Zweig and Joan Bruna. Exponential separations in symmetric neural networks. *Advances in Neural Information Processing Systems*, 35:33134–33145, 2022.

Appendix A. Theorem 8 and Its Proof

Theorem 8 ***Any*** *multivariate multiset function* $f : \mathbb{X}_{\mathbb{R}^D,N} \to \mathrm{codom}(f)$ *— over a multisets of* N *elements in* $\mathbb{R}^D$ *— is sum-decomposable via* $\mathbb{R}^{\binom{N+D}{D}-1}$*, that is,*

$$\forall X \in \mathbb{X}_{\mathbb{R}^D,N} :\ f(X) = \rho \circ \Phi(X),\ \textit{where}\ \Phi(X) \stackrel{\text{def}}{=} \sum_{x \in X} \phi(x),$$

where $\phi : \mathbb{R}^D \to \mathrm{codom}(\phi) \subseteq \mathbb{R}^{\binom{N+D}{D}-1}$ *is a continuous function and* $\rho : \mathrm{codom}(\Phi) \to \mathrm{codom}(f)$.

Note that compared to Theorem 3, the function f in Theorem 8 is not necessarily continuous, and the decoder function ρ is not necessarily continuous as well.

A.1. Proof

Let $N, D \in \mathbb{N}$. We want to prove that for any multivariate multiset function $f : \mathbb{X}_{\mathbb{R}^D,N} \to \mathrm{codom}(f)$, there exists a sum-decomposition via $\mathbb{R}^{\binom{N+D}{D}-1}$.

Trivial case of $\mathbf{N = 1}$. We define functions ϕ and ρ as follows:

$$\forall x \in \mathbb{R}^D : \phi(x) = x, \text{ and } \rho(x) = f(\{\{x\}\}),$$

where $\mathrm{dom}(\phi) = \mathbb{R}^D$. Since $\Phi(\{\{x\}\}) = \phi(x)$, $\mathrm{codom}(\rho) = \mathrm{codom}(\Phi) = \mathrm{codom}(\phi) = \mathbb{R}^D = \mathbb{R}^{\binom{1+D}{D}-1}$, $\mathrm{codom}(\rho) = \mathrm{codom}(f)$, and $f(\{\{x\}\}) = \rho \circ \Phi(\{\{x\}\})$, we arrive at the theorem's statement for $N = 1$.

Remark 7 *In our notation, depending on the context,* x_n *can mean either (1) the* n*-th coordinate (element) of vector* x *(say in* $\mathbb{R}^D$*) or (2) a vector indexed by* n*, for example,* $x_1, \ldots, x_N \in \mathbb{R}^D$*. In the latter case, we do emphasize the domain of the vector a priori, that is,* $x_n \in \mathbb{R}^D$*.*

General case of $\mathbf{N \geq 2}$. We break down our approach into two steps:

1. We show that there exists a function $\phi : \mathbb{R}^D \to \mathrm{codom}(\phi) \subseteq \mathbb{R}^{\binom{N+D}{D}-1}$ such that $\Phi(X) = \sum_{x \in X} \phi(x)$ is an injective multiset function, that is, Φ^{-1} is well-defined on $\mathrm{codom}(\Phi)$.
2. Let $\rho = f \circ \Phi^{-1}$. This immediately proves $f = \rho \circ \Phi(X) = \rho\big(\sum_{x \in X} \phi(x)\big)$.

This is an extension to the existing univariate result (that is, $D = 1$); refer Theorem 2 in (Zaheer et al., 2017). In the one-dimensional case, Zaheer et al. (2017) prove that Φ is an invertible function by showing that, given $\Phi(X)$, one can construct a univariate polynomial $p(t; \Phi(X))$ whose roots are X, that is, $\Phi^{-1} \circ \Phi(X) = \mathrm{roots} \circ p(t; \Phi(X)) = X$ where roots returns the multiset of roots of a polynomial equation. Moreover, the appropriate choice for the basis function ϕ — which makes this analysis tractable — gives a bound for the latent dimension, that is, dimension of the ambient vector space containing $\mathrm{codom}(\Phi)$.

In our approach, we arrive at the appropriate choice for ϕ constructing a **multivariate** polynomial whose parameterized roots are **related** to X. In what follows, we (1) introduce the basis function ϕ, (2) construct an appropriate multivariate polynomial $p(t; z, \Phi(X))$ — parameterized by both $t \in \mathbb{R}$ and $z \in \mathbb{R}^D$ — and (3) extract X from its **parameterized** roots. In step (3), we introduce novel

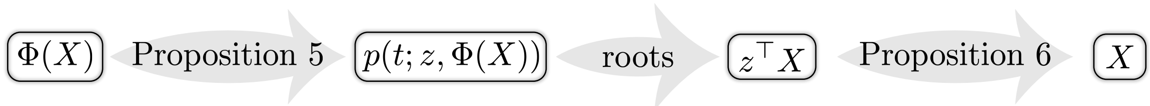

Figure 1: Proof sketch for the injectivity of Φ.

techniques for analyzing parameterized multisets — akin to computing directional derivatives for multivariate functions. We summarize these steps in Figure 1.

The following definition introduces several frequently used functions in this proof.

Definition 7 *For any multiset of real scalars $X = \{\{x_n \in \mathbb{R} : n \in [N]\}\}$ where $N \geq 2$, we let*

$$\mathrm{gap}(X) = \min_{\substack{n,n' \in [N] \\ x_n \neq x'_n}} |x_n - x_{n'}|, \qquad \mathrm{diam}(X) = \max_{\substack{n,n' \in [N] \\ n \neq n'}} |x_n - x_{n'}|,$$

$$\mathrm{unique}(X) = \{x_n : n \in [N]\}, \qquad \mathrm{sort}(X) = \big(x_{\pi(n)}\big)_{n \in [N]} \in \mathbb{R}^N,$$

where $\pi : [N] \to [N]$ is a permutation operator such that $x_{\pi(1)} \geq x_{\pi(2)} \geq \cdots \geq x_{\pi(N)}$.

Remark 8 *If $x_n, x_{n'} \in X$ where $x_n = x_{n'}$ for distinct $n, n' \in [N]$, then the permutation operator π in Definition 7 is not unique; but any such permutation π results in the same sorted vector $(x_{\pi(n)})_{n \in [N]}$. Hence,* $\mathrm{sort}(X)$ *is well-defined for any multiset of real-valued scalars X.*

Remark 9 *Let X be a multiset of real scalars. Then,* $\mathrm{gap}(X)$ *is well-defined only if the cardinality of* $\mathrm{unique}(X)$ *is strictly greater than one, that is,* $|\mathrm{unique}(X)| > 1$.

We consider a class of multivariate polynomials paramterized with $t \in \mathbb{R}$ and $z \in \mathbb{R}^D$. In Proposition 5, we introduce a function ϕ that enables us to construct each polynomial — in the aforementioned class — using only $\Phi(X) = \sum_{x \in X} \phi(x)$. In other words, knowing t, z and $\Phi(X)$, we can represent the polynomial $\prod_{x \in X}(t - z^\top x)$. This allows us to write the polynomial $\prod_{x \in X}(t - z^\top x)$ as a function depending on variables t, z and $\Phi(X)$, which we call $p(t; z, \Phi(X))$.

Proposition 5 *Let $N, D \in \mathbb{N}$ and $\phi : \mathbb{R}^D \to \mathrm{codom}(\phi) \subseteq \mathbb{R}^{\binom{N+D}{D}-1}$ be the following continuous function:*

$$\forall x \in \mathbb{R}^D : \ \phi(x) = \Big(\prod_{d=1}^{D} x_d^{k_d}\Big)_{k \in \mathcal{K}_N^D} \in \mathbb{R}^{\binom{N+D}{D}-1},$$

where $k = (k_d)_{d \in [D]}$ is a D-tuple and $\mathcal{K}_N^D = \{(k_d)_{d \in [D]} : k_1 + \ldots + k_D \in [N], k_1, \ldots, k_D \geq 0\}$. Then, for all $X \in \mathbb{X}_{\mathbb{R}^D, N}$, $\Phi(X) = \sum_{x \in X} \phi(x)$ suffices to construct the following multivariate polynomial:

$$\forall t \in \mathbb{R}, z \in \mathbb{R}^D : \ \prod_{x \in X} (t - z^\top x) = p\big(t; z, \Phi(X)\big). \tag{3}$$

To show that Φ is an invertible function, we want to argue that the multiset X can be uniquely recovered form the multivariate polynomial in equation (3), that is, $p\big(t; z, \Phi(X)\big)$. Let us proceed with the following definitions.

Definition 8 *For any $z \in \mathbb{R}^D$, multiset X of at least two D-dimensional vectors, and multivariate polynomial $p\big(t; z, \Phi(X)\big)$ in the equation (3), we formalize the following functions:*

- $\mathrm{roots} \circ p\big(t; z, \Phi(X)\big) = \{\{t : p\big(t; z, \Phi(X)\big) = 0\}\} = \{\{z^\top x : x \in X\}\} \stackrel{\text{def}}{=} z^\top X$
- $\mathrm{separators} \circ p\big(t; z, \Phi(X)\big) = \mathrm{argmax}_{z \in \mathbb{R}^D} |\mathrm{unique} \circ \mathrm{roots} \circ p\big(t; z, \Phi(X)| \,,$

where $|\cdot|$ returns the cardinality of its input set.

Definition 9 *Let X be a multiset of at least two D-dimensional vectors. If exists, the directional derivative of $\mathrm{sort}\big(z^\top X\big)$ — where $z^\top X = \{\{z^\top x : x \in X\}\}$ — at $z \in \mathbb{R}^D$ in the direction of unit norm $v \in \mathbb{R}^D$ is given as follows:*

$$\nabla_v \mathrm{sort}\big(z^\top X\big) = \lim_{\delta \to 0} \frac{1}{\delta}\Big(\mathrm{sort}\big((z + \delta v)^\top X\big) - \mathrm{sort}\big(z^\top X\big)\Big). \tag{4}$$

In Proposition 6, we show how to retrieve X from $\mathrm{roots} \circ p\big(t; z, \Phi(X)\big)$, that is, the parameterized multiset $z^\top X$.

Proposition 6 *For any $z \in \mathbb{R}^D$, multiset $X = \{\{x_n \in \mathbb{R}^D : n \in [N]\}\}$ where $N \geq 2$, and the multivariate polynomial $p\big(t; z, \Phi(X)\big)$ in the equation (3), we have*

$$\mathrm{separators} \circ p\big(t; z, \Phi(X)\big) \neq \emptyset.$$

Moreover, for any $z^* \in \mathrm{separators} \circ p\big(t; z, \Phi(X)\big)$, the directional derivative of $\mathrm{sort} \circ \mathrm{roots} \circ p\big(t; z, \Phi(X)\big)$ is well-defined and we have:

$$\forall d \in [D] : \nabla_{e_d} \mathrm{sort} \circ \mathrm{root} \circ p\big(t; z, \Phi(X)|_{z=z^*} = (e_d^\top x_{\pi_{z^*}(n)})_{n \in [N]},$$

where $e_d \in \mathbb{R}^D$ is the d-th standard basis vector for $\mathbb{R}^D$ ($d \in [D]$), and $\pi_{z^*} : [N] \to [N]$ is a permutation operator that sorts the elements ${z^*}^\top X$ — see Definition 7.

In summary, given $\Phi(X) \in \mathbb{R}^{\binom{N+D}{D}-1}$, we can construct a multivariate polynomial $p\big(t; z, \Phi(X)\big)$ with parameterized roots $z^\top X$; see Proposition 5. Then, we can pick a fixed vector $z^* \in \mathrm{separators} \circ p\big(t; z, \Phi(X)\big) \neq \emptyset$; see Definition 8 and Proposition 6. We then prove the following result:

$$\forall d \in [D] : \nabla_{e_d} \mathrm{sort} \circ \mathrm{root} \circ p\big(t; z, \Phi(X)|_{z=z^*} = (e_d^\top x_{\pi_{z^*}(n)})_{n \in [N]},$$

where ∇_{e_d} computes the directional derivative (see Definition 9) in the direction of e_d — the d-th standard basis of $\mathbb{R}^D$ — for $d \in [D]$, and $x_n \in \mathbb{R}^D$ is an element of X indexed by n. We retrieve X as follows:

$$\begin{aligned}\{\{(e_d^\top x_{\pi_{z^*}(n)})_{d \in [D]} \in \mathbb{R}^D : n \in [N]\}\} &= \{\{(e_d^\top x_n)_{d \in [D]} \in \mathbb{R}^D : n \in [N]\}\} \\ &= X.\end{aligned}$$

This result does not depend on the specific choices for the permutation operator π_{z^*} and $z^* \in \mathrm{separators} \circ p\big(t; z, \Phi(X)\big)$. Therefore, Φ is an invertible multiset function, that is,

$$\Phi^{-1} \circ \Phi(X) = \{\{ \begin{pmatrix} \big(\nabla_{e_1} \mathrm{sort} \circ \mathrm{root} \circ p\big(t; z, \Phi(X)|_{z=z^*}\big)_n \\ \vdots \\ \big(\nabla_{e_D} \mathrm{sort} \circ \mathrm{root} \circ p\big(t; z, \Phi(X)|_{z=z^*}\big)_n \end{pmatrix} \in \mathbb{R}^D : n \in [N]\}\}, \tag{5}$$

where $z^* \in \text{separators} \circ p\big(t; z, \Phi(X)\big)$, and subscript n denotes the n-the element of the N-dimensional vectors. The function Φ^{-1} is only well-defined on $\text{codom}(\Phi)$; see equation (5).
Now we let $\rho : \text{codom}(\Phi) \to \text{codom}(f)$ where

$$\forall y \in \text{codom}(\Phi) \subseteq \mathbb{R}^{\binom{N+D}{D}-1} : \rho(y) = f \circ \Phi^{-1}(y).$$

This proves the sum-decomposition representation claim of the theorem, that is, $f = \rho \circ \Phi$. In Appendices A.2 and A.3 we provide proofs of Propositions 5 and 6. In Appendix A.4, we provide two illustrative examples on computing Φ^{-1}.

A.2. Proof of Proposition 5

We expand the expression in equation (3) as follows:

$$\forall t \in \mathbb{R}, z \in \mathbb{R}^D : \prod_{x \in X} (t - z^\top x) = t^N + \sum_{n \in [N]} (-1)^n a_n(z; X) t^{N-n} \tag{6}$$

where each coefficient $a_n(z; X)$ is determined using the Newton-Girard formulae (Séroul, 2012), that is,

$$a_n(z; X) = \frac{1}{n} \det \begin{pmatrix} E_1(z; X) & 1 & 0 & \cdots & 0 \\ E_2(z; X) & E_1(z; X) & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ E_n(z; X) & E_{n-1}(z; X) & E_{n-2}(z; X) & \cdots & E_1(z; X) \end{pmatrix} \tag{7}$$

for all $n \in [N]$ and $z \in \mathbb{R}^D$, and $E_n(z; X) = \sum_{x \in X} (z^\top x)^n$. Therefore, each coefficient $a_n(z; X)$ is a polynomial function of $\{E_n(z; X)\}_{n=1}^N$ — moments of the parameterized multiset $\{\{z^\top x : x \in X\}\} \stackrel{\text{def}}{=} z^\top X$. Lemma 2 lets us relate each moment to the elementary symmetric polynomials.

Lemma 2 *For any $k_1, \cdots, k_D \in \mathbb{N} \cup \{0\}$ and $n \in \mathbb{N}$, let*

$$\binom{n}{k_1, \ldots, k_D}^{\text{ind}} = \begin{cases} \frac{n!}{k_1! \cdots k_D!} & \textit{if } k_1 + \cdots + k_D = n \\ 0 & \textit{otherwise.} \end{cases}$$

Let $x, z \in \mathbb{R}^D$ and $n \in [N]$. Then, we have $(z^\top x)^n = \langle \psi(z, n), \phi(x) \rangle$ such that

$$\psi(z, n) = \Big(\binom{n}{k_1, \ldots, k_D}^{\text{ind}} \prod_{d=1}^D z_d^{k_d}\Big)_{k \in \mathcal{K}_N^D}, \quad \phi(x) = \big(\prod_{d=1}^D x_d^{k_d}\big)_{k \in \mathcal{K}_N^D} \in \mathbb{R}^{\binom{N+D}{D}-1}, \tag{8}$$

where $k = (k_d)_{d \in [D]}$ and $\mathcal{K}_N^D = \{(k_d)_{d \in [D]} : k_1 + \ldots + k_D \in [N], k_1, \ldots, k_D \geq 0\}$.

Proof Let $x, z \in \mathbb{R}^D$ and $n \in [N]$. Then, we have

$$(z^\top x)^n = (\sum_{d \in [D]} z_d x_d)^n = \sum_{k_1 + \ldots + k_D = n} \binom{n}{k_1, \ldots, k_D} \prod_{d=1}^D z_d^{k_d} \prod_{d=1}^D x_d^{k_d} = \langle \psi(z, n), \phi(x) \rangle$$

where $\phi(x)$ and $\psi(z, n)$ are given in equation (8). Since the dimension of $\phi(x)$ — the size of $\mathcal{K}_N^D$ — is equivalent to the number of solutions to the following problem:

$$k_1, \ldots, k_D \in \mathbb{N} \cup \{0\} : 1 \leq \sum_{d=1}^{D} k_d \leq N. \tag{9}$$

We can transform the problem in equation (9) to the following form:

$$k_1, \ldots, k_D, k_\circ \in \mathbb{N} \cup \{0\}, k_\circ \neq N : \sum_{d=1}^{D} k_d + k_\circ = N. \tag{10}$$

In the occupancy problem, we ask: *how many ways can one distribute N indistinguishable objects into $D+1$ distinguishable bins?* The number of nonnegative solutions are $\binom{N+D}{D}$; refer to (Feller, 1967), section 5. However, if $k_\circ = N$, then $k_1 = k_2 = \cdots = k_D = 0$ which is not allowed. If we exclude this case, we arrive at $\binom{N+D}{D} - 1$ integer solutions for problems in equations (9) and (10) ■

Let us now prove the proposition's statement. Given $\Phi(X) = \sum_{x \in X} \phi(x)$, we compute

$$\forall z \in \mathbb{R}^D, n \in [N] : E_n(z; X) = \sum_{x \in X} \langle z, x \rangle^n = \sum_{x \in X} \langle \psi(z, n), \phi(x) \rangle = \langle \psi(z, n), \Phi(X) \rangle,$$

that are, all parameterized moments required to construct $\prod_{x \in X}(t - x^\top z)$ — refer to Lemma 2, and equation (7). Therefore, we can uniquely identify the polynomial in equation (3) with only $\Phi(X)$.

A.3. Proof of Proposition 6

Proposition 7 *For any $z \in \mathbb{R}^D$, multiset $X = \{\{x_n \in \mathbb{R}^D : n \in [N]\}\}$ where $N \geq 2$, and the multivariate polynomial $p\big(t; z, \Phi(X)\big)$ in the equation (3), we have*

$$\text{separators} \circ p\big(t; z, \Phi(X)\big) \neq \emptyset.$$

Moreover, for any $z^* \in$ separators $\circ\, p\big(t; z, \Phi(X)\big)$, the directional derivative of sort $\circ$ roots $\circ$ $p\big(t; z, \Phi(X)\big)$ *is well-defined and we have:*

$$\forall d \in [D] : \nabla_{e_d} \text{sort} \circ \text{root} \circ p\big(t; z, \Phi(X)|_{z=z^*} = (e_d^\top x_{\pi_{z^*}(n)})_{n \in [N]},$$

where $e_d \in \mathbb{R}^D$ is the d-th standard basis vector for $\mathbb{R}^D$ ($d \in [D]$), and $\pi_{z^*} : [N] \to [N]$ is a permutation operator that sorts the elements ${z^*}^\top X$ — see Definition 7.

