



G-invariant representations using coorbits: bi-Lipschitz properties

Radu Balan¹ · Efstratios Tsoukanis² 

Received: 23 September 2025 / Accepted: 30 August 2026
© The Author(s) 2026

Abstract

Let $\mathcal{V}$ be a finite-dimensional real vector space with a finite group G acting orthogonally on it. We address the general problem of constructing Euclidean-stable embeddings of the orbit space $\mathcal{V}/G$. Our approach builds on embeddings derived from subsets of sorted coorbits. The main result establishes that whenever such embeddings are injective, they are automatically bi-Lipschitz. Furthermore, we show that stable embeddings can be achieved in reduced dimensions, and that any continuous or Lipschitz G -invariant map can be factored through these embeddings.

Keywords G-invariant representations · Bi-Lipschitz embeddings · Sorted coorbits · Orbit space · Max filtering · Invariant machine learning

Mathematics Subject Classification (2020) Primary 51F30 · Secondary 46B85 · 42C15 · 20C35 · 68T05 · 94A12

1 Introduction

In many machine learning applications, the goal is to embed data into a Euclidean space $\mathbb{R}^m$ via a *symmetry-invariant embedding* Ψ , which is then used as a feature space. Ideally, this embedding should not only be *invariant under a group action* but should also *separate orbits* and satisfy appropriate *stability conditions*, ensuring that small perturbations in the input lead to only minor changes in the output or prediction.

This setting falls within the broader framework of *invariant machine learning* [4, 13, 15, 17, 18, 20, 26, 28].

Communicated by: Alex Townsend

✉ Efstratios Tsoukanis
efstratios.tsoukanis@cgu.edu

Radu Balan
rvbalan@umd.edu

¹ Department of Mathematics, University of Maryland, College Park, MD, USA

² Institute of Mathematical Sciences, Claremont Graduate University, Claremont, CA, USA

The most common group actions studied in this context include:

- Permutations [8, 16, 31],
- Reflections [27], and
- Translations [14].

A closely related problem is that of *equivariant embeddings*, where the objective is to construct mappings that intertwine group actions on the domain and codomain [12, 26, 29, 35, 36].

An important example of an invariant learning problem is the *phase retrieval problem*, originally introduced in [7], where the symmetry group is the 1-torus (a compact abelian group). This problem has been extensively studied in both real and complex settings [1, 2, 5, 6, 22].

Our work builds on and unifies two prior approaches:

- The *max filter embedding* developed in [16, 28], and
- The *permutation-invariant representation* introduced in [8].

A related analysis appears in [3], where the stability of generalized phase retrieval problems is studied under compact group actions, and conditions for stable recovery from invariant measurements are provided.

There is a deep connection between permutation-invariant embeddings and the phase retrieval problem. For example, [6] provides sharp bounds for both the upper and lower Lipschitz constants in real and complex settings, while [10] shows an isometric equivalence between the real-valued phase retrieval problem and representations invariant under the symmetric group Σ_2 .

A different line of work, represented by [21], proves that for any discrete group G of isometries acting on $\mathbb{R}^d$, there exists a *bi-Lipschitz* embedding of the quotient space $\mathbb{R}^d/G$ into $\mathbb{R}^N$, where the distortion depends only on the dimension d .

Since the initial version of this paper was made publicly available on arXiv, [30] has extended our construction. They demonstrate that for certain compact groups, and given a sufficient number of generic templates, the *coorbit filter bank* (as defined in their work) yields injective and bi-Lipschitz embeddings.

In this paper, we prove that the sorted coorbit embeddings constructed in [9] are *globally bi-Lipschitz* provided they are injective.

Organization of the Paper:

- Section 1 introduces the embedding map and the overall construction.
- Section 2 analyzes the upper Lipschitz bound and establishes the equivalence between injectivity on the quotient space and stability.
- Section 3 shows that a generic linear projection can reduce the target dimension while preserving both injectivity and stability.
- Section 4 presents universal factorization results for arbitrary continuous or Lipschitz G -invariant functions.

1.1 Notation

The *sorted coorbit embedding* was introduced in [9]. In this subsection, we review its construction. Let $(\mathcal{V}, \langle \cdot, \cdot \rangle)$ be a real inner product space of dimension $d \geq 2$, and let

G be a finite group of order $|G| = N$ acting orthogonally on $\mathcal{V}$. For each $g \in G$, we denote the group action on a vector $x \in \mathcal{V}$ by $U_g x$.