For any $z \in \mathbb{R}^D$ and multiset $X = \{\{x_n \in \mathbb{R}^D : n \in [N]\}\}$ where $N \geq 2$, we have

$$\text{sort} \circ \text{root} \circ p(t; z, X) = (z^\top x_{\pi_z(n)})_{n \in [N]} \in \mathbb{R}^N$$

where $\pi_z : [N] \to [N]$ is a permutation operator such that $z^\top x_{\pi_z(1)} \geq z^\top x_{\pi_z(2)} \geq \cdots \geq z^\top x_{\pi_z(N)}$; see Definition 7. Given such an ordered list, we want to retrieve the multiset X. **If** the order of the elements of X after sorting remains unchanged for a perturbed parameter $z + \delta e_d$ — where

$e_d \in \mathbb{R}^D$ is the d-th standard basis for $\mathbb{R}^D$ and small enough $\delta \in \mathbb{R}$, that is, $x_{\pi_{z+\delta e_d}(n)} = x_{\pi_z(n)}$ for all $n \in [N]$ and $d \in [D]$ — then we have the following equality:

$$\begin{aligned}\frac{1}{\delta}\Big(\mathrm{sort}((z+\delta e_d)^\top X) - \mathrm{sort}(z^\top X)\Big) &= \frac{1}{\delta}\Big((z+\delta e_d)^\top x_{\pi_{z+\delta e_d}(n)} - z^\top x_{\pi_z(n)}\Big)_{n\in[N]} \\ &\overset{(a)}{=} \frac{1}{\delta}\Big(z^\top x_{\pi_z(n)} + \delta e_d^\top x_{\pi_z(n)} - z^\top x_{\pi_z(n)}\Big)_{n\in[N]} \\ &= \frac{1}{\delta}(\delta e_d^\top x_{\pi_z(n)})_{n\in[N]} = (e_d^\top x_{\pi_z(n)})_{n\in[N]},\end{aligned}$$

where (a) is due to our assumption $x_{\pi_{z+\delta e_d}(n)} = x_{\pi_z(n)}$ for all $n \in [N]$ and $d \in [D]$. If this property holds true, we can compute the following limit:

$$\begin{aligned}&\lim_{\delta\to} \frac{1}{\delta}\Big(\mathrm{sort}\circ\mathrm{root}\circ p(t; z+\delta e_d, X) - \mathrm{sort}\circ\mathrm{root}\circ p(t; z, X)\Big) \\ &\qquad\qquad = \lim_{\delta\to} \frac{1}{\delta}\Big(\mathrm{sort}((z+\delta e_d)^\top X) - \mathrm{sort}(z^\top X)\Big),\end{aligned} \tag{11}$$

to retrieve the d-component of the elements in X up to a fixed but unknown permutation π_z that does not depend on e_d — that is, $(e_d^\top x_{\pi_z(n)})_{n\in[N]}$ — for all $d \in [D]$. The limit in equation (11) is well-defined and returns $(e_d^\top x_{\pi_z(n)})_{n\in[N]}$ if there exists a vector $z \in \mathbb{R}^D$ such that it admits a solution for the following feasibility problem:

$$\text{find } \delta^* > 0 \text{ such that } x_{\pi_{z+\delta e_d}(n)} = x_{\pi_z(n)}, \quad \text{for all } n \in [N], d \in [D], \delta \le \delta^*.$$

As we shall see, any vector $z^* \in \mathrm{separators}\circ p\big(t; z, \Phi(X)\big)$ admits a solution to the aforementioned problem. To prove this result, we first need to derive the following property for the separators.

Lemma 3 *For any $z \in \mathbb{R}^D$, multiset X of at least two D-dimensional vectors, and the multivariate polynomial $p\big(t; z, \Phi(X)\big)$ in the equation (3), we have* $\mathrm{separators}\circ p\big(t; z, \Phi(X)\big)$ *is a nonempty subset of $\mathbb{R}^D$ and for all $z^* \in \mathrm{separators}\circ p\big(t; z, \Phi(X)\big)$, we have*

$$\begin{aligned}|\mathrm{unique}\circ\mathrm{roots}\circ p\big(t; z^*, \Phi(X)\big) &= \max\nolimits_{z\in\mathbb{R}^D}|\mathrm{unique}\circ\mathrm{roots}\circ p\big(t; z, \Phi(X)\big) \\ &= |\mathrm{unique}(X)|.\end{aligned}$$

Proof If $|\mathrm{unique}(X)| = 1$, then $\mathrm{separators}\circ p\big(t; z, \Phi(X)\big) = \mathbb{R}^D$ and the statement is trivial. Therefore, in what follows, we assume $|\mathrm{unique}(X)| > 1$.
Let X be a multiset of (at least two distinct) D-dimensional vectors and $\mathrm{roots}\circ p\big(t; z, \Phi(X) = z^\top X$, for all $z \in \mathbb{R}^D$. If $x, x' \in X$ where $x \neq x'$, then we have $z^\top x = z^\top x'$ for $z \in (x - x')^\perp \subset \mathbb{R}^D$. Therefore, we have

$$\forall z \in \mathbb{R}^D : |\mathrm{unique}(z^\top X)| \le |\mathrm{unique}(X)|.$$

We can prove the claim if we show $|\mathrm{unique}(z^\top X)|$ achieves its upper bound $|\mathrm{unique}(X)|$ over a subset of $\mathbb{R}^D$ — namely, $\mathrm{separators}\circ p\big(t; z, \Phi(X)\big)$.
Let $P_{x,x'} = (x - x')^\perp$ for distinct $x, x' \in \mathrm{unique}(X)$ — that is, $x \neq x'$. By construction, $P_{x,x'}$ is a $(D-1)$-dimensional subspace since $x \neq x'$. Since $\mathrm{unique}(X)$ contains only distinct elements, we have

$$\forall z \in P_{x,x'} \iff \langle z, x - x'\rangle = 0,$$

for all distinct $x, x' \in \text{unique}(X)$. We now construct the following set:

$$P_X = \bigcup_{\substack{x,x' \in \text{unique}(X) \\ x \neq x'}} P_{x,x'},$$

which is a finite union of $(D-1)$-dimensional subspaces. Therefore, P_X can not be equal to $\mathbb{R}^D$, that is, $\mathbb{R}^D \setminus P_X$ is a nonempty set. For any $z^* \in \mathbb{R}^D \setminus P_X$, we have

$$\begin{aligned}
&\forall x, x' \in X, x \neq x' : \langle z^*, x - x' \rangle = {z^*}^\top x - {z^*}^\top x' \neq 0 \\
&\forall x, x' \in X, x = x' : \langle z^*, x - x' \rangle = {z^*}^\top x - {z^*}^\top x' = 0.
\end{aligned}$$

Hence, we have $|\text{unique}({z^*}^\top X)| = |\text{unique}(X)|$ for all $z^* \in \text{separators} \circ p\big(t; z, \Phi(X)\big)$ where $\text{separators} \circ p\big(t; z, \Phi(X)\big) = \mathbb{R}^D \setminus P_X$ — a nonempty subset of $\mathbb{R}^D$. ■

As a result of Lemma 3, we have

$$\forall z^* \in \text{separators} \circ p\big(t; z, \Phi(X)\big) : |\text{unique}({z^*}^\top X)| = |\text{unique}(X)|,$$

that is, repeated (or distinct) elements in ${z^*}^\top X$ correspond to identical (or distinct) elements in X. We now want to show that the following directional derivative is well-defined:

$$\nabla_v \text{sort}\big(z^\top X\big)|_{z=z^*} = \lim_{\delta \to 0} \frac{1}{\delta} \Big(\text{sort}\big((z^* + \delta v)^\top X\big) - \text{sort}\big({z^*}^\top X\big) \Big)$$

for all $z^* \in \text{separators} \circ p\big(t; z, \Phi(X)\big)$ and unit norm vector $v \in \mathbb{R}^D$.

We break down the rest of the proof in two cases.

Case 1: $|\text{unique}(\mathbf{X})| > 1$.

Limiting behavior of $(z^* + \delta v)^\top X$ as $\delta \to 0$.

Let x, x' be two distinct elements in $\text{unique}(X)$, that is, $\|x - x'\|_2 > 0$. If $z^* \in \text{separators} \circ p\big(t; z, \Phi(X)\big)$, then we have $|{z^*}^\top x - {z^*}^\top x'| = \varepsilon > 0$; see Lemma 3. Let $z_v^*(\delta) = z^* + \delta v$ where $v \in \mathbb{R}^D$ is a unit norm vector and $\varepsilon = \text{gap}({z^*}^\top X) > 0$ — which is well-defined since $|\text{unique}(X)| > 1$. Then, we have

$$\begin{aligned}
\forall \text{ distinct } x, x' \in \text{unique}(X), \delta < \frac{\varepsilon}{2\text{diam}(X)} : &\|z_v^*(\delta)^\top (x - x')\|_2 = \|(z^* + \delta v)^\top (x - x')\|_2 \\
&\overset{\text{(a)}}{\geq} \|{z^*}^\top (x - x')\|_2 - \delta \|v^\top (x - x')\|_2 \\
&\overset{\text{(b)}}{>} \varepsilon - \frac{\varepsilon}{2\text{diam}(X)} \|x - x'\|_2 \\
&\overset{\text{(c)}}{\geq} \varepsilon - \frac{\varepsilon}{2} = \frac{\varepsilon}{2} > 0,
\end{aligned}$$

where (a) is due to the triangle inequality, (b) is due to $|{z^*}^\top x - {z^*}^\top x'| = \varepsilon$ and $\delta < \frac{\varepsilon}{2\text{diam}(X)}$, and (c) is due to $\|x - x'\|_2 \leq \text{diam}(X)$. Therefore, the vector $z_v^*(\delta)$ separates distinct elements of X in $z_v^*(\delta)^\top X$ — for all unit norm vectors $v \in \mathbb{R}^D$ and $\delta < \frac{\varepsilon}{2\text{diam}(X)}$. On the other hand, if x, x' are two identical elements in X, then we have $z_v^*(\delta)^\top x = z_v^*(\delta)^\top x'$ — that is, the repeated elements in X correspond to the repeated elements in $z_v^*(\delta)^\top X$. Therefore, we have $|\text{unique}(z_v^*(\delta)^\top X)| =$

$|\mathrm{unique}(X)|$, or equivalently $z_v^*(\delta)^\top \in \mathrm{separators} \circ p(t; z, X)$.
Directional derivative of $\mathrm{sort} \circ \mathrm{root} \circ p(t; z, X)$ **at** $z = z^*$**.**
Let $z^* \in \mathrm{separators} \circ p\big(t; z, \Phi(X)\big)$. Then, we have

$$\mathrm{sort} \circ \mathrm{root} \circ p(t; z^*, X) = \big({z^*}^\top x_{\pi_{z^*}(n)}\big)_{n \in [N]} \in \mathbb{R}^N,$$

where $\pi_{z^*} : [N] \to [N]$ is a permutation operator such that ${z^*}^\top x_{\pi_{z^*}(1)} \geq {z^*}^\top x_{\pi_{z^*}(2)} \geq \cdots \geq {z^*}^\top x_{\pi_{z^*}(N)}$. The repeated elements in X do not change the value of the output of the sort function as they correspond to the repeated elements in ${z^*}^\top X$. In other words, π_{z^*} is not necessarily unique; but our results do not depend on the specific choice of the permutation operator. The minimum distance between distinct elements of ${z^*}^\top X$ is $\varepsilon = \mathrm{gap}({z^*}^\top X) > 0$. If $z_v^*(\delta)$ is the perturbed version of z^* in direction of v such that $\|z^* - z_v^*(\delta)\|_2 = \delta < \frac{\varepsilon}{2\mathrm{diam}(X)}$, then $z_v^*(\delta) \in \mathrm{separators} \circ p(t; z, X)$ — see our discussion in the previous paragraph.

Claim 1 *The following equality holds true:*

$$\forall n \in [N] : x_{\pi_{z_v^*(\delta)}(n)} = x_{\pi_{z^*}(n)},$$

for any unit norm vector $v \in \mathbb{R}^D$ *and* $\delta < \frac{\varepsilon}{2\mathrm{diam}(X)}$, *and any permutation operator* $\pi_{z_v^*(\delta)} : [N] \to [N]$ *such that* $z_v^*(\delta)^\top x_{\pi_{z_v^*(\delta)}(1)} \geq z_v^*(\delta)^\top x_{\pi_{z_v^*(\delta)}(2)} \geq \cdots \geq z_v^*(\delta)^\top x_{\pi_{z_v^*(\delta)}(N)}$.

Proof Consider $i, j \in [N]$ where $i > j$. If $x_{\pi_{z^*}(j)} = x_{\pi_{z^*}(i)}$, then we have $z_v^*(\delta)^\top x_{\pi_{z^*}(j)} \geq z_v^*(\delta)^\top x_{\pi_{z^*}(i)}$ — as both terms are equal to each other. On the other hand, if $x_{\pi_{z^*}(j)} \neq x_{\pi_{z^*}(i)}$, then we have

$$\begin{aligned}
z_v^*(\delta)^\top x_{\pi_{z^*}(j)} - z_v^*(\delta)^\top x_{\pi_{z^*}(i)} &= (z^* + \delta v)^\top (x_{\pi_{z^*}(j)} - x_{\pi_{z^*}(i)}) \\
&\overset{\text{(a)}}{\geq} {z^*}^\top (x_{\pi_{z^*}(j)} - x_{\pi_{z^*}(i)}) - \delta \|x_{\pi_{z^*}(j)} - x_{\pi_{z^*}(i)}\|_2 \\
&\overset{\text{(b)}}{\geq} \varepsilon - \delta \mathrm{diam}(X) > \varepsilon - \frac{\varepsilon}{2} = \frac{\varepsilon}{2} > 0,
\end{aligned}$$

where (a) is due to Cauchy–Schwarz inequality, and (b) is due to $\|x_{\pi_{z^*}(j)} - x_{\pi_{z^*}(i)}\|_2 \leq \mathrm{diam}(X)$ and $\delta < \frac{\varepsilon}{2\mathrm{diam}(X)}$. Therefore, the permutation π_{z^*} also sorts the elements of $z_v^*(\delta)^\top X$, that is, $x_{\pi_{z_v^*(\delta)}(n)} = x_{\pi_{z^*}(n)}$, for all $n \in [N]$. ∎

Finally, for all $d \in [D]$, we have

$$\begin{aligned}
\nabla_{e_d} \mathrm{sort} \circ \mathrm{root} \circ p\big(t; z, \Phi(X)|_{z = z^*} &\overset{\text{(a)}}{=} \lim_{\delta \to 0} \frac{1}{\delta} \Big(\mathrm{sort}((z^* + \delta e_d)^\top X) - \mathrm{sort}({z^*}^\top X) \Big) \\
&\overset{\text{(b)}}{=} \lim_{\delta \to 0} \frac{1}{\delta} \Big((z^* + \delta e_d)^\top x_{\pi_{z^* + \delta e_d}(n)} - {z^*}^\top x_{\pi_{z^*}(n)} \Big)_{n \in [N]} \\
&\overset{\text{(c)}}{=} \lim_{\delta \to 0} \frac{1}{\delta} \Big((z^* + \delta e_d)^\top x_{\pi_{z^*}(n)} - {z^*}^\top x_{\pi_{z^*}(n)} \Big)_{n \in [N]} \\
&= \lim_{\delta \to 0} \frac{1}{\delta} (\delta e_d^\top x_{\pi_{z^*}(n)})_{n \in [N]} = (e_d^\top x_{\pi_{z^*}(n)})_{n \in [N]}.
\end{aligned}$$

where (a) is due to the definition of the directional derivation, (b) follows from the definition of permutation operator in sort function, and (c) follows from Claim 1.
Case 2: $|\mathbf{unique}(\mathbf{X})| = 1$.
This directional derivative is well-defined if $|\mathrm{unique}(X)| = 1$, that is,

$$\nabla_v \mathrm{sort}\big(z^\top X\big)|_{z=z^*} = \lim_{\delta\to 0}\frac{1}{\delta}\Big(\mathrm{sort}\big((z^*+\delta v)^\top X\big) - \mathrm{sort}\big({z^*}^\top X\big)\Big)$$
$$= \lim_{\delta\to 0}\frac{1}{\delta}\Big(\big((z^*+\delta v)^\top x_{\pi_1(n)}\big)_{n\in[N]} - \big({z^*}^\top x_{\pi_2(n)}\big)_{n\in[N]}\Big) \qquad = v^\top x 1,$$

where $\pi_1, \pi_2 : [N] \to [N]$ are two permutation operators, and $X = \{\{x_n : n \in [N]\}\}$ and $x_n = x$ for all $n \in [N]$, and $1 \in \mathbb{R}^N$ is the vector of all ones. Therefore, for all $z^* \in$ separators $\circ$ $p\big(t; z, \Phi(X)\big) = \mathbb{R}^D$. And we have

$$\forall d \in [D] : \nabla_{e_d} \mathrm{sort} \circ \mathrm{root} \circ p\big(t; z, \Phi(X)|_{z=z^*} = e_d^\top x 1.$$

This readily proves the proposition's statement.