Define the *orbit space* $\hat{\mathcal{V}} = \mathcal{V} / \sim$, where two vectors $x, y \in \mathcal{V}$ are equivalent ($x \sim y$) if there exists $g \in G$ such that $y = U_g x$. The *orbit* of a vector x is denoted by $[x] = \{U_g x \mid g \in G\}$.

We endow $\hat{\mathcal{V}}$ with the following *natural metric* $\mathbf{d} : \hat{\mathcal{V}} \times \hat{\mathcal{V}} \rightarrow \mathbb{R}$:

$$\mathbf{d}([x], [y]) = \min_{h_1, h_2 \in G} \|U_{h_1} x - U_{h_2} y\| = \min_{g \in G} \|x - U_g y\|. \quad (1)$$

With this metric, $(\hat{\mathcal{V}}, \mathbf{d})$ becomes a complete metric space.

Our objective is to construct a bi-Lipschitz Euclidean embedding of the metric space $(\hat{\mathcal{V}}, \mathbf{d})$ into a Euclidean space $\mathbb{R}^m$. Specifically, we aim to construct a function $\Psi : \mathcal{V} \rightarrow \mathbb{R}^m$ such that

1. Invariance: $\Psi(U_g x) = \Psi(x)$, $\forall x \in \mathcal{V}$, $\forall g \in G$,
2. Orbit separation: If $x, y \in \mathcal{V}$ are such that $\Psi(x) = \Psi(y)$, then there exists $g \in G$ such that $y = U_g x$,
3. Stability: there exist constants $0 < a \leq b < \infty$ such that for all $x, y \in \mathcal{V}$

$$a \mathbf{d}([x], [y])^2 \leq \|\Psi(x) - \Psi(y)\|^2 \leq b \mathbf{d}([x], [y])^2.$$

The *invariance* property (Item 1 above) allows us to construct a well-defined function on the quotient space $\hat{\mathcal{V}} = \mathcal{V} / \sim$ via

$$\hat{\Psi} : \hat{\mathcal{V}} \rightarrow \mathbb{R}^m, \quad \hat{\Psi}([x]) = \Psi(x), \quad \forall [x] \in \hat{\mathcal{V}}.$$

If the G -invariant map Ψ satisfies the orbit separation property (Item 2 above), we say that it *separates G -orbits* in $\mathcal{V}$.

Our construction of the embedding Ψ is based on a non-linear sorting operator $\downarrow$, defined as follows.

Notation 1.1 Let $\downarrow : \mathbb{R}^r \rightarrow \mathbb{R}^r$ be the map that takes a vector and returns its coordinates sorted in *non-increasing order*:

$$x \in \mathbb{R}^r \mapsto \downarrow x = (x_{\sigma(i)})_{1 \leq i \leq r}, \quad x_{\sigma(1)} \geq \cdots \geq x_{\sigma(r)}$$

for some permutation $\sigma : \{1, \dots, r\} \rightarrow \{1, \dots, r\}$.

For an integer $p \in \mathbb{N}$, let $[p] = \{1, 2, \dots, p\}$. For a set S , $|S|$ denotes its cardinality. Fix a p -tuple of vectors $\mathbf{w} = (w_1, \dots, w_p) \in \mathcal{V}^p$. For each $i \in [p]$ and $j \in [N]$, define the scalar-valued function $\Phi_{w_i, j} : \mathcal{V} \rightarrow \mathbb{R}$ as follows:

$$\Phi_{w_i, j}(x) = j\text{-th coordinate of } \downarrow(\langle U_g w_i, x \rangle)_{g \in G}.$$

Next, let $S \subset [N] \times [p]$ be an index set of size $|S| = m$, and for each $i \in [p]$, define $S_i = \{k \in [N] \mid (k, i) \in S\}$ (the i -th column of S). Let $m_i = |S_i|$. Then, $m = \sum_{i=1}^p m_i$.

Notation 1.2 The *sorted coorbit embedding* $\Phi_{\mathbf{w},S}$ associated with the windows $\mathbf{w} \in \mathcal{V}^p$ and index set $S \subset [N] \times [p]$ is given by

$$\Phi_{\mathbf{w},S} : \mathcal{V} \rightarrow \mathbb{R}^m, \quad \Phi_{\mathbf{w},S}(x) = [\{\Phi_{w_1,j}(x)\}_{j \in S_1}, \dots, \{\Phi_{w_p,j}(x)\}_{j \in S_p}] \in \mathbb{R}^m. \quad (2)$$

Let $\ell : \mathbb{R}^m \rightarrow \mathbb{R}^q$ be a linear map.