A.4. Two Illustrative examples

[Repeated Roots] Let $N = D = 2$, and $\Phi(X) = \begin{pmatrix}2 & 0 & 2 & 0 & 0\end{pmatrix}^\top \in \mathbb{R}^{\binom{N+D}{D}-1} = \mathbb{R}^5$ for a multiset X. The goal is to recover X. In the proof of Proposition 5, Lemma 2 relates parameterized moments of the multivariate polynomial $p(t; z, \Phi(X))$ to $\Phi(X)$ using the following functions:

$$\forall z = (z_1, z_2)^\top \in \mathbb{R}^2 :\ \psi(z,1) = \begin{pmatrix}z_1 & z_2 & 0 & 0 & 0\end{pmatrix}^\top,\ \psi(z,2) = \begin{pmatrix}0 & 0 & z_1^2 & 2z_1z_2 & z_2^2\end{pmatrix}^\top.$$

Since $E_n(z, X) = \langle \psi(z, n), \Phi(X)\rangle$, for $n \in [2]$, then we have $E_1(z, X) = 2z_1$ and $E_2(z, X) = 2z_1^2$. We now can use Girard's formula (see equation (7)):

$$a_1(z; X) = E_1(z, X), a_2(z; X) = \frac{1}{2}\det\begin{pmatrix}E_1(z, X) & 1\\ E_2(z, X) & E_1(z, X),\end{pmatrix}.$$

to compute the coefficients of the multivariate polynomial as $a_1(z; X) = 2z_1$ and $a_2(z; X) = z_1^2$. We arrive at the following multivariate polynomial:

$$p\big(t; z, \Phi(X)\big) = t^2 - a_1(z; X)t + a_2(z; X) = t^2 - 2z_1 t + z_1^2 = (t - z_1)^2,$$

and roots $\circ\, p\big(t; z, \Phi(X) = z^\top X = \{\{z_1, z_1\}\}$, for all $z \in \mathbb{R}^2$. Since $\mathrm{unique}(z^\top X) = 1$ — $\forall z \in \mathbb{R}^2$ —- then we have separators $\circ\, p\big(t; z, \Phi(X) = \mathbb{R}^2$. Let $z^* = (1, 1)^\top \in \mathbb{R}^2$ be a separator vector. Therefore, we have $\mathrm{sort}(z^\top X)|_{z=z^*} = (1, 1)^\top$. We also have

$$\mathrm{sort}(z + \delta e_1)^\top X|_{z=z^*} = (1 + \delta, 1 + \delta)^\top,\ \ \mathrm{sort}(z + \delta e_2)^\top X|_{z=z^*} = (1, 1)^\top.$$

for all $\delta > 0$. These quantities let us compute the directional derivatives in Proposition 6 as follows:

$$(e_1^\top x_{\pi_{z^*}}(n))_{n\in[2]} = (1, 1)^\top,\ (e_2^\top x_{\pi_{z^*}}(n))_{n\in[2]} = (0, 0)^\top,$$

see equation (4). Finally, we arrive at $X = \Phi^{-1} \circ \Phi(X) = \{\{(1, 0), (1, 0)\}\}$. [Unique Roots] Let $N = D = 2$, and

$$\Phi(X) = \begin{pmatrix}-2 & 1 & 10 & -7 & 5\end{pmatrix}^\top \in \mathbb{R}^{\binom{N+D}{D}-1} = \mathbb{R}^5$$

for a multiset X. The goal is to recover X. Since $E_n(z,X) = \langle \psi(z,n), \Phi(X)\rangle$, for $n \in [2]$, then we have $E_1(z,X) = -2z_1 + z_2$ and $E_2(z,X) = 10z_1^2 - 14z_1z_2 + 5z_2^2$. We now can use Girard's formula; see Lemma 2 and equation (7)):

$$a_1(z;X) = E_1(z,X), a_2(z;X) = \frac{1}{2}\det\begin{pmatrix} E_1(z,X) & 1 \\ E_2(z,X) & E_1(z,X), \end{pmatrix}.$$

to compute the coefficients of the multivariate polynomial as $a_1(z;X) = -2z_1 + z_2$ and $a_2(z;X) = -3z_1^2 - 2z_2^2 + 5z_1z_2$. We then have the following multivariate polynomial:

$$p\big(t; z, \Phi(X) = t^2 + (2z_1 - z_2)t - 3z_1^2 - 2z_2^2 + 5z_1z_2.$$

To compute the roots of $p\big(t; z, \Phi(X)$, we use the quadratic formula. The discriminant is given as follows:

$$\Delta(z,X) = a_1(z;X)^2 - 4a_2(z;X) = 16z_1^2 + 9z_2^2 - 24z_1z_2 = (4z_1 - 3z_2)^2.$$

The parametric roots are $r_1(z,X) = \frac{1}{2}(-a_1(z;X) + \sqrt{\Delta(z,X)}) = z_1 - z_2$ and $r_1(z,X) = \frac{1}{2}(-a_1(z;X) - \sqrt{\Delta(z,X)}) = -3z_1 + 2z_2$, that is, $\mathrm{roots} \circ p\big(t; z, \Phi(X) = z^\top X = \{\{z_1 - z_2, -3z_1 + 2z_2\}\}$, for all $z \in \mathbb{R}^2$. Since $\mathrm{unique}(z^\top X) = 2$ — $\forall z \in \mathbb{R}^2 \setminus \{z \in \mathbb{R}^2 : z_1 - z_2 = -3z_1 + 2z_2\}$ —- then we have $\mathrm{separators} \circ p\big(t; z, \Phi(X) = \{z \in \mathbb{R}^2 : z_1 \neq \frac{3}{4}z_2\}$. Let $z^* = (1,1)^\top \in \mathbb{R}^2$ be a separator vector. Therefore, we have $\mathrm{sort}(z^\top X)|_{z=z^*} = (0,-1)^\top$.

$$\mathrm{sort}(z + \delta e_1)^\top X|_{z=z^*} = (\delta, -1 - 3\delta)^\top, \;\; \mathrm{sort}(z + \delta e_2)^\top X|_{z=z^*} = (-\delta, -1 + 2\delta)^\top.$$

for all $0 < \delta < \frac{1}{3}$. Now we can compute the directional derivatives in Proposition 6 as follows:

$$(e_1^\top x_{\pi_{z^*}}(n))_{n\in[2]} = (1,-3)^\top, \; (e_2^\top x_{\pi_{z^*}}(n))_{n\in[2]} = (-1,2)^\top,$$

see equation (4). Finally, we arrive at $X = \Phi^{-1} \circ \Phi(X) = \{\{(1,-1), (-3,2)\}\}$.

Appendix B. Proof of Theorem 3

The function $f : \mathbb{X}_{\mathbb{D},N} \to \mathrm{codom}(f)$ is continuous over its domain, that is, $\rho \circ \Phi$ is continuous over $\mathbb{X}_{\mathbb{D},N}$, and we have $\rho = f \circ \Phi^{-1}$; see Theorem 8 and its proof for the definition of Φ and its inverse. Before proceeding with the proof, let us introduce the following set.

Definition 10 *For any mutiset function Φ, we let $\Phi(\mathbb{X}_{\mathbb{D},N}) \stackrel{\text{def}}{=} \{\Phi(X) : X \in \mathbb{X}_{\mathbb{D},N}\}$.*

With the notation in Definition 10, $\rho = f \circ \Phi^{-1}$ is a map from $\Phi(\mathbb{X}_{\mathbb{D},N})$ to $\mathrm{codom}(f)$. Using Lemmas 4 and 5 and Fact 1, we show first that $\Phi(\mathbb{X}_{\mathbb{D},N})$ is a compact set.

Lemma 4 *$\Phi : \mathbb{X}_{\mathbb{D},N} \to \Phi(\mathbb{X}_{\mathbb{D},N})$ is a continuous and injective function.*

Lemma 5 *$\mathbb{X}_{\mathbb{D},N}$ is a compact set.*

Fact 1 *(Pugh and Pugh 2002) The image of a compact set under continuous map is a compact set.*

In Proposition 8, we prove that Φ^{-1} is a continuous function over the compact set $\Phi(\mathbb{X}_{\mathbb{D},N})$.

Proposition 8 *The function Φ^{-1} is continuous on the compact set $\Phi(\mathbb{X}_{\mathbb{D},N})$.*

As a direct result of Proposition 8, $\rho = f \circ \Phi^{-1}$ is a continuous function on the compact subset $\Phi(\mathbb{X}_{\mathbb{D},N}) \subset \mathbb{R}^{\binom{N+D}{D}-1}$.

Fact 2 *Since $\Phi(\mathbb{X}_{\mathbb{D},N})$ is a compact subset of $\mathbb{R}^{\binom{N+D}{D}-1}$, the continuous function $\rho : \Phi(\mathbb{X}_{\mathbb{D},N}) \to \mathrm{codom}(f)$ has a continuous extension to $\mathbb{R}^{\binom{N+D}{D}-1}$, that is, there exists a continuous function $\rho_{\mathrm{e}} : \mathbb{R}^{\binom{N+D}{D}-1} \to \mathrm{codom}(\rho_{\mathrm{e}})$ where*

$$\forall u \in \Phi(\mathbb{X}_{\mathbb{D},N}) : \rho_{\mathrm{e}}(u) = \rho(u),$$

and $\mathrm{codom}(f) \subseteq \mathrm{codom}(\rho_{\mathrm{e}})$. To see the continuous extension theorem, refer to (Deimling, 2010).

From Fact 2, there exists a continuous function $\rho_{\mathrm{e}} : \mathbb{R}^{\binom{N+D}{D}-1} \to \mathrm{codom}(\rho_{\mathrm{e}})$ where $f(X) = \rho_{\mathrm{e}} \circ \Phi(X)$ for all $X \in \mathbb{X}_{\mathbb{D},N}$. Finally, if we rename ρ_{e} to ρ, we arrive at the theorem's statement.

B.1. Proof of Lemma 4

As a direct result of Theorem 8, Φ is an injective function as it is invertible over its domain. The continuity of Φ follows from the continuity of ϕ — see Lemma 6.

Lemma 6 *Let $\phi : \mathbb{D} \to \mathrm{codom}(\phi) \subset \mathbb{R}^K$ be a continuous function on metric space $(\mathbb{D}, \mathrm{d})$ and $\Phi : \mathbb{X}_{\mathbb{D},N} \to \mathrm{codom}(\Phi) \subset \mathbb{R}^K$, $\Phi(X) = \sum_{x\in X} \phi(x)$ for $K, N \in \mathbb{N}$. Then, Φ is a continuous multiset function on $\mathbb{X}_{\mathbb{D},N}$. The same result is also valid on domain $\mathbb{X}_{\mathbb{D},[N]}$.*

Proof We use the following the notion of distance between multisets with elements in $\mathbb{D}$:

$$d_M(X, X') = \begin{cases} \min_{\pi\in\Pi(N_\circ)} \sqrt{\sum_{n\in[N_\circ]} d(x_n, x'_{\pi(n)})^2} & \text{if } |X| = |X'| = N_\circ \\ \infty & \text{otherwise,} \end{cases} \tag{12}$$

where $N_\circ \in [N]$, $|\cdot|$ returns the cardinality of its input multiset, $\Pi(N_\circ)$ is the set of permutation operators on $[N_\circ]$, $X = \{\{x_n : n \in [|X|]\}\}$, and $X' = \{\{x'_n : n \in [|X'|]\}\}$.

Following the definition of continuity, for any $\varepsilon > 0$, we want to find a $\delta(\varepsilon)$ such that if $d_M(X, X') < \delta(\varepsilon)$, then $\|\Phi(X) - \Phi(X')\|_2 < \varepsilon$.

For any $\delta > 0$ and $X \in \mathbb{X}_{\mathbb{D},[N]}$, let $X' \in \mathbb{X}_{\mathbb{D},[N]}$ be such that $d_M(X, X') < \delta$, that is, both multisets have the same size of $|X| = |X'| = N_\circ \in [N]$ and there is a permutation operator $\pi : [N_\circ] \to [N_\circ]$ such that $d_M(X, X') = \sqrt{\sum_{n\in[N_\circ]} d(x_n, x'_{\pi(n)})^2} < \delta$. It suffices to show the following:

$$\|\Phi(X) - \Phi(X')\|_2 = \|\sum_{x\in X} \phi(x) - \sum_{x'\in X'} \phi(x')\|_2 \overset{(a)}{\leq} \sum_{n\in[N_\circ]} \|\phi(x_n) - \phi(x'_{\pi(n)})\|_2$$
$$\overset{(b)}{\leq} \sum_{n\in[N_\circ]} \max_{v\in\mathbb{R}^D:\|v\|_2<\delta} \|\phi(x_n) - \phi(x_n + v)\|_2 < \varepsilon,$$

where (a) is due to the triangle inequality, (b) is due to $\|x_n - x'_{\pi(n)}\|_2 \leq \delta$, for all $n \in [N_\circ]$. It suffices to show that for any $\varepsilon > 0$, there exits a $\delta(\varepsilon)$ such that

$$\forall n \in [N], v \in \mathbb{R}^D, \|v\|_2 < \delta(\varepsilon) : \|\phi(x_n) - \phi(x_n + v)\|_2 < N^{-1}\varepsilon < N_\circ^{-1}\varepsilon.$$

Since ϕ is a continuous function, there exists a $\delta_\phi(x_n, N^{-1}\varepsilon) > 0$ such that

$$\forall v \in \mathbb{R}^D, \|v\|_2 < \delta_\circ(x_n, N^{-1}\varepsilon) : \|\phi(x_n) - \phi(x_n + v)\|_2 < N^{-1}\varepsilon.$$

If we let $\delta(\varepsilon) = \min_{n\in[N_\circ]} \delta_\phi(x_n, N^{-1}\varepsilon) > 0$, then we have $\|\Phi(X) - \Phi(X')\|_2 \leq \varepsilon$. Therefore, Φ is a continuous function. The same result is also valid on domain $\mathbb{X}_{\mathbb{D},N}$. ■

B.2. Proof of Lemma 5

Let $\mathrm{OC}(S)$ be the set of all open covers of a topological space S.

Fact 3 *(Engelking 1989) A topological space S is compact if any open cover of S has a finite subcover.*

Definition 11 *We define the following maps between subsets of $\mathbb{X}_{\mathbb{D},N}$ and $\mathbb{D} \subseteq \mathbb{R}^D$.*

- *Let $\mathbb{U} \subseteq \mathbb{D}^N$ and $T = [x_1, \ldots, x_N] \in \mathbb{U}$. Then, we let $\mathrm{set}(T) \overset{\text{def}}{=} \{\{x_n : n \in [N]\}\} \in \mathbb{X}_{\mathbb{D},N}$ and $\mathrm{set}(\mathbb{U}) \overset{\text{def}}{=} \{\mathrm{set}(T) : T \in \mathbb{U}\} \subseteq \mathbb{X}_{\mathbb{D},N}$*
- *Let $\mathbb{V} \subseteq \mathbb{X}_{\mathbb{D},N}$ and $X = \{\{x_n : n \in [N]\}\} \in \mathbb{V}$. Then, we let*

$$\mathrm{mat}(X) \overset{\text{def}}{=} \{[x_{\pi(1)}, \ldots, x_{\pi(N)}] : \pi \in \Pi(N)\} \subseteq \mathbb{R}^D$$

and $\mathrm{mat}(\mathbb{V}) \overset{\text{def}}{=} \bigcup_{X\in\mathbb{V}} \mathrm{mat}(X) \subseteq \mathbb{R}^D$,

where $\Pi(N)$ is the set of permutation operators $\pi : [N] \to [N]$ for $N \in \mathbb{N}$.

Given a matrix, the function set maps it to a multiset. In contrast, the function mat creates all possible matrices by rearranging elements of its input multiset.

Claim 2 *If* $\{\mathbb{V}_\lambda : \lambda \in \Lambda\} \in \mathrm{OC}(\mathbb{X}_{\mathbb{D},N})$*, then* $\{\mathrm{mat}(\mathbb{V}_\lambda) : \lambda \in \Lambda\} \in \mathrm{OC}(\mathbb{D}^N)$.

Claim 3 *If* $\{\mathbb{U}_\lambda : \lambda \in \Lambda\} \in \mathrm{OC}(\mathbb{D}^N)$*, then* $\{\mathrm{set}(\mathbb{U}_\lambda) : \lambda \in \Lambda\} \in \mathrm{OC}(\mathbb{X}_{\mathbb{D},N})$

Let $\{\mathbb{V}_\lambda : \lambda \in \Lambda\}$ be an open cover for $\mathbb{X}_{\mathbb{D},N}$. From Claim 2, $\{\mathrm{mat}(\mathbb{V}_\lambda) : \lambda \in \Lambda\}$ is an open cover for $\mathbb{D}^N$ — a closed and bounded subset of $\mathbb{R}^D$. Therefore, there is a finite subsequence $\{\mathrm{mat}(\mathbb{V}_{\lambda_k}) : k \in [K]\}$ that forms an open cover for $\mathbb{D}^N$. From Claim 3, $\{\mathrm{set} \circ \mathrm{mat}(\mathbb{V}_{\lambda_k}) : k \in [K]\} = \{\mathbb{V}_{\lambda_k} : k \in [K]\}$, is a finite open cover for $\mathbb{X}_{\mathbb{D},N}$. Therefore, $\mathbb{X}_{\mathbb{D},N}$ is a compact set.

Proof of Claim 2 To prove $\{\mathrm{mat}(\mathbb{V}_\lambda) : \lambda \in \Lambda\}$ is an open cover for $\mathbb{D}^N$, we first show that for all $T \in \mathbb{D}^N \subseteq \mathbb{R}^{N\times D}$, we have $T \in \mathrm{mat}(\mathbb{V}_\lambda)$ for a $\lambda \in \Lambda$.

Let $T = [x_1, \ldots, x_N] \in \mathbb{D}^N$. Then, we have $\mathrm{set}(T) = \{\{x_n : n \in [N]\}\} \in \mathbb{V}_\lambda \subseteq \mathbb{X}_{\mathbb{D},N}$ for a $\lambda \in \Lambda$. Since the following holds true:

$$\forall \pi \in \Pi(N) : [x_{\pi(1)}, \ldots, x_{\pi(N)}] \in \mathrm{mat}(\mathbb{V}_\lambda),$$

then, we have $T \in \mathrm{mat}(\mathbb{V}_\lambda)$. Therefore, $\{\mathrm{mat}(\mathbb{V}_\lambda) : \lambda \in \Lambda\}$ forms a cover for $\mathbb{D}^N$.

Next, we prove that $\mathrm{mat}(\mathbb{V}_\lambda)$ is an open set. Let $T = [x_1, \ldots, x_N] \in \mathrm{mat}(\mathbb{V}_\lambda)$, $\varepsilon > 0$, and $\mathcal{N}(T, \varepsilon) = \{T' \in \mathbb{R}^{N\times D} : \|T - T'\|_F \leq \varepsilon\}$. We want to show that for small enough $\varepsilon > 0$, $\mathcal{N}(T, \varepsilon) \subseteq \mathrm{mat}(\mathbb{V}_\lambda)$.

For all $T' = [x'_1, \ldots, x'_N] \in \mathcal{N}(T, \varepsilon)$, we have

$$d_M(X, X') = \min_{\pi\in\Pi(N)} \sqrt{\sum_{n\in[N]} \|x_n - x'_{\pi(n)}\|_2^2} \leq \|T - T'\|_F \leq \varepsilon, \text{ where } X' = \{\{x'_n : n \in [N]\}\}.$$

Since $\mathbb{V}_\lambda$ is an open set, for any $X \in \mathbb{V}_\lambda$, there exists $\varepsilon > 0$ such that $X' \in V_\lambda$ where $d_M(X, X') < \varepsilon$. Therefore, we have $T' \in \mathrm{mat}(\mathbb{V}_\lambda)$. Since this is the case for all $T' \in \mathcal{N}(T, \varepsilon)$, we have $\mathcal{N}(T, \varepsilon) \subseteq \mathrm{mat}(\mathbb{V}_\lambda)$, that is, $\mathrm{mat}(\mathbb{V}_\lambda)$ is an open set.

Proof of Claim 3 To prove $\{\mathrm{set}(\mathbb{U}_\lambda) : \lambda \in \Lambda\}$ is an open cover for $\mathbb{X}_{\mathbb{D},N}$, we first show that for all $X \in \mathbb{X}_{\mathbb{D},N}$, we have $X \in \mathrm{set}(\mathbb{U}_\lambda)$ for a $\lambda \in \Lambda$.

Let $X = \{\{x_n : n \in [N]\}\} \in \mathbb{X}_{\mathbb{D},N}$. Since $T_\pi = [x_{\pi(1)}, \ldots, x_{\pi(N)}] \in \mathbb{D}^N$ — for all $\pi \in \Pi(N)$ — we have $T_\pi \in \mathbb{U}_\lambda$ for a $\lambda \in \Lambda$. Therefore, we have $\mathrm{set}(T_\pi) = \{\{x_{\pi(n)} : n \in [N]\}\} = X \in \mathrm{set}(\mathbb{U}_\lambda)$. This proves that $\{\mathrm{set}(\mathbb{U}_\lambda) : \lambda \in \Lambda\}$ is a cover for $\mathbb{X}_{\mathbb{D},N}$.

We now prove that $\mathrm{set}(\mathbb{U}_\lambda)$ is an open set. Let $X = \{\{x_n : n \in [N]\}\} \in \mathrm{set}(\mathbb{U}_\lambda)$, $\varepsilon > 0$, $\mathcal{N}(X, \varepsilon) = \{X' \in \mathbb{X}_{\mathbb{R}^D,N} : d_M(X, X') \leq \varepsilon\}$, and $T = [x_1, \ldots, x_N]$. We want to show that for small enough $\varepsilon > 0$, $\mathcal{N}(X, \varepsilon) \subseteq \mathrm{set}(\mathbb{U}_\lambda)$.