Notation 1.3 The embedding $\Psi_{\mathbf{w},S,\ell}$ associated with the windows $\mathbf{w} \in \mathcal{V}^p$, index set $S \subset [N] \times [p]$ and linear map $\ell : \mathbb{R}^m \rightarrow \mathbb{R}^q$ is given by

$$\Psi_{\mathbf{w},S,\ell} = \ell \circ \Phi_{\mathbf{w},S} : \mathcal{V} \rightarrow \mathbb{R}^q, \quad \Psi_{\mathbf{w},S,\ell}(x) = \ell(\Phi_{\mathbf{w},S}(x)). \quad (3)$$

In this paper, we focus on stability and universality properties of the maps $\Phi_{\mathbf{w},S}$ and $\Psi_{\mathbf{w},S,\ell}$.

Informally, our main results (Theorems 1.5, 1.7, 1.8 and 1.9) can be summarized as follows:

1. **Injectivity implies (bi-Lipschitz) stability:** If the embedding separates orbits, then it is automatically bi-Lipschitz with respect to the quotient space metric.
2. **Dimension efficiency:** A stable, bi-Lipschitz embedding can be achieved in a Euclidean space of dimension at most twice that of the input space.
3. **Universality:** Any continuous or Lipschitz G -invariant map can be factored through the coorbit embedding $\Phi_{\mathbf{w},S}$.

Remark 1.4 Figure 1 provides a schematic illustration of the proposed embedding scheme.

For the rest of the paper, we will use $\Phi_{i,j}$ as shorthand for $\Phi_{w_i,j}$, and the two notations will be used interchangeably.

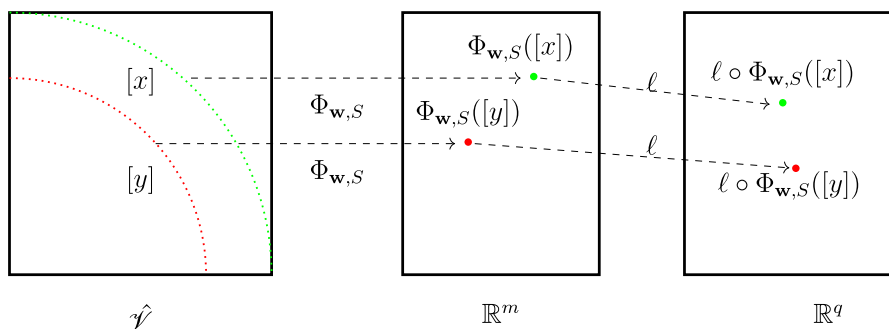


Fig. 1 Proposed embedding scheme

1.2 Main results

In this subsection, we summarize the key results of this paper, which concern bi-Lipschitz stability of embeddings, dimensionality reduction, and universality of G -invariant maps.

Theorem 1.5 *Let G be a finite group acting orthogonally on the vector space $\mathcal{V}$. For fixed $\mathbf{w} \in \mathcal{V}^p$ and $S \subset [N] \times [p]$, where $|S| = m$, suppose that the map $\hat{\Phi}_{\mathbf{w},S} : \hat{\mathcal{V}} \rightarrow \mathbb{R}^m$ is injective on the quotient space $\hat{\mathcal{V}} = \mathcal{V}/G$. Then, there exist $0 < a \leq b < \infty$ such that for all $x, y \in \mathcal{V}$,*

$$a \mathbf{d}([x], [y]) \leq \|\Phi_{\mathbf{w},S}(x) - \Phi_{\mathbf{w},S}(y)\|_2 \leq b \mathbf{d}([x], [y]).$$