For all $X' = \{\{x'_n : n \in [N]\}\} \in \mathcal{N}(X, \varepsilon)$, we have

$$\|T - T'_\pi\|_F = d_M(X, X') \leq \varepsilon \text{ where } T'_\pi = [x'_{\pi(1)}, \ldots, x'_{\pi(N)}],$$

for a permutation operator $\pi : [N] \to [N]$ that best match elements of X and X'. Since $\mathbb{U}_\lambda$ is an open subset, there exists $\varepsilon > 0$ such that $T_\pi \in \mathbb{U}_\lambda$. Therefore, we have $X' = \mathrm{set}(T'_\pi) \in \mathrm{set}(\mathbb{U}_\lambda)$. Since this is the case for all $X' \in \mathcal{N}(X, \varepsilon)$, we have $\mathcal{N}(X, \varepsilon) \subseteq \mathrm{set}(\mathbb{U}_\lambda)$, that is, $\mathrm{set}(\mathbb{U}_\lambda)$ is an open set.

B.3. Proof of Proposition 8

By definition of continuity, we want to show that, for any $\varepsilon > 0$ and $X \in \mathbb{X}_{\mathbb{D},N}$, there exists $\delta_f(\varepsilon) > 0$ such that

$$\begin{aligned}\forall X' \in \mathbb{X}_{\mathbb{D},N}, \|\Phi(X) - \Phi(X')\|_2 < \delta_f(\varepsilon) &: d_M(\Phi^{-1} \circ \Phi(X) - \Phi^{-1} \circ \Phi(X')) < \varepsilon \\ &: d_M(X, X') < \varepsilon,\end{aligned}$$

where d_M is given in equation (12). We use the result in Lemma 7 to establish the continuity of Φ^{-1} over $\Phi(\mathbb{X}_{\mathbb{D},N})$.

Lemma 7 *Let $X \in \mathbb{X}_{\mathbb{D},N}$. The parameterized multiset that consists of the root of the polynomial $p\big(t; z, \Phi(X)\big)$ in equation (3) (that is, $z^\top X$) varies continuously with $\Phi(X)$. More precisely, for all $\varepsilon > 0$, there exists $\delta(\varepsilon) > 0$ such that*

$$\forall X' \in \mathbb{X}_{\mathbb{D},N}, \|\Phi(X) - \Phi(X')\|_2 < \delta(\varepsilon) : \max_{z \in \mathbb{R}^D : \|z\|_2 = 1} d_M(z^\top X, z^\top X') < \varepsilon.$$

Let $\varepsilon > 0$, $X = \{\{x_n : n \in [N]\}\}$ and $X' = \{\{x_n^{]\prime} : n \in [N]\}\} \in \mathbb{X}_{\mathbb{D},N}$. From Lemma 7, if $\|\Phi(X) - \Phi(X')\|_2 < \delta(\varepsilon)$, then we have

$$\forall z \in \mathbb{R}^D, \|z\|_2 = 1 : d_M(z^\top X, z^\top X') = \sqrt{\sum_{n \in [N]} |z^\top x_n - z^\top x'_{\pi^*(n)}|^2} < \varepsilon$$

for a permutation operator $\pi^* : [N] \to [N]$. Then, we have

$$\forall z \in \mathbb{R}^D, \|z\|_2 = 1, n \in [N] : |z^\top x_n - z^\top x'_{\pi^*(n)}|^2 < \varepsilon.$$

If $x_n - x'_{\pi^*(n)} \neq 0$ and $z = \|x_n - x'_{\pi^*(n)}\|_2^{-1}(x_n - x'_{\pi^*(n)})$, then we have arrive at $\|x_n - x'_{\pi^*(n)}\|_2 < \varepsilon$, where $n \in [N]$. If $x_n - x'_{\pi^*(n)} = 0$, then $\|x_n - x'_{\pi^*(n)}\|_2 < \varepsilon$ is trivially the case. Therefore, we have

$$d_M(X, X') = \min_{\pi \in \Pi(N)} \sqrt{\sum_{n \in [N]} \|x_n - x'_{\pi(n)}\|_2^2} \leq \sqrt{N}\varepsilon,$$

where $\Pi(N)$ is the set of permutation operators on $[N]$. Finally, we establish the continuity of Φ^{-1} on $\Phi(\mathbb{X}_{\mathbb{D},N})$ by letting $\delta_f(\varepsilon) = \delta(\frac{\varepsilon}{\sqrt{N}})$, that is,

$$\begin{aligned}\forall X' \in \mathbb{X}_{\mathbb{D},N}, \|\Phi(X) - \Phi(X')\|_2 < \delta_f(\varepsilon) &: \max_{z \in \mathbb{R}^D : \|z\|_2 = 1} d_M(z^\top X, z^\top X') < \frac{\varepsilon}{\sqrt{N}} \\ &: d_M(X, X') < \varepsilon.\end{aligned}$$

Proof of Lemma 7. We construct the the polynomial $p\big(t; z, \Phi(X)\big)$ in equation (3), that is,

$$\forall t \in \mathbb{R}, z \in \mathbb{R}^D : p\big(t; z, \Phi(X)\big) = t^N + \sum_{n \in [N]} (-1)^n a_n(z; X) t^{N-n}$$

by first computing the following parameterized moments:

$$\forall n \in [N], z \in \mathbb{R}^D : E_n(z, X) = \langle \psi(z, n), \Phi(X) \rangle.$$

Fact 4 *For a fixed $z \in \mathbb{R}^D$ and $n \in [N]$, $E_n(z, X)$ is a linear function of $\Phi(X)$. Furthermore, $E_n(z, X)$ is a continuous functions of $(z, \Phi(X))$.*

The coefficients of $p\big(t; z, \Phi(X)\big)$ are polynomial functions of the moments $\big(E_n(z, X)\big)_{n\in[N]}$; see the Newton-Girard equation (7).

Fact 5 *The coefficients of the polynomial $p\big(t; z, \Phi(X)\big)$ in equation (3) vary continuously with the moments $\big(E_n(z, X)\big)_{n\in[N]}$.*

Therefore, the coefficients of the polynomial $p\big(t; z, \Phi(X)\big)$ in equation (3) vary continuously with $(z, \Phi(X))$; see Facts 4 and 5.

Theorem 9 *(Ćurgus and Mascioni, 2006) The function $f : \mathbb{C}^N \to \mathbb{C}^N$, which associates every $a = (a_n)_{n\in[N]} \in \mathbb{C}^N$ to the multiset of roots, $f(a) \in \mathbb{C}^N$, of the monic polynomial formed using a as the coefficient, i.e., $t^N + a_1 t^{N-1} + \cdots + (-1)^{N-1} a_{N-1} x + (-1)^N a_N$, is a homeomorphism.*

From Theorem 9, Facts 4, and 5, the parameterized root multiset of $p\big(t; z, \Phi(X)\big)$ (that is, $z^\top X$) vary continuously with $(z, \Phi(X))$. Therefore, for all $X \in \mathbb{X}_{\mathbb{D},N}$, $z \in \mathbb{R}^D$ and $\varepsilon > 0$, there exists $\delta(\varepsilon, z) > 0$ such

$$\forall X' \in \mathbb{X}_{\mathbb{D},N}, \|\Phi(X) - \Phi(X')\|_2 < \delta(\varepsilon, z) : d_M(z^\top X, z^\top X') < \varepsilon.$$

We may fix the norm of the vector z to one, since by definition of d_M in equation (12), we have

$$\forall \alpha \in \mathbb{R} : d_M((\alpha z)^\top X, (\alpha z)^\top X') = |\alpha| d_M(z^\top X, z^\top X'.$$

After this normalization, for all $\varepsilon > 0$, we have

$$\forall X' \in \mathbb{X}_{\mathbb{D},N}, z \in \mathbb{R}^D, \|z\|_2 = 1, \|\Phi(X) - \Phi(X')\|_2 < \delta(\varepsilon, z) : d_M(z^\top X, z^\top X') < \varepsilon.$$

Let $z^* \in \operatorname{argmax}_{z\in\mathbb{R}^D:\|z\|_2=1} d_M(z^\top X, z^\top X')$. Then, we have

$$\forall X' \in \mathbb{X}_{\mathbb{D},N}, \|\Phi(X) - \Phi(X')\|_2 < \delta(\varepsilon, z^*) : \max_{z\in\mathbb{R}^D:\|z\|_2=1} d_M(z^\top X, z^\top X') < \varepsilon,$$

which proves the statement if z^* exists. Therefore, we need to prove the existence of z^*.
The set $\{z \in \mathbb{R}^D : \|z\|_2 = 1\}$ is compact. If we prove that $d_M(z^\top X, z^\top X')$ is a continuous function of z, then by the extreme-value theorem (Stein and Shakarchi, 2010), z^* does exist. To this end, we show that $d_M^2(z^\top X, z^\top X')$ (and hence $d_M(z^\top X, z^\top X')$) is continuous. We use the following first-order perturbation analysis:

$$d_M^2((z + \mathrm{dz})^\top X, (z + \mathrm{dz})^\top X') = \sum_{n\in[N]} |(z + \mathrm{dz})^\top x_n - (z + \mathrm{dz})^\top x'_{\pi_{z+\mathrm{dz}}(n)}|^2$$

where $\pi_{z+\mathrm{dz}} : [N] \to [N]$ is a permutation operator that best matches elements of perturbed multisets $(z+\mathrm{dz})^\top X$ and $(z+\mathrm{dz})^\top X'$. Let $X'' = \{\{x_n - x'_{\pi_z(n)} : n \in [N]\}\}$. As we discussed in the proof of Proposition 6, if $\|\mathrm{dz}\|_2 < \frac{\mathrm{gap}(z^\top X'')}{\mathrm{diam}(\mathbb{D})}$ — $\mathrm{gap}(z^\top X'') \neq 0$ since $X \neq X'$ — then $x'_{\pi_z(n)} = x'_{\pi_{z+\mathrm{dz}}(n)}$ for all $n \in [N]$. Therefore, we have

$$d_M^2((z + \mathrm{dz})^\top X, (z + \mathrm{dz})^\top X') = d_M^2(z^\top X, z^\top X') + O(\|\mathrm{dz}\|_2^2),$$

that is, $d_M(z^\top X, z^\top X')$ is a continuous function of z. This concludes the proof.

Appendix C. Proof of Theorem 4

C.1. Extension of Theorem 8

Let $\mathbb{D}$ be a compact subset of $\mathbb{R}^D$, that is, compact $\mathbb{D} \neq \mathbb{R}^D$. The encoding function $\Phi(X) = \sum_{x \in X} \phi(x)$ — where $\phi : \mathbb{D} \to \mathrm{codom}(\phi)$ — is an injective map over multisets with exactly N elements, that is, $\Phi^{-1} \circ \Phi(X) = X$ where $X \in \mathbb{X}_{\mathbb{D},N}$. To extend the result to multisets of variable sizes, we follow the proof sketch for the one-dimensional case (Wagstaff et al., 2019). Let $x_\circ \in \mathbb{R}^D \setminus \mathbb{D}$. Then, we define $\phi'(x) = \phi(x) - \phi(x_\circ)$. For a multiset $X \in \mathbb{X}_{\mathbb{D},[N]}$ with $|X| \leq N$ elements, we have

$$
\begin{aligned}
\forall X \in \mathbb{X}_{\mathbb{D},[N]} : \Phi'(X) &= \sum_{x \in X} \phi'(x) = \sum_{x \in X} \phi(x) - |X|\phi(x_\circ) \\
&= \Phi(X \cup \{\{\underbrace{x_\circ, \ldots, x_\circ}_{N-|X|}\}\}) - N\phi(x_\circ) \\
&= \Phi(X \cup \{\{\underbrace{x_\circ, \ldots, x_\circ}_{N-|X|}\}\}) + \mathrm{const}
\end{aligned}
$$

where $\mathrm{const} = -N\phi(x_\circ)$. Since Φ is injective over $\mathbb{X}_{\mathbb{D},N}$, Φ' is an injective map. That is to say,

$$
\begin{aligned}
\forall X \in \mathbb{X}_{\mathbb{D},[N]} : \Big(\Phi^{-1} \circ (\Phi'(X) - \mathrm{const})\Big) \cap \mathbb{D} &= \Big(\Phi^{-1} \circ \Phi(X \cup \{\{\underbrace{x_\circ, \ldots, x_\circ}_{N-|X|}\}\})\Big) \cap \mathbb{D} \\
&= (X \cup \{\{\underbrace{x_\circ, \ldots, x_\circ}_{N-|X|}\}\}) \cap \mathbb{D} \\
&= X.
\end{aligned}
$$

Therefore, we have $\Phi'^{-1}(U) = \Phi^{-1}\big(U - \mathrm{const}\big) \cap \mathbb{D}$ for all $U \in \Phi'(\mathbb{X}_{\mathbb{D},[N]}) = \{\Phi'(X) : X \in \mathbb{X}_{\mathbb{D},[N]}\}$. If we define $\rho = f \circ (\Phi')^{-1}$ where $\mathrm{dom}(\rho) = \Phi'(\mathbb{X}_{\mathbb{D},[N]})$, then we have $f(X) = \rho \circ \Phi'(X)$ for all $X \in \mathbb{X}_{\mathbb{D},[N]}$. We arrive at the theorem's exact statement by renaming Φ' to Φ.

C.2. Extension of Theorem 3

Let $\mathbb{D}$ be a compact subset of $\mathbb{R}^D$, that is, compact $\mathbb{D} \neq \mathbb{R}^D$. In Lemma 5, we prove that $\mathbb{X}_{\mathbb{D},n}$ is a compact set, for all $n \in \mathbb{N}$. Since $\mathbb{X}_{\mathbb{D},[N]}$ is a finite union of compact sets, that is,

$$
\mathbb{X}_{\mathbb{D},[N]} = \bigcup_{n=1}^{N} \mathbb{X}_{\mathbb{D},n},
$$

itself is a compact set (Sutherland, 2009). Since Φ' is a continuous map (see Lemma 6), $\Phi'(\mathbb{X}_{\mathbb{D},[N]})$ is also a compact set (Pugh and Pugh, 2002).

Now let us show that Φ'^{-1} is a continuous map over compact set $\mathrm{codom}(\Phi') = \Phi'(\mathbb{X}_{\mathbb{D},[N]})$. We have to show that for all $\varepsilon > 0$ and all $X, X' \in \mathbb{X}_{\mathbb{D},[N]}$ such that $\|\Phi'(X) - \Phi'(X')\|_2 < \delta(\varepsilon)$ we have $d_M(\Phi'^{-1} \circ \Phi'(X), \Phi'^{-1} \circ \Phi'(X')) < \varepsilon$ where $\delta(\varepsilon) > 0$ and d_M is the matching distance between multisets, that is,

$$
d_M(X, X') = \begin{cases} \min_{\text{bijection } \pi:[N_\circ]\to[N_\circ]} \sqrt{\sum_{n \in [N_\circ]} \|x_n - x'_{\pi(n)}\|_2^2} & \text{if } |X| = |X'| = N_\circ \\ \infty & \text{if } |X| \neq |X'|, \end{cases}
$$

where $X = \{\{x_n : n \in [N_\circ]\}\}$, $X' = \{\{x'_n : n \in [N_\circ]\}\}$, $N_\circ \in [N]$. On the other hand, we have $\Phi'^{-1}(U) = \Phi^{-1}\big(U - \text{const}\big) \cap \mathbb{D}$ for all $U \in \Phi'(\mathbb{X}_{\mathbb{D},[N]})$ where Φ^{-1} is a continuous function; see Proposition 8.

Consider the continuous function $\Psi(U) = \Phi^{-1}\big(U - \text{const}\big)$ where $U \in \Phi'(\mathbb{X}_{\mathbb{D},[N]})$. By definition of continuity, for all $\varepsilon > 0$ and all $X, X' \in \mathbb{X}_{\mathbb{D},[N]}$ such that $\|\Phi'(X) - \Phi'(X')\|_2 < \delta(\varepsilon)$ we have $d_M(\Psi \circ \Phi'(X), \Psi \circ \Phi'(X')) < \varepsilon$ where $\delta(\varepsilon) > 0$. Since we have,

$$\Psi \circ \Phi'(X) = X \cup \{\{\underbrace{x_\circ, \ldots, x_\circ}_{N-|X|}\}\}$$
$$\Psi \circ \Phi'(X') = X' \cup \{\{\underbrace{x_\circ, \ldots, x_\circ}_{N-|X'|}\}\},$$

we can simplify $d_M(\Psi \circ \Phi'(X), \Psi \circ \Phi'(X')) < \varepsilon$ as follows:

$$d_M(X \cup \{\{\underbrace{x_\circ, \ldots, x_\circ}_{N-|X|}\}\}, X' \cup \{\{\underbrace{x_\circ, \ldots, x_\circ}_{N-|X'|}\}\}) < \varepsilon.$$

If X and X' have different number of elements in $\mathbb{D}$, then we have $\varepsilon > \inf_{x\in\mathbb{D}} \|x - x_\circ\|_2$. Let $\varepsilon_\circ > 0$ be such that $\varepsilon_\circ < \inf_{x\in\mathbb{D}} \|x - x_\circ\|_2$. If we pick $0 < \varepsilon < \varepsilon_\circ$, then X and X' have the same number of elements in $\mathbb{D}$ and

$$\begin{aligned} d_M((\Phi')^{-1} \circ \Phi'(X), (\Phi')^{-1} \circ \Phi'(X')) &= d_M(\Psi \circ \Phi'(X) \cap \mathbb{D}, \Psi \circ \Phi'(X') \cap \mathbb{D}) \\ &= d_M(\Psi \circ \Phi'(X), \Psi \circ \Phi'(X')) < \varepsilon \end{aligned}$$

That is, $(\Phi')^{-1}$ is a continuous function over $\Phi'(\mathbb{X}_{\mathbb{D},[N]})$. Therefore, $\rho = f \circ (\Phi')^{-1}$ is a continuous function on compact set $\Phi'(\mathbb{X}_{\mathbb{D},[N]}) \subset \mathbb{R}^{\binom{N+D}{D}-1}$, and it has a continuous extension to $\mathbb{R}^{\binom{N+D}{D}-1}$; refer to Fact 2. We arrive at the theorem's statement by renaming Φ' to Φ.

Appendix D. Proof of Proposition 1

Let $\Phi' : \mathbb{X}_{\mathbb{D},[N]} \to \mathrm{codom}(\Phi')$ where $N = \max\{N_1, N_2\}$, $\Phi'(X) = \sum_{x \in X} \phi'(x)$, and ϕ' is given in the proof of Theorem 4. The function Φ' is injective on $\mathbb{X}_{\mathbb{D},[N]}$ and $(\Phi')^{-1}$ is continuous on compact set $\Phi'(\mathbb{X}_{\mathbb{D},[N]})$. Since $\mathbb{X}_{\mathbb{D},[N_1]}$ and $\mathbb{X}_{\mathbb{D},[N_2]}$ are compact subsets of $\mathbb{X}_{\mathbb{D},[N]} \subseteq \mathbb{R}^{\binom{N+D}{D}-1}$, the function Φ' is injective on $\Phi'(\mathbb{X}_{\mathbb{D},[N_1]})$ and $\Phi'(\mathbb{X}_{\mathbb{D},[N_2]})$ and $(\Phi')^{-1}$ is continuous on both domains. We define the following function:

$$\forall U_1 \in \Phi'(\mathbb{X}_{\mathbb{D},[N_1]}), U_2 \in \Phi'(\mathbb{X}_{\mathbb{D},[N_2]}) : \rho(U_1, U_2) = f\big((\Phi')^{-1}(U_1), (\Phi')^{-1}(U_2)\big).$$

If f is a continuous multiset function, ρ (defined above) is a continuous function on its compact domain $\Phi'(\mathbb{X}_{\mathbb{D},[N_1]}) \times \Phi'(\mathbb{X}_{\mathbb{D},[N_2]})$ as it is the composition of continuous functions. Therefore, it has a continuous extension to $\mathbb{R}^{\binom{N+D}{D}-1} \times \mathbb{R}^{\binom{N+D}{D}-1}$; refer to Fact 2.