Corollary 1.6 *When injective, the max filter bank embedding is bi-Lipschitz. Specifically: Let G be a finite group acting orthogonally on $\mathcal{V}$. Let $S_{\max} = \{(1, i) \mid i \in [p]\} \subset [N] \times [p]$ be the index set that selects, for each window w_i , the maximal (first sorted) coorbit coordinate. Suppose the max filter bank embedding $\hat{\Phi}_{\mathbf{w},S_{\max}} : \hat{\mathcal{V}} \rightarrow \mathbb{R}^p$, defined by*

$$\hat{\Phi}_{\mathbf{w},S_{\max}}([x]) = \left(\max_{g \in G} \langle U_g w_i, x \rangle \right)_{i \in [p]},$$

is injective. Then, $\hat{\Phi}_{\mathbf{w},S_{\max}}$ is bi-Lipschitz, that is, there are $0 < a \leq b < \infty$ such that for every $x, y \in \mathcal{V}$,

$$a \mathbf{d}([x], [y]) \leq \|\Phi_{\mathbf{w},S_{\max}}(x) - \Phi_{\mathbf{w},S_{\max}}(y)\|_2 \leq b \mathbf{d}([x], [y]).$$

Theorem 1.7 *Let $\{U_g \mid g \in G\}$ denote a representation of a finite group G of order N acting by isometries on the real vector space $\mathcal{V}$ of dimension d . Let $V_G = \{x \in \mathcal{V} \mid U_g x = x \ \forall g \in G\}$ denote the linear space of vectors invariant under this representation, and let $d_G = \dim(V_G)$ denote its dimension. Let $\mathbf{w} \in \mathcal{V}^p$ and $S \subset [N] \times [p]$ so that $\hat{\Phi}_{\mathbf{w},S} : \hat{\mathcal{V}} \rightarrow \mathbb{R}^m$ is injective on the quotient space $\hat{\mathcal{V}}$. Then, for a generic linear map $\ell : \mathbb{R}^m \rightarrow \mathbb{R}^{2d-d_G}$, the map $\hat{\Psi}_{\mathbf{w},S,\ell} = \ell \circ \hat{\Phi}_{\mathbf{w},S} : \hat{\mathcal{V}} \rightarrow \mathbb{R}^{2d-d_G}$ is injective and bi-Lipschitz. Here, “generic” refers to a property that holds on an open dense subset in the Zariski topology over the space of matrices.*

Theorem 1.8 *Assume that for some $S \subset [N] \times [p]$ and $\mathbf{w} = (w_1, \dots, w_p) \in \mathcal{V}^p$, the map $\hat{\Phi}_{\mathbf{w},S} : \hat{\mathcal{V}} \rightarrow \mathbb{R}^m$ is bi-Lipschitz on the quotient space $\hat{\mathcal{V}} = \mathcal{V}/G$, with (upper) Lipschitz constant $b = \text{Lip}(\hat{\Phi}_{\mathbf{w},S})$ and lower Lipschitz constant $a > 0$.*

1. *For every Lipschitz map $F : \hat{\mathcal{V}} \rightarrow H$ into a real Hilbert space H , there exists a Lipschitz map $T : \mathbb{R}^m \rightarrow H$ such that $F([x]) = T(\Phi_{\mathbf{w},S}(x))$ for all $x \in \mathcal{V}$, with Lipschitz constant $\text{Lip}(T) \leq \frac{1}{a} \text{Lip}(F)$.*
2. *Conversely, for any Lipschitz map $T : \mathbb{R}^m \rightarrow H$, the map $F : \hat{\mathcal{V}} \rightarrow H$, $F([x]) = T(\Phi_{\mathbf{w},S}(x))$, is well defined (as $T \circ \Phi_{\mathbf{w},S}$ is G -invariant) and Lipschitz, with Lipschitz constant $\text{Lip}(F) \leq b \text{Lip}(T)$.*

Theorem 1.9 Assume that, for fixed $S \subset [N] \times [p]$ and $\mathbf{w} = (w_1, \dots, w_p) \in \mathcal{V}^p$, the map $\hat{\Phi}_{\mathbf{w},S} : \hat{\mathcal{V}} \rightarrow \mathbb{R}^m$ is bi-Lipschitz on the quotient space $\hat{\mathcal{V}} = \mathcal{V}/G$. Let L denote a locally convex topological vector space. For every continuous map $F : \hat{\mathcal{V}} \rightarrow L$, there exists a continuous map $T : \mathbb{R}^m \rightarrow L$ such that $F([x]) = T(\Phi_{\mathbf{w},S}(x))$ for all $x \in \mathcal{V}$. Furthermore, $T(\mathbb{R}^m)$ is contained in the convex hull of $F(\hat{\mathcal{V}})$.

Remark 1.10 In [11], we generalize Theorem 1.5 to the case where the embedding has the form $A \circ \Phi_{\mathbf{w},S}$, where A is a linear map.

Remark 1.11 In [9], we study under which conditions the map $\hat{\Phi}_{\mathbf{w},S} : \hat{\mathcal{V}} \rightarrow \mathbb{R}^m$ is injective on the quotient space $\hat{\mathcal{V}} = \mathcal{V}/G$.

2 Stability of the embedding

Suppose that for $\mathbf{w} = (w_1, \dots, w_p) \in \mathcal{V}^p$ and $S \subset [N] \times [p]$ the map $\hat{\Phi}_{\mathbf{w},S}$ is injective. In this case, we claim the map $\hat{\Phi}_{\mathbf{w},S}$ is also bi-Lipschitz (Theorem 1.5). An estimate of the upper Lipschitz constant b is given in Lemma 2.1.

Corollary 1.6 follows from Theorem 1.5.

The proof of Theorem 1.5 is contained in the next two subsections.

2.1 Upper Lipschitz bound

Note that $\Phi_{\mathbf{w},S}$ is Lipschitz because it is a composition of Lipschitz maps. It remains to estimate the Lipschitz constant.

Lemma 2.1 (Upper Lipschitz bound) Consider a finite group G of size N acting orthogonally on $\mathcal{V}$. Let $\mathbf{w} \in \mathcal{V}^p$ and $S \subset [N] \times [p]$. Denote

$$B = \max_{\substack{\sigma_1, \dots, \sigma_p \subset G \\ |\sigma_i| = m_i, \forall i}} \lambda_{\max} \left(\sum_{i=1}^p \sum_{g \in \sigma_i} U_g w_i w_i^T U_g^T \right),$$

where $S_i = \{k \in [N] \mid (k, i) \in S\}$ and $m_i = |S_i|$. Then, $\hat{\Phi}_{\mathbf{w},S} : (\hat{\mathcal{V}}, \mathbf{d}) \rightarrow \mathbb{R}^m$ is Lipschitz with upper Lipschitz constant $\sqrt{B}$.

Proof For fixed $x, y \in \mathcal{V}$, $i \in [p]$, $j \in S_i$, let $\psi_{i,j} : [0, 1] \rightarrow \mathbb{R}$, where

$$\psi_{i,j}(t) = \Phi_{i,j}((1-t)x + ty) = \langle (1-t)x + ty, U_{g(t)} w_i \rangle.$$

Here $g(t)$ denotes a group element that achieves the j -th coordinate of the sorted coorbit of the vector $(1-t)x + ty$, that is, $\Phi_{i,j}((1-t)x + ty) = \langle (1-t)x + ty, U_{g(t)} w_i \rangle$.

The function $\psi_{i,j}$ is Lipschitz, being the composition of the $\|w_i\|$ -Lipschitz map $\Phi_{i,j}$ with an affine path; hence it is absolutely continuous and differentiable almost

everywhere. Notice that

$$\frac{d}{dt}\psi_{i,j}(t) = \langle y - x, U_{g(t)}w_i \rangle$$

for almost every t . Specifically, $\psi_{i,j}(t)$ is differentiable at all $t \in [0, 1]$ such that there exists $\epsilon > 0$ so that $g|_{(t-\epsilon, t+\epsilon)}$ can be chosen to be constant. Such a choice is possible for all but finitely many t : the N functions $t \mapsto \langle (1-t)x + ty, U_g w_i \rangle$ are affine, so two of them either coincide identically or cross at a single point, and away from the finitely many crossing points a single sorting permutation of the coorbit remains valid on a neighborhood.

By the fundamental theorem of calculus, we get

$$\Phi_{i,j}(y) - \Phi_{i,j}(x) = \int_0^1 \frac{d}{dt} \Phi_{i,j}((1-t)x + ty) dt.$$

Therefore, writing $g_{j,t}$ for the element $g(t)$ chosen above, to make its dependence on j explicit (the dependence on i is kept implicit),

$$\Phi_{i,j}(y) - \Phi_{i,j}(x) = \int_0^1 \langle y - x, U_{g_{j,t}}w_i \rangle dt.$$