Appendix E. Proof of Theorem 5

We define $\phi : \mathbb{R}^D \to \mathrm{codom}(\phi) \subset \mathbb{C}^{D\times N}$ as follows:

$$\forall x \in \mathbb{R}^D : \phi(x) = \begin{pmatrix} r(x) & r(x)^{\odot 2} & \cdots & r(x)^{\odot N} \end{pmatrix} \in \mathbb{C}^{D\times N}, \tag{13}$$

where $r(x) = x + 1l(x)j$, $1 \in \mathbb{R}^D$ is a vector of all ones, $l : \mathbb{R}^D \to \mathbb{R}$ is a continuous function, $j = \sqrt{-1}$, and $\odot$ computes elementwise exponents.

Fact 6 *The function ϕ is continuous.*

Lemma 8 *Let ϕ be the function defined in equation (13). Then, the function $\Phi(X) = \sum_{x\in X} \phi(x)$ is injective on $\mathbb{X}^l_{\mathbb{R}^D,N}$.*

Let $\Phi(\mathbb{X}^l_{\mathbb{R}^D,N}) \overset{\text{def}}{=} \{\Phi(X) : X \in \mathbb{X}^l_{\mathbb{R}^D,N}\}$. From Lemma 8, there exists an inverse function $\Phi^{-1} : \Phi(\mathbb{X}^l_{\mathbb{R}^D,N}) \to \mathbb{X}^l_{\mathbb{R}^D,N}$, that is, $\Phi^{-1} \circ \Phi(X) = X$ for all $X \in \mathbb{X}^l_{\mathbb{R}^D,N}$. We construct $\rho : \Phi(\mathbb{X}^l_{\mathbb{R}^D,N}) \to \mathrm{codom}(f)$ as $\rho = f \circ \Phi^{-1}$. This completes the proof as follows:

$$\forall X \in \mathbb{X}^l_{\mathbb{R}^D,N} : \rho \circ \Phi(X) = f \circ \Phi^{-1} \circ \Phi(X) = f(X).$$

E.1. Proof of Lemma 8

From equation (13), we have

$$\forall X \in \mathbb{X}^l_{\mathbb{R}^D,N} :\ \Phi(X) = \sum_{x\in X} \begin{pmatrix} r(x) & r(x)^{\odot 2} & \cdots & r(x)^{\odot N} \end{pmatrix} \in \mathbb{C}^{D\times N}. \tag{14}$$

Definition 12 *Let Φ^{-1}_{deep} be the continuous function introduced in Deep Sets paper (Zaheer et al., 2017), viz., $\Phi^{-1}_{\text{deep}} \circ \Phi_{\text{deep}} = X$ where $\Phi_{\text{deep}}(X) = (\sum_{x\in X} x, \ldots, \sum_{x\in X} x^N)$, where $X \in \mathbb{X}_{\mathbb{C},N}$ is a multiset of N scalars in $\mathbb{C}$. With slight abuse of notation, we generalize this definition to the following row-wise function:*

$$\forall X_1, \ldots, X_D \in \mathbb{X}_{\mathbb{C},N} :\ \Phi^{-1}_{\text{deep}}(\begin{pmatrix} \Phi_{\text{deep}}(X_1) \\ \vdots \\ \Phi_{\text{deep}}(X_D) \end{pmatrix}) = \begin{pmatrix} \Phi^{-1}_{\text{deep}} \circ \Phi_{\text{deep}}(X_1) \\ \vdots \\ \Phi^{-1}_{\text{deep}} \circ \Phi_{\text{deep}}(X_D) \end{pmatrix} = \begin{pmatrix} X_1 \\ \vdots \\ X_D \end{pmatrix}$$

Definition 13 *Let $X = \{\{x_n \in \mathbb{C} : n \in [N]\}\} \in \mathbb{X}_{\mathbb{C},N}$ be a multiset of N complex-valued elements. We then define the function* sort *as follows:*

$$\mathrm{sort}(X) = \big(\mathrm{Re}(x_{\pi(n)})\big)_{n\in[N]} \in \mathbb{R}^N,$$

where $\pi : [N] \to [N]$ is any permutation operator such that $\mathrm{Im}(x_{\pi(1)}) \le \cdots \le \mathrm{Im}(x_{\pi(N)})$.

Definition 14 *Let $X_1, \ldots, X_D \in \mathbb{X}_{\mathbb{C},N}$ be multisets of N complex-valued elements. We then define the function* sortvec *as follows:*

$$\mathrm{sortvec}(\begin{pmatrix} X_1 \\ \vdots \\ X_D \end{pmatrix}) = \{\{\begin{pmatrix} e_n^\top \mathrm{sort}(X_1) \\ \vdots \\ e_n^\top \mathrm{sort}(X_D) \end{pmatrix} \in \mathbb{R}^D : n \in [N]\}\} \in \mathbb{X}_{\mathbb{R}^D,N},$$

where $e_n \in \mathbb{R}^N$ is the n-th standard basis vector for $\mathbb{R}^N$.

Remark 10 *Permutation operators in Definitions 13 and 14 may not be unique. This happens if the input multiset has at least two elements with nonunique imaginary parts. If this is the case, functions* sort *and* sortvec *are both ill-defined. In what follows, we show that for certain inputs of interest both functions are indeed well-defined.*

Let $\Psi : \Phi(\mathbb{X}^l_{\mathbb{R}^D,N}) \to \mathbb{X}^l_{\mathbb{R}^D,N}$ where $\Psi = \mathrm{sortvec} \circ \Phi^{-1}_{\text{deep}}$. Then, for all $X \in \mathbb{X}^l_{\mathbb{R}^D,N}$, we have

$$\begin{aligned}
\Psi \circ \Phi(X) &= \mathrm{sortvec} \circ \Phi^{-1}_{\text{deep}} \circ \Phi(X) \\
&\overset{\text{(a)}}{=} \mathrm{sortvec} \circ \Phi^{-1}_{\text{deep}}\Big(\sum_{x \in X} \big((x + 1l(x)j) \quad (x + 1l(x)j)^{\odot 2} \quad \cdots \quad (x + 1l(x)j)^{\odot N} \big) \Big) \\
&\overset{\text{(b)}}{=} \mathrm{sortvec} \circ \Phi^{-1}_{\text{deep}}\Big(\begin{pmatrix} \sum_{x\in X} e_1^\top x + l(x)j & \cdots & \sum_{x \in X} (e_1^\top x + l(x)j)^N \\ & \vdots & \\ \sum_{x\in X} e_D^\top x + l(x)j & \cdots & \sum_{x \in X} (e_D^\top x + l(x)j)^N \end{pmatrix} \Big) \\
&\overset{\text{(c)}}{=} \mathrm{sortvec} \circ \Phi^{-1}_{\text{deep}}\Big(\begin{pmatrix} \Phi_{\text{deep}}(\{\{e_1^\top x + l(x)j : x \in X\}\}) \\ \vdots \\ \Phi_{\text{deep}}(\{\{e_D^\top x + l(x)j : x \in X\}\}) \end{pmatrix} \Big) \\
&\overset{\text{(d)}}{=} \mathrm{sortvec}\Big(\begin{pmatrix} \{\{e_1^\top x + l(x)j : x \in X\}\} \\ \vdots \\ \{\{e_D^\top x + l(x)j : x \in X\}\} \end{pmatrix} \Big).
\end{aligned}$$

where (a) is due to equations (14) and (13), (b) follows from explicitly writing the elements of $\Phi(X)$, (c) follows from the definition of Φ_{deep} (see Definition 12), and finally (d) is due to the fact that we allow Φ^{-1}_{deep} to operate elementwise.

Case 1 (Distinct Identifiers). Let $X = \{\{x_n : n \in [N]\}\}$. If all elements of $l(X) = \{\{l(x) : x \in X\}\}$ are unique, then we have

$$\forall d \in [D] : \mathrm{sort}(\{\{e_d^\top x + l(x)j : x \in X\}\}) = \big(e_d^\top x_{\pi(n)}\big)_{n \in [N]} \in \mathbb{R}^N$$

where $\pi : [N] \to [N]$ is the permutation operator such that $l(x_{\pi(1)}) < \cdots < l(x_{\pi(N)})$. Then, we have

$$\Psi \circ \Phi(X) = \{\{ \begin{pmatrix} e_1^\top x_{\pi(n)} \\ \vdots \\ e_D^\top x_{\pi(n)} \end{pmatrix} : n \in [N]\}\} = \{\{x_{\pi(n)} : n \in [N]\}\} = X.$$

Case 2 (Repeated Identifiers). If $l(X)$ has repeated elements, then there exists at least two distinct permutation operators π and π' ($\pi \neq \pi'$) that sort the elements of $l(X)$, that is,

$$\begin{aligned}
l(x_{\pi(1)}) \leq l(x_{\pi(2)}) \leq \cdots \leq l(x_{\pi(N)}) \\
l(x_{\pi'(1)}) \leq l(x_{\pi'(2)}) \leq \cdots \leq l(x_{\pi'(N)}).
\end{aligned}$$

In this case, we have $l(x_{\pi(n)}) = l(x_{\pi'(n)})$ for all $n \in [N]$ — even though $\pi(n) \neq \pi'(n)$ for some $n \in [N]$. From Definition 2, since $l(x_{\pi(n)}) = l(x_{\pi'(n)})$, we have $x_{\pi'(n)} = x_{\pi(n)}$ where $n \in [N]$. Consequently, we have

$$\begin{aligned}
\forall d \in [D] : \mathrm{sort}(\{\{e_d^\top x + l(x)j : x \in X\}\}) &= \big(e_d^\top x_{\pi(n)}\big)_{n \in [N]} \\
&= \big(e_d^\top x_{\pi'(n)}\big)_{n \in [N]} \in \mathbb{R}^N.
\end{aligned}$$

Therefore, even though there are multiple permutation operators that sorts the elements of $\{\{e_d^\top x + l(x)j : x \in X\}\}$, the output of the sort function remains unchanged, that is, sort is a well-defined function for any element of $X \in \mathbb{X}^l_{\mathbb{R}^D,N}$. Consequently, $\mathrm{sortvec}$ is well-defined on $\mathbb{X}^l_{\mathbb{R}^D,N}$ and we have

$$\begin{aligned}\forall X \in \mathbb{X}^l_{\mathbb{R}^D,N} : \Psi \circ \Phi(X) &= \{\{\begin{pmatrix} e_1^\top x_{\pi_1(n)} \\ \vdots \\ e_D^\top x_{\pi_D(n)} \end{pmatrix} : n \in [N]\}\} \\ &\overset{\text{(a)}}{=} \{\{\begin{pmatrix} e_1^\top x_{\pi_1(n)} \\ \vdots \\ e_D^\top x_{\pi_1(n)} \end{pmatrix} : n \in [N]\}\} = X,\end{aligned}$$

where π_d is a permutation operator that sorts the elements of $\{\{e_d^\top x + l(x)j : x \in X\}\}$ — for all $d \in [D]$ — and (a) is due to $x_{\pi_i(n)} = x_{\pi_j(n)}$ for all $i, j \in [D]$ and $n \in [N]$. Therefore, we have

$$\forall X \in \mathbb{X}^l_{\mathbb{R}^D,N} = \Psi \circ \Phi(X) = \mathrm{sortvec} \circ \Phi_{\text{deep}}^{-1} \circ \Phi(X) = X,$$

that is, $\Psi = \mathrm{sortvec} \circ \Phi_{\text{deep}}^{-1}$ is well-defined on $\Phi(\mathbb{X}^l_{\mathbb{R}^D,N})$ and $\Psi = \Phi^{-1} : \Phi(\mathbb{X}^l_{\mathbb{R}^D,N}) \to \mathbb{X}^l_{\mathbb{R}^D,N}$. This proves that Φ is an injective function on $\mathbb{X}^l_{\mathbb{R}^D,N}$.

Appendix F. Proof of Proposition 2

The proof is similar to that of Theorem 4. Let $\mathbb{D}$ be a compact subset of $\mathbb{R}^D$, that is, $\mathbb{D} \neq \mathbb{R}^D$.

The encoding function $\phi : \mathbb{D} \to \mathrm{codom}(\phi)$ (defined in the proof Theorem 5) such that $\Phi(X) = \sum_{x \in X} \phi(x)$ is an injective map over multisets with exactly N elements, that is, $\Phi^{-1} \circ \Phi(X) = X$ where $X \in \mathbb{X}^l_{\mathbb{D},N}$ and $l : \mathbb{D} \to \mathbb{R}$ is the continuous identifier function. Let $x_\circ \in \mathbb{R}^D \setminus \mathbb{D}$. Then, we define $\phi'(x) = \phi(x) - \phi(x_\circ)$. For a multiset $X \in \mathbb{X}^l_{\mathbb{D},[N]}$ with $|X| \leq N$ elements, we have

$$\begin{aligned}
\forall X \in \mathbb{X}^l_{\mathbb{D},[N]} : \Phi'(X) &= \sum_{x \in X} \phi'(x) = \sum_{x \in X} \phi(x) - |X|\phi(x_\circ) \\
&= \Phi(X \cup \{\{\underbrace{x_\circ, \ldots, x_\circ}_{N-|X|}\}\}) - N\phi(x_\circ) \\
&= \Phi(X \cup \{\{\underbrace{x_\circ, \ldots, x_\circ}_{N-|X|}\}\}) + \mathrm{const}
\end{aligned}$$

where $\mathrm{const} = -N\phi(x_\circ)$. Since Φ is injective over $\mathbb{X}_{\mathbb{D},N}$, Φ' is an injective map. That is to say,

$$\begin{aligned}
\forall X \in \mathbb{X}^l_{\mathbb{D},[N]} : \Big(\Phi^{-1} \circ (\Phi'(X) - \mathrm{const})\Big) \cap \mathbb{D} &= \Big(\Phi^{-1} \circ \Phi(X \cup \{\{\underbrace{x_\circ, \ldots, x_\circ}_{N-|X|}\}\})\Big) \cap \mathbb{D} \\
&= (X \cup \{\{\underbrace{x_\circ, \ldots, x_\circ}_{N-|X|}\}\}) \cap \mathbb{D} \\
&= X.
\end{aligned}$$

Therefore, we have $\Phi'^{-1}(U) = \Phi^{-1}\big(U - \mathrm{const}\big) \cap \mathbb{D}$ for all $U \in \Phi'(\mathbb{X}^l_{\mathbb{D},[N]}) = \{\Phi'(X) : X \in \mathbb{X}^l_{\mathbb{D},[N]}\}$. If we define $\rho = f \circ (\Phi')^{-1}$ where $\mathrm{dom}(\rho) = \Phi'(\mathbb{X}^l_{\mathbb{D},[N]})$, then we have $f(X) = \rho \circ \Phi'(X)$ for all $X \in \mathbb{X}^l_{\mathbb{D},[N]}$. We arrive at the exact form of sum-decomposition by renaming Φ' to Φ.

Appendix G. Proof of Proposition 3

Let $X \in \mathbb{X}_{\mathbb{Q}^D,N}$. For any rational-valued vectors $x, x' \in X$ such that $l(x) = l(x')$, we have

$$\text{const} \sum_{d\in[D]} (x_d - x'_d) \log \zeta(d) = 0, \tag{15}$$

where x_d and x'_d are d-th elements of x and x', and $\text{const} \in \mathbb{N}$ is such that $y_d \stackrel{\text{def}}{=} \text{const}(x_d - x'_d) \in \mathbb{Z}$ — for all $d \in [D]$. From equation (15), we have

$$\sum_{d\in[D]} y_d \log \zeta(d) = 0 \;\rightarrow\; \prod_{d\in[D]} \zeta(d)^{y_d} = 1.$$

Therefore, we have

$$\prod_{\substack{d\in[D]\\ y_d>0}} \zeta(d)^{y_d} = \prod_{\substack{d\in[D]\\ -y_d>0}} \zeta(d)^{-y_d} = n \in \mathbb{N} \tag{16}$$

Both sides of equation (16) are prime number decompositions of an integer $n \in \mathbb{N}$ with completely exclusive set of prime numbers. Therefore, we have $n = 1$ which results in $y_d = \text{const}(x_d - x'_d) = 0$ for all $d \in [D]$, that is, $x = x'$. This proves the following result:

$$\forall x, x' \in X\big(\in \mathbb{X}_{\mathbb{Q}^D,N}\big) : \; l(x) = l(x') \longrightarrow \; x = x'.$$

Finally, since l is a continuous linear function on $\mathbb{R}^D$, $\mathbb{X}_{\mathbb{Q}^D,N}$ is an l-identifiable set.

Appendix H. Proof of Lemma 1

We need to show for any $X \in \mathbb{X}_{\mathbb{D},N}$, there exists a sequence $\{X_n \in \mathbb{X}_{Q(\mathbb{D}),N} : n \in \mathbb{N}\}$ such that $\lim_{n\to\infty} \Phi(X_n) = \Phi(X)$. From Lemma 6, Φ is a continuous map. Therefore, we simply need to prove the following property:

$$\forall X \in \mathbb{X}_{\mathbb{D},N} : \lim_{n\to\infty} X_n = X,$$

where $X_n \in \mathbb{X}_{Q(\mathbb{D}),N}$ for all $n \in \mathbb{N}$. By definition, we want show $\forall \varepsilon > 0, \exists N(\varepsilon) \in \mathbb{N}$ such that

$$\forall n \geq N(\varepsilon) : d_M(X_n^m, X) < \varepsilon.$$

Let $\mathcal{N}_n(x) = \{y \in Q(\mathbb{D}) : \|x-y\|_2 \leq \frac{1}{n}\}$ be a bounded set centered at $x \in \mathbb{D}$ and $n \in \mathbb{N}$. It is important to note that $\mathcal{N}_n(x)$ is a nonempty set for all $n \in \mathbb{N}$, that is, the intersection of $Q(\mathbb{D})$ with the nonempty interior of $\mathbb{D}$ is nonempty because $Q(\mathbb{D})$ is a dense subset of $\mathbb{D}$. We let $q_n(x)$ be **any** random point in $\mathcal{N}_n(x)$. Then, for any $X \in \mathbb{X}_{\mathbb{D},N}$, we let $X_n = \{\{q_n(x) : x \in X\}\} \in \mathbb{X}_{Q(\mathbb{D}),N}$. By construction, we have

$$d_M(X_n, X) \leq \sqrt{N \max_{x\in\mathbb{D}} \|x - q_n(x)\|_2^2} \leq \sqrt{N} n^{-1}.$$

If we let $N(\varepsilon) = \lfloor \frac{\sqrt{N}}{\varepsilon} \rfloor + 1$, then $d_M(X_n, X) \leq \varepsilon$ for all $n \geq N(\varepsilon)$. Therefore, we have $\lim_{n\to\infty} d_M(X_n, X) = 0$, that is, $\lim_{n\to\infty} \mathcal{X}_n^m = X$. Any realization of the random process $\{X_n\}_{n\in\mathbb{N}}$ forms a sequence in $\mathbb{X}_{Q(\mathbb{D}),N}$ that converges to X, that is, $\mathbb{X}_{Q(\mathbb{D}),N}$ is a dense subset of $\mathbb{X}_{\mathbb{D},N}$.

The function Φ in Theorem 5 is continuous; see Fact 6 and Lemma 6. Furthermore, we showed that $\mathbb{X}_{Q(\mathbb{D}),N}$ is a dense subset of $\mathbb{X}_{\mathbb{D},N}$. Therefore, for any $U \in \Phi(\mathbb{X}_{\mathbb{D},N})$ there exists (at least) a $X \in \mathbb{X}_{\mathbb{D},N}$ such that $U = \Phi(X)$. Let $\{X_n \in \mathbb{X}_{Q(\mathbb{D}),N} : n \in \mathbb{N}\}$ be such that $\lim_{n\to\infty} X_n = X$. Since Φ is a continuous map, we have $\lim_{n\to\infty} \Phi(X_n) = \Phi(X)$. That is, there exists a sequence $\{U_n = \Phi(X_n) \in \Phi(\mathbb{X}_{Q(\mathbb{D}),N}) : n \in \mathbb{N}\}$ such that $\lim_{n\to\infty} U_n = U$. This proves that $\Phi(\mathbb{X}_{Q(\mathbb{D}),N})$ is a dense subset of $\Phi(\mathbb{X}_{\mathbb{D},N})$. This completes the proof.

Appendix I. Proof of Theorem 6

Fact 7 *Let $\mathbb{D}$ be a compact subset of $\mathbb{R}^D$ with nonempty interior. The function ϕ in Proposition 3 is continuous. Its associated multiset function $\Phi : \mathbb{X}_{\mathbb{D},N} \to \mathrm{codom}(\Phi)$ is a continuous function (see Lemma 4)*

From Fact 7 and Corollary 1, there exist a continuous multiset function and

$$\Phi : \mathbb{X}_{\mathbb{D},N} \to \mathrm{codom}(\Phi) \subset \mathbb{C}^{D\times N},$$

and $\rho : \Phi(\mathbb{X}_{Q(\mathbb{D}),N}) \to \mathrm{codom}(\rho)$ such that

$$\forall X \in \mathbb{X}_{Q(\mathbb{D}),N} :\ f(X) = \rho\big(\sum_{x\in X} \phi(x)\big) = \rho \circ \Phi(X).$$

In this proof, we want to define the function $\rho_e : \Phi(\mathbb{X}_{\mathbb{D},N}) \to \mathrm{codom}(\rho_e)$ as follows:

$$\forall Z \in \Phi(\mathbb{X}_{\mathbb{D},N}) : \rho_e(Z) = \lim_{Z_n \to Z} \rho(Z_n), \tag{17}$$

where $Z_n \in \Phi(\mathbb{X}_{Q(\mathbb{D}),N})$ for all $n \in \mathbb{N}$. The goal is to show that ρ_e is (1) well-defined and (2) continuous over its compact domain $\Phi(\mathbb{X}_{\mathbb{D},N})$. If these two conditions are valid, we let $Z = \Phi(X) \in \Phi(\mathbb{X}_{Q(\mathbb{D}),N})$ where $X \in \mathbb{X}_{Q(\mathbb{D}),N}$ and $\{Z_n \in \Phi(\mathbb{X}_{Q(\mathbb{D}),N}) : Z_n = Z, n \in \mathbb{N}\}$. Then, we have

$$\forall X \in \mathbb{X}_{Q(\mathbb{D}),N} :\ f(X) = \rho_e \circ \Phi(X) = \rho \circ \Phi(X).$$

Proposition 9 (Well-definedness) *Let $\mathcal{Z} \stackrel{\text{def}}{=} \{Z_n \in \Phi(\mathbb{X}_{Q(\mathbb{D}),N}) : n \in \mathbb{N}\}$ be the convergent sequence, that is, $\lim_{n\to\infty} Z_n = Z$. Given a continuous multiset function $f : \mathbb{X}_{\mathbb{D},N} \to \mathrm{codom}(f)$, let $\rho : \Phi(\mathbb{X}_{Q(\mathbb{D}),N}) \to \mathrm{codom}(\rho) \subset f(\mathbb{X}_{\mathbb{D},N})$ be defined in Corollary 1. Then, the sequence $\rho(\mathcal{Z}) \stackrel{\text{def}}{=} \{\rho(Z_n) : n \in \mathbb{N}\}$ is convergent to a unique point in $f(\mathbb{X}_{\mathbb{D},N})$. The term $\lim_{n\to\infty} \rho(Z_n)$ only depends on Z, and not specific choice of the sequence $\mathcal{Z}$.*

As a result of Proposition 9, the function $\rho_e : \Phi(\mathbb{X}_{\mathbb{D},N}) \to \mathrm{codom}(\rho_e) \subseteq f(\mathbb{X}_{\mathbb{D},N})$ is well-defined. That is, $\lim_{Z_n\to Z} \rho(Z_n)$ does not depend on the specific convergent sequence $\mathcal{Z}$ so long as its limiting point — $\lim_{n\to\mathbb{N}} Z_n = Z$ — is fixed.

Proposition 10 (Continuity) *The function ρ_e is continuous on the compact domain $\Phi(\mathbb{X}_{\mathbb{D},N})$.*

In summary, we have

$$\forall X \in \mathbb{X}_{Q(\mathbb{D}),N} : f(X) = \rho_e \circ \Phi(X).$$

where $\rho_e : \Phi(\mathbb{X}_{\mathbb{D},N}) \to \mathrm{codom}(\rho_e)$ and $\Phi : \mathbb{X}_{\mathbb{D},N} \to \mathrm{codom}(\Phi)$ are continuous functions. Therefore, $\rho_e \circ \Phi$ is a continuous function on $\mathbb{X}_{\mathbb{D},N}$. Since $\mathbb{X}_{Q(\mathbb{D}),N}$ is a dense subset of $\mathbb{X}_{\mathbb{D},N}$ (see Lemma 1) and $f : \mathbb{X}_{\mathbb{D},N} \to \mathrm{codom}(f)$ is a continuous multiset function, we have

$$\forall X \in \mathbb{X}_{\mathbb{D},N} :\ f(X) = \lim_{n\to\infty} f(X_n)$$

for any sequence $\{X_n \in \mathbb{X}_{Q(\mathbb{D}),N} : n \in \mathbb{N}\}$ where $\lim_{n\to\infty} X_n = X$. Therefore, we have

$$\forall X \in \mathbb{X}_{\mathbb{D},N} :\ f(X) = \lim_{n\to\infty} \rho_e \circ \Phi(X_n).$$

Since $\rho_e \circ \Phi$ is a continuous function on $\mathbb{X}_{\mathbb{D},N}$, we have

$$\forall X \in \mathbb{X}_{\mathbb{D},N} : \; f(X) = \lim_{n\to\infty} \rho_e \circ \Phi(X_n) = \rho_e \circ \Phi(\lim_{n\to\infty} X_n) = \rho_e \circ \Phi(X).$$

We argue that ρ_e has a continuous extension to $\mathbb{C}^{D\times N}$. The set $\mathbb{X}_{\mathbb{D},N}$ is a compact set. From Lemma 6, $\Phi(\mathbb{X}_{\mathbb{D},N})$ is also a compact set. Finally, Fact 1 shows this continuous extension is admitted. After renaming ρ_e to ρ, we arrive at the exact statement of the theorem.

I.1. Proof of Proposition 9

Lemma 9 *Let $\mathcal{Z} \overset{\text{def}}{=} \{Z_n \in \Phi(\mathbb{X}_{Q(\mathbb{D}),N}) : n \in \mathbb{N}\}$ be the convergent sequence, that is,*

$$\lim_{n\to\infty} Z_n = Z \in \Phi(\mathbb{X}_{\mathbb{D},N}).$$

Given a continuous multiset function $f : \mathbb{X}_{\mathbb{D},N} \to \mathrm{codom}(f)$, let

$$\rho : \Phi(\mathbb{X}_{Q(\mathbb{D}),N}) \to \mathrm{codom}(\rho) \subset f(\mathbb{X}_{\mathbb{D},N})$$

be defined in Corollary 1. The sequence $\rho(\mathcal{Z}) \overset{\text{def}}{=} \{\rho(Z_n) : n \in \mathbb{N}\}$ is Cauchy in compact metric space $(f(\mathbb{X}_{\mathbb{D},N}), \|\cdot\|_2)$.

Theorem 10 *(Attenborough, 2003) A Cauchy sequence in a compact metric space is convergent to a point in the metric space.*

Lemma 10 *Let $\mathcal{Z} \overset{\text{def}}{=} \{Z_n \in \Phi(\mathbb{X}_{Q(\mathbb{D}),N}) : n \in \mathbb{N}\}$ be the convergent sequence, that is, $\lim_{n\to\infty} Z_n = Z \in \Phi(\mathbb{X}_{\mathbb{D},N})$. Given a continuous multiset function $f : \mathbb{X}_{\mathbb{D},N} \to \mathrm{codom}(f)$, let $\rho : \Phi(\mathbb{X}_{Q(\mathbb{D}),N}) \to \mathrm{codom}(\rho) \subset f(\mathbb{X}_{\mathbb{D},N})$ be defined in Corollary 1. The sequence $\rho(\mathcal{Z}) \overset{\text{def}}{=} \{\rho(Z_n) : n \in \mathbb{N}\}$ is convergent to a unique point in $f(\mathbb{X}_{\mathbb{D},N})$. The term $\lim_{n\to\infty} \rho(Z_n)$ only depends on $Z = \lim_{n\to} Z_n$.*

I.1.1. Proof of Lemma 9

Fact 8 *Every convergent sequence is Cauchy. Hence, the convergent sequence $\mathcal{Z} \overset{\text{def}}{=} \{Z_n : n \in \mathbb{N}\}$ is Cauchy in $(\Phi(\mathbb{X}_{Q(\mathbb{D}),N}), \|\cdot\|_F)$.*

From Fact 8, for any $\delta > 0$, there exists $N(\delta) \in \mathbb{N}$ such that $\|Z_{n_1} - Z_{n_2}\|_F < \delta$ for all $n_1, n_2 > N(\delta)$. Therefore, we have

$$\forall n_1, n_2 > N(\delta) : \|\Phi(X_{n_1}) - \Phi(X_{n_2})\|_F < \delta. \tag{18}$$

where $X_n = \Phi^{-1}(Z_n)$ for all $n \in \mathbb{N}$. The set $\mathbb{X}_{Q(\mathbb{D}),N}$ is an l-identifiable subset of $\mathbb{X}_{\mathbb{D},N}$.

Proposition 11 *Let $\mathbb{X}_{\mathbb{R}^D/l,N}$ be an l-identifiable set, and $\Phi(\mathbb{X}_{\mathbb{R}^D/l,N}) = \{\Phi(X) : X \in \mathbb{X}_{\mathbb{R}^D/l,N}\}$ where Φ is defined in equations (13) and (14). The function $\Phi^{-1} : \Phi(\mathbb{X}_{\mathbb{R}^D/l,N}) \to \mathbb{X}_{\mathbb{R}^D/l,N}$ is defined in the proof of Lemma 8. We claim that Φ^{-1} is a continuous function on its domain.*

From Proposition 11, Φ^{-1} is a continuous function on $\Phi(\mathbb{X}_{Q(\mathbb{D}),N})$. Since f is a continuous multiset function on its domain, the function $\rho = f \circ \Phi^{-1}$ is continuous on $\Phi(\mathbb{X}_{Q(\mathbb{D}),N})$. By definition of continuity, for any $\varepsilon > 0$ and $\Phi(X)$ where $X \in \mathbb{X}_{Q(\mathbb{D}),N}$, there exists $\delta(\varepsilon) > 0$ such that

$$\forall X' \in \mathbb{X}_{Q(\mathbb{D}),N} : \|\Phi(X) - \Phi(X')\|_F < \delta(\varepsilon) \to \|f(X) - f(X')\|_2 < \varepsilon, \tag{19}$$

where $X = \Phi^{-1} \circ \Phi(X)$ and $X' = \Phi^{-1} \circ \Phi(X')$.

Comparing equation (18) to the left-hand-side of equation (19) — letting $Z_{n_1} = \Phi(X)$ and $Z_{n_2} = \Phi(X')$ — we have

$$\forall n_1, n_2 > N(\delta(\varepsilon)) : \|f \circ \Phi^{-1}(Z_{n_1}) - f \circ \Phi^{-1}(Z_{n_2})\|_2 < \varepsilon,$$

that is, $f \circ \Phi^{-1}(\mathcal{Z}) = \rho(\mathcal{Z})$ is a Cauchy sequence; see Corollary 1 and proof of Theorem 5. Finally, Lemma 5 and the following fact show that $f(\mathbb{X}_{\mathbb{D},N})$ is indeed a compact set, that is, $(f(\mathbb{X}_{\mathbb{D},N}), \|\cdot\|_2)$ is a sequentially compact metric space.

Fact 9 *(Pugh and Pugh 2002) The image of a compact set under continuous map is a compact set.*

Proof of Proposition 11 Let $\varepsilon > 0$ and $Z \in \Phi(\mathbb{X}_{\mathbb{R}^D/l,N})$ — that is, $Z = \Phi(X) \in \mathbb{C}^{D\times N}$ for a unique $X \in \mathbb{X}_{\mathbb{R}^D/l,N}$. We define $\mathbb{D}_\Phi(\varepsilon, Z) = \{Z' \in \Phi(\mathbb{X}_{\mathbb{R}^D/l,N}) : \|Z - Z'\|_F < \varepsilon\}$. For any $Z' \in \mathbb{D}_\Phi(\varepsilon, Z)$, we have

$$\begin{aligned}
d_M(\Phi^{-1}(Z), \Phi^{-1}(Z')) &\overset{(a)}{\leq} \sup_{\substack{\mathrm{dZ}\in\mathbb{C}^{D\times N}:\\ Z+\mathrm{d}Z\in\mathbb{D}_\Phi(\varepsilon,Z)}} d_M(X, \Phi^{-1}\circ(Z+\mathrm{d}Z))\\
&\overset{(b)}{=} \sup_{\substack{\mathrm{dZ}\in\mathbb{C}^{D\times N}:\\ Z+\mathrm{d}Z\in\mathbb{D}_\Phi(\varepsilon,Z)}} d_M\Big(X, \Phi^{-1}\Big(\begin{pmatrix}\Phi_{\text{deep}}(\{\{e_1^\top x + l(x)j : x\in X\}\})\\ \vdots \\ \Phi_{\text{deep}}(\{\{e_D^\top x + l(x)j : x\in X\}\})\end{pmatrix} + \mathrm{dZ}\Big)\Big)\\
&\overset{(c)}{=} \sup_{\substack{\mathrm{dZ}\in\mathbb{C}^{D\times N}:\\ Z+\mathrm{d}Z\in\mathbb{D}_\Phi(\varepsilon,Z)}} d_M\Big(X, \mathrm{sortvec}\Big(\begin{pmatrix}\{\{e_1^\top x + l(x)j + r_1(x,\mathrm{dz}_1;X) : x\in X\}\}\\ \vdots \\ \{\{e_D^\top x + l(x)j + r_D(x,\mathrm{dz}_D;X) : x\in X\}\}\end{pmatrix}\Big)\Big)
\end{aligned}$$

where (a) is due to the fact that we have $\|Z - Z'\|_F < \varepsilon$ and $\Phi^{-1} = \mathrm{sortvec} \circ \Phi_{\text{deep}}^{-1}$ is well-defined for $Z, Z' \in \Phi(\mathbb{X}_{\mathbb{R}^D/l,N})$, (b) is due to the definition of Φ — see equations (13) and (14) — and the fact that $Z = \Phi(X)$, (c) is due letting $\mathrm{dz}_d = e_D^\top \mathrm{d}Z \in \mathbb{C}^N$ where e_d is the d-th standard basis of $\mathbb{R}^D$ — for all $d \in [D]$ — and the fact that Φ_{deep}^{-1} is a continuous function (Zaheer et al., 2017), that is,

$$\forall d \in [D], X \in \mathbb{X}_{\mathbb{R}^D/l,N}, x \in X : \lim_{\mathrm{dz}\in\mathbb{C}^N:\mathrm{dz}\to 0} r_d(x, \mathrm{dz}; X) = 0.$$

For any $\varepsilon > 0$, $x \in X$, and $d \in [D]$, there exists a finite $\delta_d(\varepsilon, x; X) = \sup_{\mathrm{dz}\in\mathbb{D}(\varepsilon)} |r_d(x, \mathrm{dz}; X)|$ where $\mathbb{D}(\varepsilon) = \{z \in \mathbb{C}^N : \|z\|_2 < \varepsilon\}$ and $\lim_{\varepsilon\to 0} \delta_d(\varepsilon, x; X) = 0$. For any $\varepsilon > 0$ and $X \in \mathbb{X}_{\mathbb{R}^D/l,N}$, we have

$$\delta^*(\varepsilon, X) \overset{\text{def}}{=} \max_{d\in[D],x\in X} \delta_d(\varepsilon, x; X) = \sup_{d\in[D],x\in X,\mathrm{dz}\in\mathbb{D}(\varepsilon)} |r_d(x, \mathrm{dz}; X)|, \tag{20}$$

where $\lim_{\varepsilon\to 0} \delta^*(\varepsilon, X) = 0$. Let $X = \{\{x_n : n \in [N]\}\}$. Then, we have

$$\forall d \in [D] : \mathrm{sort}(\{\{e_d^\top x + l(x)j : x \in X\}\}) = \big(e_d^\top x_{\pi(n)}\big)_{n\in[N]} \in \mathbb{R}^N$$

where $\pi : [N] \to [N]$ is a permutation operator such that $l(x_{\pi(1)}) \leq \cdots \leq l(x_{\pi(N)})$. Even though the permutation operator π may not be unique, sort is a well-defined function; see the proof of Theorem 2. We let $S(X) \overset{\text{def}}{=} \{\varepsilon : \delta^*(\varepsilon, X) < \psi(X)\}$ and $\psi(X) \overset{\text{def}}{=} \min_{\substack{x,x' \in X \\ x \neq x'}} \frac{1}{2}|l(x) - l(x')| > 0$ where $l : \mathbb{R}^D \to \mathbb{R}$ is the continuous identifier function. From equation (20), we have

$$\forall d \in [D], \varepsilon \in S(X), x \in X, \mathrm{dz} \in \mathbb{D}(\varepsilon) : \mathrm{Im}(r_d(x, \mathrm{dz}; X)) < \delta^*(\varepsilon, X) < \psi(X),$$

that is,

$$\forall d \in [D], \varepsilon \in S(X), x \in X, \mathrm{dz} \in \mathbb{D}(\varepsilon) : \mathrm{Im}(r_d(x, \mathrm{dz}; X)) < \min_{\substack{x,x' \in X \\ x \neq x'}} \frac{1}{2}|l(x) - l(x')|.$$

Therefore, dZ perturbs the imaginary components of $\{\{e_d^\top x + l(x)j + r_d(x, \mathrm{dz}_d; X) : x \in X\}\}$ (for any $d \in [D]$) by at most $\delta^*(\varepsilon, X) < \min_{\substack{x,x' \in X \\ x \neq x'}} \frac{1}{2}|l(x) - l(x')|$ and distinct elements do not switch place after adding the perturbation dZ. More precisely, for all $d \in [D]$, $\varepsilon \in S(X)$, and $\mathrm{dz} \in \mathbb{D}(\varepsilon)$, we have

$$\mathrm{sort}(\{\{e_d^\top x + l(x)j + r_d(x, \mathrm{dz}; X) : x \in X\}\}) = \left(e_d^\top x_{\pi'(n)} + \mathrm{Re}(r_d(x_{\pi'(n)}, \mathrm{dz}; X))\right)_{n \in [N]} \in \mathbb{R}^N,$$

where $\pi' : [N] \to [N]$ is such that for all $\varepsilon \in S(X)$, we have

$$\forall \mathrm{dz} \in \mathbb{D}(\varepsilon) : l(x_{\pi'(1)}) + \mathrm{Im}(r_d(x_{\pi'(1)}, \mathrm{dz}; X)) \leq \cdots \leq l(x_{\pi'(N)}) + \mathrm{Im}(r_d(x_{\pi'(N)}, \mathrm{dz}; X)).$$

Since $\mathrm{Im}(r_d(x, \mathrm{dz}; X)) < \min_{\substack{x,x' \in X \\ x \neq x'}} \frac{1}{2}|l(x) - l(x')|$, we also have the following inequalities:

$$l(x_{\pi'(1)}) \leq \cdots \leq l(x_{\pi'(N)}). \tag{21}$$

Remark 11 *The permutation operator π' may vary with dz and x, if two (or more) elements of $\{\{l(x) : x \in X\}\}$ are identical. A proper notation should be $\pi'(\mathrm{dz}, x; X)$. For simplicity in notation, we avoid expressing this proper parameterization.*

Remark 12 *The perturbation dz may switch the rank (or position) of two elements only if they are equal to each other, that is, if $l(x_{\pi'(1)}) = l(x_{\pi'(2)})$, then we may have*

$$l(x_{\pi'(1)}) + \mathrm{Im}(r_d(x_{\pi'(1)}, \mathrm{dz}; X)) < l(x_{\pi'(2)}) + \mathrm{Im}(r_d(x_{\pi'(2)}, \mathrm{dz}; X)).$$

This does not provide any issue, since $l(x_{\pi'(1)}) \leq l(x_{\pi'(2)})$. In short, independent of dz and x, π' is such that $l(x_{\pi'(1)}) \leq \cdots \leq l(x_{\pi'(N)})$.