Note that, for each fixed t , the group elements $g_{j,t}$, $j \in S_i$, can be chosen pairwise distinct (choose a sorting permutation of the coorbit), so that $\sigma_i(t) := \{g_{j,t} \mid j \in S_i\}$ satisfies $|\sigma_i(t)| = m_i$. By Minkowski's integral inequality, applied jointly over all coordinates (i, j) with $i \in [p]$, $j \in S_i$,

$$\begin{aligned} \|\Phi_{\mathbf{w},S}(x) - \Phi_{\mathbf{w},S}(y)\| &= \left(\sum_{i=1}^p \sum_{j \in S_i} \left| \int_0^1 \langle y - x, U_{g_{j,t}}w_i \rangle dt \right|^2 \right)^{1/2} \\ &\leq \int_0^1 \left(\sum_{i=1}^p \sum_{j \in S_i} \langle y - x, U_{g_{j,t}}w_i \rangle^2 \right)^{1/2} dt \leq \sqrt{B} \|x - y\|, \end{aligned}$$

where the last inequality follows from

$$\sum_{i=1}^p \sum_{j \in S_i} \langle y - x, U_{g_{j,t}}w_i \rangle^2 = (y - x)^T \left(\sum_{i=1}^p \sum_{g \in \sigma_i(t)} U_g w_i w_i^T U_g^T \right) (y - x) \leq B \|y - x\|^2.$$

Hence, choosing representatives $x \in [x]$, $y \in [y]$ so that $\mathbf{d}([x], [y]) = \|x - y\|$,

$$\|\Phi_{\mathbf{w},S}(x) - \Phi_{\mathbf{w},S}(y)\|^2 \leq B \|x - y\|^2 = B \mathbf{d}([x], [y])^2.$$

□

2.2 Lower Lipschitz bound

We prove the lower bound by contradiction. Assuming the lower Lipschitz constant vanishes, we produce a pair of sequences whose $\Phi_{w,S}$ -images collapse relative to their quotient distance, concentrate them near a single unit vector z_1 (Lemma 2.7), and then, one direction at a time, extend this single direction to an orthogonal family $z_1, \dots, z_d$ along which separation keeps failing, the “robotic arm” of Fig. 2. Once the family spans $\mathcal{V}$, the vanishing of the lower Lipschitz constant forces a unit vector $\tilde{z}$ to be orthogonal to all relevant coorbit filters. A small translation of $\sum_{r=1}^d z_r$ along $\tilde{z}$ then yields two distinct points with the same embedding, contradicting injectivity.

Each step of this construction perturbs a partial sum $\sum_r z_r$ by a small amount, so we first record how the sorted coorbit coordinates react to such perturbations. Two quantities attached to each coordinate (i, j) govern this: the set $L^{i,j}(x)$ of group elements that realize the j -th sorted value, and the gap $\Delta^{i,j}(x)$ separating that value from the others. Lemmas 2.2–2.5 describe how $L^{i,j}$ and $\Delta^{i,j}$ behave under rescaling and under small additive perturbations, and these are the invariances the induction uses.

2.2.1 Geometric analysis of coorbits

First, let us introduce some additional notation. For fixed $i \in [p]$, $j \in [N]$ and $x \in \mathcal{V}$, we define the following non-empty subset of the group G :

$$L^{i,j}(x) = \{g \in G \mid \langle x, U_g w_i \rangle = \Phi_{i,j}(x)\}. \quad (4)$$

This represents the collection of group elements that achieve the j -th position of the sorted coorbit $\downarrow (\langle U_g w_i, x \rangle)_{g \in G}$.

Consider also the map

$$\Delta^{i,j}(x) = \begin{cases} \min_{g \notin L^{i,j}(x)} (|\langle U_g w_i, x \rangle - \Phi_{i,j}(x)|) \frac{1}{\|w_i\|}, & \text{if } L^{i,j}(x) \neq G \\ \frac{\|x\|}{\|w_i\|}, & \text{if } L^{i,j}(x) = G. \end{cases} \quad (5)$$

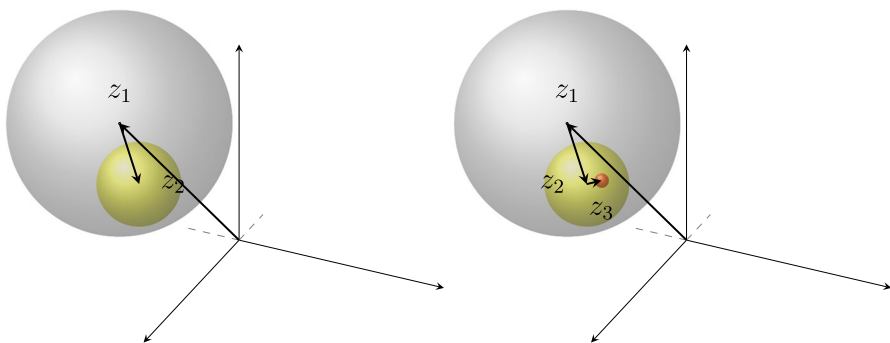


Fig. 2 The robotic arm method

Here and in what follows, we assume that $w_i \neq 0$ for every $i \in [p]$. This loses no generality: a window $w_i = 0$ contributes constant (zero) coordinates to $\Phi_{\mathbf{w},S}$, which affect neither injectivity on the quotient space nor any Lipschitz bound, so such windows can be discarded.