From equation (21) and the definition of π, we have $l(x_{\pi(n)}) = l(x_{\pi'(n)})$ for all $n \in [N]$ — even though, we may have $\pi \neq \pi'$. From Definition 2, since $l(x_{\pi(n)}) = l(x_{\pi'(n)})$, we have $x_{\pi'(n)} = x_{\pi(n)}$ for all $n \in [N]$. Consequently, for all $d \in [D]$, $\varepsilon \in S(X)$ and $\mathrm{dz} \in \mathbb{D}(\varepsilon)$, we have

$$\mathrm{sort}(\{\{e_d^\top x + l(x)j + r_d(x, \mathrm{dz}; X) : x \in X\}\}) = \left(e_d^\top x_{\pi'(n)} + \mathrm{Re}(r_d(x_{\pi'(n)}, \mathrm{dz}; X))\right)_{n \in [N]}$$
$$\mathrm{sort}(\{\{e_d^\top x + l(x)j : x \in X\}\}) = \left(e_d^\top x_{\pi(n)}\right)_{n \in [N]} = \left(e_d^\top x_{\pi'(n)}\right)_{n \in [N]}.$$

Therefore, even though there are multiple permutation operators that sorts the elements of $\{\{e_d^\top x + l(x)j + r_d(x, \mathrm{dz}; X) : x \in X\}\}$, the output of the sort function gives an ordering that remains unchanged for distinct elements of X for any $x \in X \in \mathbb{X}_{\mathbb{R}^D/l,N}$ and $\mathrm{dz} \in \mathbb{D}(\varepsilon)$ where $\varepsilon \in S(X)$. Consequently, we have

$$\begin{aligned}
&d_M(\Phi^{-1}(Z), \Phi^{-1}(Z')) \\
&\leq \sup_{\substack{\mathrm{dZ}\in\mathbb{C}^{D\times N}: \\ Z+\mathrm{d}Z\in\mathbb{D}_\Phi(\varepsilon,Z)}} d_M(X, \mathrm{sortvec}(\begin{pmatrix} \{\{e_1^\top x + l(x)j + r_1(x, \mathrm{dz}_1; X) : x \in X\}\} \\ \vdots \\ \{\{e_D^\top x + l(x)j + r_D(x, \mathrm{dz}_D; X) : x \in X\}\} \end{pmatrix})) \\
&\overset{\text{(a)}}{\leq} \sup_{\substack{\mathrm{dZ}\in\mathbb{C}^{D\times N}: \\ Z+\mathrm{d}Z\in\mathbb{D}_\Phi(\varepsilon,Z)}} d_M(X, \{\{\begin{pmatrix} e_1^\top x_{\pi_1(n)} + \mathrm{Re}(r_1(x_{\pi_1(n)}, \mathrm{dz}_1; X)) \\ \vdots \\ e_D^\top x_{\pi_D(n)} + \mathrm{Re}(r_D(x_{\pi_D(n)}, \mathrm{dz}_D; X)) \end{pmatrix} : n \in [N]\}\}) \\
&\overset{\text{(b)}}{\leq} \sup_{\substack{\mathrm{dZ}\in\mathbb{C}^{D\times N}: \\ Z+\mathrm{d}Z\in\mathbb{D}_\Phi(\varepsilon,Z)}} d_M\Big(\{\{\begin{pmatrix} e_1^\top x_{\pi_1(n)}) \\ \vdots \\ e_D^\top x_{\pi_D(n)}) \end{pmatrix} : n \in [N]\}\}, \\
&\qquad\qquad \{\{\begin{pmatrix} e_1^\top x_{\pi_1(n)} + \mathrm{Re}(r_1(x_{\pi_1(n)}, \mathrm{dz}_1; X)) \\ \vdots \\ e_D^\top x_{\pi_D(n)} + \mathrm{Re}(r_D(x_{\pi_D(n)}, \mathrm{dz}_D; X)) \end{pmatrix} : n \in [N]\}\}\Big) \\
&\overset{\text{(c)}}{\leq} \sup_{\substack{\mathrm{dZ}\in\mathbb{C}^{D\times N}: \\ Z+\mathrm{d}Z\in\mathbb{D}_\Phi(\varepsilon,Z)}} \sqrt{\sum_{n\in[N]} \| \begin{pmatrix} \mathrm{Re}(r_1(x_{\pi_1(n)}, \mathrm{dz}_1; X)) \\ \vdots \\ \mathrm{Re}(r_D(x_{\pi_D(n)}, \mathrm{dz}_D; X)) \end{pmatrix} \|_2^2} \\
&\overset{\text{(d)}}{\leq} \sqrt{DN} \sup_{d\in[D], x\in X, \mathrm{dz}\in\mathbb{D}(\varepsilon)} |r_d(x, \mathrm{dz}; X)| = \sqrt{DN}\delta^*(\varepsilon, X)
\end{aligned}$$

where (a) uses permutation operators $\pi_d : [N] \to [N]$ and it depends on elements of $\{\{r_d(x, \mathrm{dz}_1; X) : x \in X\}$, but $x_{\pi_d(n)} = x_{\pi(n)}$ for all $n \in [N]$ and $d \in [D]$, (b) follows from the fact that $x_{\pi_d(n)} = x_{\pi(n)}$ for all $n \in [N]$ and $d \in [D]$, (c) follows from the definition of the matching distance d_M, and (d) follows from the fact that if $\mathrm{dZ} \in \mathbb{C}^{D\times N}$ is such that $Z + \mathrm{dZ} \in \mathbb{D}_\Phi(\varepsilon, Z)$, then its individual rows $\mathrm{dz}_1, \ldots, \mathrm{dz}_D \in \mathbb{R}^N$ have norms upper bounded by ε, that is, $\mathrm{dz}_d \in \mathbb{D}(\varepsilon)$ and $|\mathrm{Re}(r_d(x, \mathrm{dz}_d; X))| \leq |r_d(x, \mathrm{dz}_d; X)|$ for all $d \in [D]$.

Continuity Statement. For any $Z = \Phi(X) \in \Phi(\mathbb{X}_{\mathbb{R}^D,N})$ and $\delta > 0$, there exists a positive $\epsilon(\delta) \in \{\varepsilon' \in S(X) : \sqrt{DN}\delta^*(\varepsilon', X) < \delta\}$ where

$$\forall Z' \in \Phi(\mathbb{X}_{\mathbb{R}^D,N}) : \|Z - Z'\|_F < \epsilon(\delta) \to d_M(\Phi^{-1}(Z), \Phi^{-1}(Z')) < \delta.$$

I.1.2. PROOF OF LEMMA 10

Let $\mathcal{Z}_1 = \{Z_{1,n} \in \Phi(\mathbb{X}_{Q(\mathbb{D}),N}) : n \in \mathbb{N}\}$ and $\mathcal{Z}_2 = \{Z_{2,n} \in \Phi(\mathbb{X}_{Q(\mathbb{D}),N}) : n \in \mathbb{N}\}$ be two sequences such that $\lim_{n\to\infty} Z_{1,n} = \lim_{n\to\infty} Z_{2,n} = Z$.
From Lemma 9, the following limits are well-defined:

$$\lim_{n\to\infty} \rho(Z_{1,n}) = f_1, \ \lim_{n\to\infty} \rho(Z_{2,n}) = f_2 \in f(\mathbb{X}_{\mathbb{D},N}).$$

We construct $\mathcal{Z} = \{Z_n : n \in \mathbb{N}\}$ in $\Phi(\mathbb{X}_{Q(\mathbb{D}),N})$ where $Z_{2n} = Z_{1,n}$ and $Z_{2n+1} = Z_{2,n}$ for all $n \in \mathbb{N}$. By construction, we have $\lim_{n\to\infty} Z_n = Z$. Since all convergent sequences are Cauchy, $\mathcal{Z}$ is a Cauchy sequence. Therefore, from our discussion the proof of Lemma 9, the sequence $\rho(\mathcal{Z})$ must converge to $f^* \in f(\mathbb{X}_{\mathbb{D},N})$.

Fact 10 *Every subsequence of a convergent sequence converges to the same limit as the original sequence.*

Both $\rho(\mathcal{Z}_1)$ and $\rho(\mathcal{Z}_2)$ are subsequences of the convergent sequence $\rho(\mathcal{Z})$. Therefore, we have

$$\lim_{n\to\infty} \rho(Z_{1,n}) = \lim_{n\to\infty} \rho(Z_{2,n}) = \lim_{n\to\infty} \rho(Z_n) = f^*,$$

that is, $f_1 = f_2$. Therefore, the limit of $\rho(\mathcal{Z})$ only depends on the limit of the sequence $\mathcal{Z}$.

I.2. Proof of Proposition 10

We want to show that, for any $\Phi(X) \in \Phi(\mathbb{X}_{\mathbb{D},N})$ and $\varepsilon > 0$, there is $\delta(\varepsilon) > 0$ such that

$$\forall X' \in \mathbb{X}_{\mathbb{D},N} : \|\Phi(X) - \Phi(X')\|_F < \delta(\varepsilon) \to \|\rho_e \circ \Phi(X) - \rho_e \circ \Phi(X')\| < \varepsilon. \tag{22}$$

We first use the definition of ρ_e to reforulate the left-hand-side of equation (22) in terms of convergent sequences in $\Phi(\mathbb{X}_{Q(\mathbb{D}),N})$. This is formalized in Lemma 11.

Lemma 11 *Let $X, X' \in \mathbb{X}_{\mathbb{D},N}$. There exist convergent sequences $\mathcal{Z}_x \stackrel{\text{def}}{=} \{Z_{x,n} \in \Phi(\mathbb{X}_{Q(\mathbb{D}),N}) : n \in \mathbb{N}\}$ and $\mathcal{Z}_y \stackrel{\text{def}}{=} \{Z_{y,n} \in \Phi(\mathbb{X}_{Q(\mathbb{D}),N}) : n \in \mathbb{N}\}$ and $N_x(\delta), N_y(\delta) \in \mathbb{N}$ such that*

$$\begin{aligned} \forall n > N_x(\delta) :\ & \|Z_{x,n} - \Phi(X)\|_2 < \delta \\ \forall n > N_y(\delta) :\ & \|Z_{y,n} - \Phi(X')\|_2 < \delta. \end{aligned}$$

for any $\delta > 0$. If $\|\Phi(X) - \Phi(X')\|_2 < \delta$, then we have

$$\forall n > N(\delta) \stackrel{\text{def}}{=} \max\{N_x(\delta), N_y(\delta)\} :\ \|Z_{x,n} - Z_{y,n}\|_2 < 3\delta.$$

As the result of Lemma 11, the left-hand-side of equation (22) gives us the following inequality:

$$\forall X' \in \mathbb{X}_{\mathbb{D},N} : \|\Phi(X) - \Phi(X')\|_F < \delta, n > N(\delta) \to \|Z_{x,n} - Z_{y,n}\|_2 < 3\delta,$$

where $N(\delta) \in \mathbb{N}$, $\mathcal{Z}_x = \{Z_{x,n} : n \in \mathbb{N}\}$ and $\mathcal{Z}_y = \{Z_{y,n} : n \in \mathbb{N}\}$ are convergent sequences in $\Phi(\mathbb{X}_{Q(\mathbb{D}),N})$ (in Lemma 11), that is,

$$\lim_{n\to\infty} Z_{x,n} = \Phi(X),\ \lim_{n\to\infty} Z_{y,n} = \Phi(X') \in \Phi(\mathbb{X}_{\mathbb{D},N}).$$

In Lemma 11 we prove that convergent sequences $\mathcal{Z}_x$ and $\mathcal{Z}_y$ become arbitrary close to each other as $\delta \to 0$. In Lemma 12, we use the fact that ρ (not ρ_e) is a continuous function on noncompact domain $\Phi(\mathbb{X}_{Q(\mathbb{D}),N})$, and argue that $\|\rho(Z_{x,n}) - \rho(Z_{y,n})\|_2$ converges to zero as $\delta \to 0$.

Lemma 12 *For all $Z \in \Phi(\mathbb{X}_{Q(\mathbb{D}),N})$ and $u > 0$, there exists a $\gamma(u) > 0$ such that*

$$\forall Z' \in \Phi(\mathbb{X}_{Q(\mathbb{D}),N}) : \|Z - Z'\|_F < \gamma(u) \to \|\rho(Z) - \rho(Z')\|_2 < u.$$

Note that γ shows the **functional** *dependence of the upper bound for $\|Z - Z'\|_F$ for any value of $\|\rho(Z) - \rho(Z')\|_2$. For any $\delta > 0$, we have*

$$\forall n > N^{'}(\delta) : \|\rho(Z_{x,n}) - \rho(Z_{y,n})\|_2 < \delta.$$

where $N^{'}(\delta) = N(\min\{\frac{\delta}{3}, \frac{\gamma(\delta)}{3}\})$, and $N(\delta) \in \mathbb{N}$, convergent sequences $\mathcal{Z}_x = \{Z_{x,n} : n \in \mathbb{N}\}$ and $\mathcal{Z}_y = \{Z_{y,n} : n \in \mathbb{N}\}$ are defined in Lemma 11.

We now use Lemma 12 to show that for all $\delta > 0$, we have

$$\forall X' \in \mathbb{X}_{\mathbb{D},N} : \|\Phi(X) - \Phi(X')\|_F < \delta, n > N^{'}(\delta) \to \|\rho(Z_{x,n}) - \rho(Z_{y,n})\|_2 < \delta,$$

where $\mathcal{Z}_x = \{Z_{x,n} : n \in \mathbb{N}\}$ and $\mathcal{Z}_y = \{Z_{y,n} : n \in \mathbb{N}\}$ are the convergent sequences in $\Phi(\mathbb{X}_{Q(\mathbb{D}),N})$ and $N^{'}(\delta)$ is defined in Lemma 12.

Lemma 13 *Let $\mathcal{Z}_x = \{Z_{x,n} \in \Phi(\mathbb{X}_{Q(\mathbb{D}),N}) : n \in \mathbb{N}\}$ and $\mathcal{Z}_y = \{Z_{y,n} \in \Phi(\mathbb{X}_{Q(\mathbb{D}),N}) : n \in \mathbb{N}\}$ be the convergent sequences in Lemma 11. For any $\delta > 0$, there exists $N_x'(\delta), N_y'(\delta) \in \mathbb{N}$ such that*

$$\begin{aligned}\forall n > N_x'(\delta) :&\ \|\rho \circ Z_{x,n} - \rho_e \circ \Phi(X)\|_2 < \delta \\ \forall n > N_y'(\delta) :&\ \|\rho \circ Z_{y,n} - \rho_e \circ \Phi(X')\|_2 < \delta.\end{aligned}$$

Let $\|\rho(Z_{x,n}) - \rho(Z_{y,n})\|_2 < \delta$ for all $n > N^{''}(\delta) \stackrel{\text{def}}{=} \max\{N_x'(\delta), N_y'(\delta), N^{'}(\delta)\}$. Then, we have

$$\forall n > N^{''}(\delta) :\ \|\rho_e \circ \Phi(X) - \rho_e \circ \Phi(X')\|_2 < 3\delta.$$

Combining the results of Lemmas 11 to 13 we arrive at the following result:

$$\forall X' \in \mathbb{X}_{\mathbb{D},N} : \|\Phi(X) - \Phi(X')\|_F < \delta \to \|\rho_e \circ \Phi(X) - \rho_e \circ \Phi(X')\|_2 < 3\delta,$$

that is, $\delta(\varepsilon) = \frac{\varepsilon}{3}$ in equation (22), and ρ_e is a continuous function on the compact domain $\Phi(\mathbb{X}_{\mathbb{D},N})$.

I.2.1. PROOF OF LEMMA 11

Let $X, X' \in \mathbb{X}_{\mathbb{D},N}$. Since $\Phi(\mathbb{X}_{Q(\mathbb{D}),N})$ is a dense subset of $\Phi(\mathbb{X}_{\mathbb{D},N})$ (see Lemma 1), there exists sequences $\mathcal{Z}_x \stackrel{\text{def}}{=} \{Z_{x,n} \in \Phi(\mathbb{X}_{Q(\mathbb{D}),N}) : n \in \mathbb{N}\}$ and $\mathcal{Z}_y \stackrel{\text{def}}{=} \{Z_{y,n} \in \Phi(\mathbb{X}_{Q(\mathbb{D}),N}) : n \in \mathbb{N}\}$ such that

$$\lim_{n\to\infty} Z_{x,n} = \Phi(X),\ \lim_{n\to\infty} Z_{y,n} = \Phi(X') \in \Phi(\mathbb{X}_{\mathbb{D},N}),$$

and

$$\rho_e \circ \Phi(X) = \lim_{n\to\infty} \rho(Z_{x,n}),\ \rho_e \circ \Phi(X') = \lim_{n\to\infty} \rho(Z_{y,n}) \in \operatorname{codom}(\rho_e) \subseteq f(\mathbb{X}_{\mathbb{D},N}).$$

That is, there exists $N_x(\delta), N_y(\delta) \in \mathbb{N}$ such that

$$\begin{aligned}\forall n > N_x(\delta) :&\ \|Z_{x,n} - \Phi(X)\|_2 < \delta \\ \forall n > N_y(\delta) :&\ \|Z_{y,n} - \Phi(X')\|_2 < \delta.\end{aligned}$$

for any $\delta > 0$. If $\|\Phi(X) - \Phi(X')\|_2 < \delta$, then for all $n > N(\delta)$, we have

$$\begin{aligned}\|Z_{x,n} - Z_{y,n}\|_2 &\overset{\text{(a)}}{\leq} \|Z_{x,n} - \Phi(X)\|_2 + \|Z_{y,n} - \Phi(X')\|_2 + \|\Phi(X) - \Phi(X')\|_2 \\ &< \delta + \delta + \delta = 3\delta,\end{aligned}$$

where $N(\delta) \overset{\text{def}}{=} \max\{N_x(\delta), N_y(\delta)\}$ and (a) follows from the triangle inequality.

I.2.2. PROOF OF LEMMA 12

The function Φ^{-1} is continuous on its noncopact domain $\Phi(\mathbb{X}_{Q(\mathbb{D}),N})$; see Proposition 11. Therefore, $\rho = f \circ \Phi^{-1}$ is a continuous function on $\Phi(\mathbb{X}_{Q(\mathbb{D}),N})$. By definition of continuity, for all $Z \in \Phi(\mathbb{X}_{Q(\mathbb{D}),N})$ and $u > 0$, there exists a $\gamma(u) > 0$ such that

$$\forall Z' \in \Phi(\mathbb{X}_{Q(\mathbb{D}),N}) : \|Z - Z'\|_F < \gamma(u) \rightarrow \|\rho(Z) - \rho(Z')\|_2 < u. \tag{23}$$

Let $\mathcal{Z}_x = \{Z_{x,n} : n \in \mathbb{N}\}$ and $\mathcal{Z}_y = \{Z_{y,n} : n \in \mathbb{N}\}$ be the convergent sequences in Lemma 11. For all $\delta > 0$, we have

$$\forall X' \in \mathbb{X}_{\mathbb{D},N} : \|\Phi(X) - \Phi(X')\|_F < \delta, n > N(\delta) \rightarrow \|Z_{x,n} - Z_{y,n}\|_2 < 3\delta.$$

For any $\delta > 0$, we let $N'(\delta) = N(\min\{\frac{\delta}{3}, \frac{\gamma(\delta)}{3}\})$ where $N(\delta) \in \mathbb{N}$ is defined in Lemma 11. By definition, we have

$$\forall n > N'(\delta) : \ \|Z_{x,n} - Z_{y,n}\|_2 < \min\{\delta, \gamma(\delta)\} \leq \gamma(\delta).$$

Since ρ is a continuous map, from equation (23), we arrive at the following inequality:

$$\forall n > N'(\delta) : \|\rho(Z_{x,n}) - \rho(Z_{y,n})\|_2 < \delta.$$

I.2.3. PROOF OF LEMMA 13

The sequences $\mathcal{Z}_x = \{Z_{x,n} : n \in \mathbb{N}\}$ and $\mathcal{Z}_y = \{Z_{y,n} : n \in \mathbb{N}\}$ are convergent, that is,

$$\lim_{n\to\infty} Z_{x,n} = \Phi(X), \ \lim_{n\to\infty} Z_{y,n} = \Phi(X') \in \Phi(\mathbb{X}_{\mathbb{D},N}).$$

Since we have $\rho_e \circ \Phi(X) = \lim_{n\to\infty} \rho(Z_{x,b})$, $\rho_e \circ \Phi(X') = \lim_{n\to\infty} \rho(Z_{y,b})$, there exists $N'_x(\delta), N'_y(\delta) \in \mathbb{N}$ such that

$$\begin{aligned}&\forall n > N'_x(\delta) : \ \|\rho \circ Z_{x,n} - \rho_e \circ \Phi(X)\|_2 < \delta \\ &\forall n > N'_y(\delta) : \ \|\rho \circ Z_{y,n} - \rho_e \circ \Phi(X')\|_2 < \delta.\end{aligned}$$

If $\|\rho(Z_{x,n}) - \rho(Z_{y,n})\|_2 < \delta$ for all $n > N''(\delta) \overset{\text{def}}{=} \max\{N'_x(\delta), N'_y(\delta), N'(\delta)\}$, then, from the triangle inequality and Lemma 12, we have

$$\begin{aligned}&\|\rho_e \circ \Phi(X) - \rho_e \circ \Phi(X')\|_2 \\ &\leq \|\rho(Z_{x,n}) - \rho(Z_{y,n})\|_2 + \|\rho(Z_{x,n}) - \rho_e \circ \Phi(X)\|_2 + \|\rho(Z_{y,n}) - \rho_e \circ \Phi(X')\|_2 \\ &< \delta + \delta + \delta = 3\delta.\end{aligned}$$

Appendix J. Proof of Proposition 4

For $K, N \in \mathbb{N}$, let $T, T' \in \mathbb{T}^l_{N,K}$ be such that $S(T) = S(T')$, that is,

$$\{(e_{n_1}^\top l(T), \alpha^1_{n_1}(T)) : n_1 \in [N]\} = \{(e_{n_1}^\top l(T'), \alpha^1_{n_1}(T')) : n_1 \in [N]\},$$

where e_n is the n-th standard basis vector of $\mathbb{R}^N$, for $n \in [N]$. By definition of ℓ-identifiable tensors, all elements of $\{\{e_n^\top l(T) : n \in [N]\}\}$ are unique. Therefore, we we have

$$\forall n_1 \in [N] : \; e_{n_1}^\top l(T) = e_{\pi(n_1)}^\top l(T'), \text{ and } \alpha^1_{n_1}(T) = \alpha^1_{\pi(n_1)}(T'),$$

for a unique permutation operator $\pi : [N] \to [N]$.

Lemma 14 *For all $k \in [K]$, we have*

$$\forall n_1, n_2, \ldots, n_k \in [N] : \alpha^k_{n_1, n_2, \ldots, n_k}(T) = \alpha^k_{\pi(n_1), \pi(n_2), \ldots, \pi(n_k)}(T'),$$

where $\pi : [N] \to [N]$ is a unique permutation operator.

Proof The claim holds for $k = 1$. We prove this statement by induction. Let us assume the claim is true for $k \in [K-1]$. We want to show that it also holds for $k+1$, that is,

$$\forall n_1, n_2, \ldots, n_{k+1} \in [N] : \alpha^{k+1}_{n_1, n_2, \ldots, n_{k+1}}(T) = \alpha^{k+1}_{\pi(n_1), \pi(n_2), \ldots, \pi(n_{k+1})}(T').$$

From the definition of α^k, we have

$$\forall n_1, \ldots, n_{k+1} \in [N] : \; e_{n_{k+1}}^\top l(T) = e_{\pi(n_{k+1})}^\top l(T'), \text{ and } \alpha^{k+1}_{n_1, \ldots, n_{k+1}}(T) = \alpha^{k+1}_{\pi(n_1), \ldots, \pi(n_{k+1})}(T')$$

— which follows from the fact that elements of $\{\{e_n^\top l(T) : n \in [N]\}\}$ are unique. This concludes the proof. ■

From Lemma 14, we have

$$\forall n_1, \ldots, n_K \in [N] : \alpha^{k+1}_{n_1, \ldots, n_K}(T) = \alpha^K_{\pi(n_1), \ldots, \pi(n_K)}(T'),$$

that is, $T = \pi(T')$.

Now let $T, T' \in \mathbb{T}^l_{N,K}$ be such that $T = \pi(T')$ where $\pi : [N] \to [N]$ is a permutation operator. By definition, we have

$$\forall n_1, \ldots, n_K \in [N] : \alpha^K_{n_1, \ldots, n_K}(T) = \alpha^K_{\pi(n_1), \ldots, \pi(n_K)}(T'),$$

and

$$\forall n \in [N] : e_n^\top l(T) = e_n^\top l(\pi(T')) = e_{\pi(n)}^\top l(T').$$

For all $n_1, \ldots, n_{K-1} \in [N]$, we have

$$\begin{aligned}
\alpha^{K-1}_{n_1, \ldots, n_{K-1}}(T) &= \{(e_{n_K}^\top l(T), \alpha^K_{n_1, \ldots, n_K}(T)) : n_K \in [N]\} \\
&= \{(e_{\pi(n_K)}^\top l(T'), \alpha^K_{\pi(n_1), \ldots, \pi(n_K)}(T')) : n_K \in [N]\} \\
&= \{(l_{n_K}(T'), \alpha^K_{\pi(n_1), \ldots, \pi(n_{K-1}), n_K}(T')) : n_K \in [N]\} \\
&= \alpha^{K-1}_{\pi(n_1), \ldots, \pi(n_{K-1})}(T')
\end{aligned}$$

Using a simple argument by induction, we arrive at the statement in Lemma 14. Therefore, we have

$$\begin{aligned}
S(T) &= \{(e_{n_1}^\top l(T), \alpha^1_{n_1}) : n_1 \in [N]\} = \{(e_{\pi(n_1)}^\top l(T'), \alpha^1_{\pi(n_1)}(T')) : n_1 \in [N]\} \\
&= \{(e_{n_1}^\top l(T'), \alpha^1_{n_1}(T')) : n_1 \in [N]\} \\
&= S(T')
\end{aligned}$$

Appendix K. Proof of Theorem 7

Definition 15 *Let $K, N \in \mathbb{N}$. For all $k \in [K]$, let $\mathbb{D}_k$ be a domain and $\phi_k : \mathbb{D}_k \to \operatorname{codom}(\phi_k)$, we define the following multiset function*

$$\Phi_k\big(\{\{x_n \in \mathbb{D}_k : n \in [N]\}\}\big) = \sum_{n \in [N]} \phi_k(x_n),$$

and $\operatorname{codom}(\Phi_k) = \{\sum_{n\in[N]} \phi_k(x_n) : x_n \in \mathbb{D}_k, \forall n \in [N]\}$.

Let us first show that the proposed sum-decomposable model is injective on $\mathbb{T}^l_{N,K}$. Let $K, N \in \mathbb{N}$ and $T, T' \in \mathbb{T}^l_{N,K}$ where

$$\sum_{n_1 \in [N]} \phi_1(e_{n_1}^\top l(T), \beta^1_{n_1}(T)) = \sum_{n_1 \in [N]} \phi_1(e_{n_1}^\top l(T'), \beta^1_{n_1}(T')),$$

that is,

$$\Phi_1(\{\{(e_{n_1}^\top l(T), \beta^1_{n_1}(T)) : n_1 \in [N]\}\}) = \Phi_1(\{\{(e_{n_1}^\top l(T'), \beta^1_{n_1}(T')) : n_1 \in [N]\}\})$$

Let us **assume** that ϕ_1 is such that the corresponding Φ_1 is an injective multiset function (see Definition 15) — we shall discuss its sufficient condition later in the proof. Since $\{\{e_n^\top l(T) : n \in [N]\}\}$ has all distinct elements for all $T \in \mathbb{T}^l_{N,K}$, we have $e_{n_1}^\top(T) = e_{\pi(n_1)}^\top l(T')$ for a unique permutation operator $\pi : [N] \to [N]$ and for all $n_1 \in [N]$. Therefore, we have

$$\forall n_1 \in [N] : \beta^1_{n_1}(T) = \beta^1_{\pi(n_1)}(T')$$

Lemma 15 *Let $k \in [K]$ and* **assume** $\{\Phi_k : k \in [K]\}$ *are injective multiset functions over their domains, that is,*

$$\forall k \in [K] : \phi_k : \mathbb{D}_k \to \operatorname{codom}(\phi_k)$$

where $\mathbb{D}_k = \operatorname{codom}(l) \times \operatorname{codom}(\Phi_{k+1})$ and $\mathbb{D}_K = \operatorname{codom}(l) \times \mathbb{R}^D$. Then, for all $n_1, \ldots, n_k \in [N]$, we have $\beta^k_{n_1,\ldots,n_k}(T) = \beta^k_{\pi(n_1),\ldots,\pi(n_k)}(T')$ where $\pi : [N] \to [N]$ is a unique permutation operator and $k \in [K]$.

Proof The claim holds for $k = 1$. We prove this statement by induction. Let us assume this claim is true for $k \in [K-1]$. We want to show that it also holds for $k+1$, that is,

$$\forall n_1, n_2, \ldots, n_{k+1} \in [N] : \beta^{k+1}_{n_1,n_2,\ldots,n_{k+1}}(T) = \beta^{k+1}_{\pi(n_1),\pi(n_2),\ldots,\pi(n_{k+1})}(T').$$

From the definition of β^k, for all $n_1, n_2, \ldots, n_k \in [N]$, we have

$$\sum_{n_{k+1} \in [N]} \phi_{k+1}(e_{n_{k+1}}^\top l(T), \beta^{k+1}_{n_1\ldots n_{k+1}}(T)) = \sum_{n_{k+1} \in [N]} \phi_{k+1}(e_{n_{k+1}}^\top l(T'), \beta^{k+1}_{n_1\ldots n_{k+1}}(T')),$$

that is,

$$\begin{aligned}&\Phi_{k+1}(\{\{(e_{n_{k+1}}^\top l(T), \beta^{k+1}_{n_1\ldots n_{k+1}}(T)) : n_{k+1} \in [N]\}\}) \\ &= \Phi_{k+1}(\{\{(e_{n_{k+1}}^\top l(T'), \beta^{k+1}_{n_1\ldots n_{k+1}}(T')) : n_{k+1} \in [N]\}\}).\end{aligned}$$

for all $n_1, n_2, \ldots, n_k \in [N]$. Since Φ_k is an injective multiset function, we have

$$\forall n_1, \ldots, n_{k+1} \in [N]: \; e_{n_{k+1}}^\top l(T) = e_{\pi(n_{k+1})}^\top l(T'), \text{ and } \beta^{k+1}_{n_1,\ldots,n_{k+1}}(T) = \beta^{k+1}_{\pi(n_1),\ldots,\pi(n_{k+1})}(T')$$

— which follows from the fact that elements of $\{\{e_n^\top l(T) : n \in [N]\}\}$ are unique. This concludes the proof. ∎

From Lemma 15, we arrive at

$$\forall n_1, \ldots, n_K \in [N] : \beta^K_{n_1,n_2,\ldots,n_K}(T) = \beta^K_{\pi(n_1),\pi(n_2),\ldots,\pi(n_K)}(T'),$$

for a unique permutation operator $\pi : [N] \to [N]$, that is, $T = \pi(T')$ and $S(T) = S(T')$; see Proposition 4.

Using induction, one can easily verify that given $S(T)$, we can compute

$$\sum_{n_1 \in [N]} \phi_1(e_{n_1}^\top l(T), \beta^1_{n_1}(T)).$$

Therefore, the following function is well-defined and injective:

$$\forall T \in \mathbb{T}^l_{N,K} : m \circ S(T) = \sum_{n_1 \in [N]} \phi_1(l_{n_1}(T), \beta^1_{n_1}(T)),$$

that is, if $m \circ S(T) = m \circ S(T')$ then we have $S(T) = S(T')$ where $T, T' \in \mathbb{T}^l_{N,K}$.

Now we define the function $f_s : S(\mathbb{T}^l_{N,K}) \to \operatorname{codom}(f)$ as follows:

$$\forall T \in \mathbb{T}^l_{N,K} : f_s \circ S(T) \stackrel{\text{def}}{=} f(T).$$

Since f is a permutation-invariant, the function f_s is well-defined, that is, $f_s \circ S(T) = f(T) = f(\pi(T)) = f_s \circ S(\pi(T)) = f_s \circ S(T)$ for any permutation operator $\pi : [N] \to [N]$. Since m is an injective function over its domain, it is invertible on it. Now we define the following function:

$$\forall u \in m \circ S(\mathbb{T}^l_{N,K}) : \rho(u) \stackrel{\text{def}}{=} f_s \circ m^{-1}(u).$$

For any $u \in m \circ S(\mathbb{T}^l_{N,K})$, we have $u = \sum_{n_1 \in [N]} \phi_1(e_{n_1}^\top l(T), \beta^1_{n_1}(T))$ where $T \in \mathbb{T}^l_{N,K}$, that is,

$$\forall T \in \mathbb{T}^l_{N,K} : \rho(u) = \rho\big(\sum_{n_1 \in [N]} \phi_1(e_{n_1}^\top l(T), \beta^1_{n_1}(T))\big) = f_s \circ m^{-1} \circ m \circ S(T) = f(T).$$

Sufficient conditions for injective multiset functions $\{\Phi_k : k \in [K]\}$.

(1) If $l(T) \in \mathbb{R}^{N \times M}$, then we use the result in Theorem 8 to ensure the injectivity of Φ_k, for all $k \in [K]$. The function ϕ_k is defined on domain $\mathbb{D}_k = \operatorname{codom}(l) \times \operatorname{codom}(\Phi_{k+1})$ where $\mathbb{D}_K = \operatorname{codom}(l) \times \mathbb{R}^D$, $\operatorname{codom}(l) \subset \mathbb{R}^M$, and $\operatorname{codom}(\Phi_{k+1}) \subset \mathbb{R}^{D_{k+1}}$. From Theorem 8, $D_k = \binom{N+D_{k+1}}{N} - 1$ ensures the injectivity of Φ_k, for all $k \in [K]$.

(2) If $l(T) \in \mathbb{Q}^{N \times M}$, then we use the result in Theorem 5 to ensure the injectivity of Φ_k, for all $k \in [K]$. This is due the fact that rational-valued vectors are identifiable (see Proposition 3). From Theorem 5, $D_k = 2N(M + D_{k+1})$ ensures the injectivity of Φ_k, for all $k \in [K]$.

Appendix L. Supplementary Discussion

As discussed in the main text, the imporant step in showing the existence of sum-decomposable representation is proving that the multiset encoding function $\Phi : \mathrm{dom}(\Phi) \to \mathrm{codom}(\Phi)$ is an injective map, that is, $\rho = f \circ \Phi^{-1}$ is well-defined over its admissible inputs, that is, $\mathrm{codom}(\Phi)$.

Proposition 12 (Fereydounian et al. 2022) *Consider the following continuous map* [1]*:*

$$\forall x \in \mathbb{R}^{\mathbb{D}}, d_1, d_2 \in [D], n \in [N] : \big(\phi(x)\big)_{d_1,d_2,n} = \begin{cases} \mathrm{Re}\{(x_{d_1} + x_{d_2}\sqrt{-1})^n\} & \textit{if } d_2 > d_1 \\ \mathrm{Im}\{(x_{d_1} + x_{d_2}\sqrt{-1})^n\} & \textit{if } d_1 > d_2 \\ 0 & \textit{otherwise}. \end{cases}$$

The map $\phi : \mathbb{R}^D \to \mathrm{codom}(\phi) \subset \mathbb{R}^{D \times D \times N}$ defines the following injective multiset function:

$$\forall X \in \mathbb{X}_{\mathbb{R}^D,N} : \Phi(X) = \sum_{x \in X} \phi(x).$$

We argue that the result in Proposition 12 is not valid for all multisets, as the following example suggests. Consider the following distinct sets:

$$X = \{\{\begin{bmatrix}1\\1\\1\end{bmatrix}, \begin{bmatrix}3\\2\\1\end{bmatrix}, \begin{bmatrix}1\\2\\2\end{bmatrix}, \begin{bmatrix}3\\1\\2\end{bmatrix}\}\},\ X' = \{\{\begin{bmatrix}1\\2\\1\end{bmatrix}, \begin{bmatrix}3\\1\\1\end{bmatrix}, \begin{bmatrix}3\\2\\2\end{bmatrix}, \begin{bmatrix}1\\1\\2\end{bmatrix}\}\}.$$

One can be readily verify that $\Phi(X) = \Phi(X')$, where Φ is defined in Proposition 12. The main insight behind this example is the fact that $\{\{(e_d^\top x_n, e_{d'}^\top x_n) : n \in [N]\}\} = \{\{(e_d^\top x_n, e_{d'}^\top x'_n) : n \in [N]\}\}$ for all distinct $d, d' \in [D]$, where $X = \{\{x_n : n \in [N]\}\}$, $X' = \{\{x'_n : n \in [N]\}\}$, $D = 3$, and $N = 4$. This later equality does indeed show $X = X'$ **if** both multisets contains distinct vectors with **distinct** elements, namely, sets with distinct vectors. The key elements in proving Theorem 8 is to construct an injective Φ which guarantees the existence of $\rho = f \circ \Phi^{-1}$. Even assuming input multisets contain vectors with distinct elements, the above result doe not guarantee the continuity of ρ. Furthermore, one can easily show that $\mathrm{codom}(\Phi)$ is not a compact set. To show this, note that the domain of Φ does not include a single point X in the example above. Now, one can construct a sequence of (multi)ests X_n where $\lim_{n\to\infty} X_n = X$ such that, for all $n \in \mathbb{N}$, all elements of X_n are distinct and have distinct values, that is, $X_n \in \mathrm{dom}(\Phi)$. Since Φ is a continuous map, $\{\Phi(X_n) : n \in \mathbb{N}\}$ is a Cauchy sequence in $\mathrm{codom}(\Phi)$ whose limit does not belong to $\mathrm{codom}(\Phi)$, that is, the co domain of Φ is not compact.

1. This is a trivially altered version of the function in (Fereydounian et al., 2022).