Lemma 2.2 *For any $x \in \mathcal{V}$, $i \in [p]$, and $j \in [N]$,*

$$|\{g \in G \mid \langle U_g w_i, x \rangle > \Phi_{i,j}(x)\}| \leq j - 1 \quad (6)$$

and

$$|\{g \in G \mid \langle U_g w_i, x \rangle < \Phi_{i,j}(x)\}| \leq N - j. \quad (7)$$

Proof By definition $\Phi_{i,j}(x)$ is the j -th largest among the values $\{\langle U_g w_i, x \rangle\}_{g \in G}$. Hence at most $j - 1$ of them are strictly larger than $\Phi_{i,j}(x)$, which is Equation (6); and at least j of them are $\geq \Phi_{i,j}(x)$ (the top j sorted entries), so at most $N - j$ are strictly smaller, which is Equation (7). $\square$

Note that for any $w_1, \dots, w_p \in \mathcal{V} \setminus \{0\}$ the subset $L^{i,j}(x) \subset G$ has the following “nestedness” property.

Lemma 2.3 (Nestedness of active sets) *For any $x, y \in \mathcal{V}$ such that $\|y\| < \frac{1}{2}\Delta^{i,j}(x)$, we have that $L^{i,j}(x + y) \subset L^{i,j}(x)$. Furthermore,*

$$\{g \in G \mid \langle U_g w_i, x \rangle > \Phi_{i,j}(x)\} \subset \{g \in G \mid \langle U_g w_i, x + y \rangle > \Phi_{i,j}(x + y)\},$$

$$\{g \in G \mid \langle U_g w_i, x \rangle < \Phi_{i,j}(x)\} \subset \{g \in G \mid \langle U_g w_i, x + y \rangle < \Phi_{i,j}(x + y)\},$$

$$\{g \in G \mid \langle U_g w_i, x + y \rangle \geq \Phi_{i,j}(x + y)\} \subset \{g \in G \mid \langle U_g w_i, x \rangle \geq \Phi_{i,j}(x)\}$$

and

$$\{g \in G \mid \langle U_g w_i, x + y \rangle \leq \Phi_{i,j}(x + y)\} \subset \{g \in G \mid \langle U_g w_i, x \rangle \leq \Phi_{i,j}(x)\}.$$

Proof Assume there exists $g \in G$ such that $g \in L^{i,j}(x + y)$ but $g \notin L^{i,j}(x)$. Assume first that $\langle U_g w_i, x \rangle < \Phi_{i,j}(x)$; the opposite case $\langle U_g w_i, x \rangle > \Phi_{i,j}(x)$ is handled by the mirror argument, run with the set $\cup_{k \geq j} L^{i,k}(x)$ (which contains the bottom $N - j + 1$ sorted entries, hence at least $N - j + 1$ elements) and (7) in place of $\cup_{k \leq j} L^{i,k}(x)$ and (6). So, for every $h \in \cup_{k \leq j} L^{i,k}(x)$

$$\begin{aligned} \langle U_h w_i, x + y \rangle - \langle U_g w_i, x + y \rangle &\geq \langle U_h w_i, x \rangle - \langle U_g w_i, x \rangle - 2\|y\|\|w_i\| \\ &\geq \|w_i\|\Delta^{i,j}(x) - 2\|y\|\|w_i\| > 0. \end{aligned}$$

Here the second inequality holds because $\langle U_h w_i, x \rangle \geq \Phi_{i,j}(x)$ while, by the definition of $\Delta^{i,j}(x)$ and since $g \notin L^{i,j}(x)$, $\Phi_{i,j}(x) - \langle U_g w_i, x \rangle \geq \|w_i\|\Delta^{i,j}(x)$; the last (strict) inequality follows from the assumption $\|y\| < \frac{1}{2}\Delta^{i,j}(x)$. On the other hand, $\langle U_g w_i, x + y \rangle = \Phi_{i,j}(x + y)$. Thus,

$$\cup_{k \leq j} L^{i,k}(x) \subset \{h \in G \mid \langle U_h w_i, x + y \rangle > \Phi_{i,j}(x + y)\}.$$

But the set $\cup_{k \leq j} L^{i,k}(x)$ contains at least j elements (the top j sorted entries), and so we obtain a contradiction with Lemma 2.2, Equation (6).

It remains to verify the four displayed inclusions. Inclusions three and four are the contrapositives of the second and first, so it suffices to prove the first two. Since the action is orthogonal, $\Phi_{i,j}$ and each map $\langle U_g w_i, \cdot \rangle$ are Lipschitz with constant $\|w_i\|$, so both $|\Phi_{i,j}(x+y) - \Phi_{i,j}(x)|$ and $|\langle U_g w_i, x+y \rangle - \langle U_g w_i, x \rangle|$ are $< \frac{1}{2} \|w_i\| \Delta^{i,j}(x)$. If $\langle U_g w_i, x \rangle > \Phi_{i,j}(x)$, then $g \notin L^{i,j}(x)$ gives $\langle U_g w_i, x \rangle \geq \Phi_{i,j}(x) + \|w_i\| \Delta^{i,j}(x)$, whence

$$\langle U_g w_i, x+y \rangle > \Phi_{i,j}(x) + \frac{1}{2} \|w_i\| \Delta^{i,j}(x) > \Phi_{i,j}(x+y),$$

which is the first inclusion; the second is identical with all inequalities reversed. $\square$

Lemma 2.4 *For $i \in [p]$ and $j \in [N]$, fix vectors $x, y \in \mathcal{V}$ and positive numbers c_1, c_2 such that $\max(c_1, c_2) \|y\| < \frac{1}{4} \Delta^{i,j}(x)$. Then, $L^{i,j}(x + c_1 y) = L^{i,j}(x + c_2 y)$.*

Proof Assume there exists $g_1 \in L^{i,j}(x + c_2 y)$ with $g_1 \notin L^{i,j}(x + c_1 y)$. Assume first that $\langle U_{g_1} w_i, x + c_1 y \rangle < \Phi_{i,j}(x + c_1 y)$; the opposite case is handled by the same mirror argument as in the proof of Lemma 2.3, via $\cup_{r \geq j} L^{i,r}(x + c_1 y)$ and (7). The set $\cup_{r \leq j} L^{i,r}(x + c_1 y)$ consists of those h with $\langle U_h w_i, x + c_1 y \rangle \geq \Phi_{i,j}(x + c_1 y)$, so it contains the top j sorted entries and $|\cup_{r \leq j} L^{i,r}(x + c_1 y)| \geq j$; moreover $g_1 \notin \cup_{r \leq j} L^{i,r}(x + c_1 y)$, since $\langle U_{g_1} w_i, x + c_1 y \rangle < \Phi_{i,j}(x + c_1 y)$. On the other hand, from Lemma 2.2, Equation (6),

$$|\{h \in G \mid \langle U_h w_i, x + c_2 y \rangle > \Phi_{i,j}(x + c_2 y)\}| \leq j - 1.$$

Hence

$$\cup_{r \leq j} L^{i,r}(x + c_1 y) \setminus \{h \in G \mid \langle U_h w_i, x + c_2 y \rangle > \Phi_{i,j}(x + c_2 y)\} \neq \emptyset.$$

Therefore there exists $h \in \cup_{r \leq j} L^{i,r}(x + c_1 y)$ such that

$$\langle U_h w_i, x + c_2 y \rangle \leq \Phi_{i,j}(x + c_2 y) = \langle U_{g_1} w_i, x + c_2 y \rangle. \quad (8)$$

Moreover, by Lemma 2.3, $g_1 \in L^{i,j}(x)$.

But if $\langle U_h w_i, x \rangle - \langle U_{g_1} w_i, x \rangle > 0$, then

$$\langle U_h w_i, x + c_2 y \rangle - \langle U_{g_1} w_i, x + c_2 y \rangle \geq \|w_i\| (\Delta^{i,j}(x) - 2c_2 \|y\|) > 0,$$

which contradicts (8). If $\langle U_h w_i, x \rangle - \langle U_{g_1} w_i, x \rangle < 0$, then

$$\langle U_{g_1} w_i, x + c_1 y \rangle - \langle U_h w_i, x + c_1 y \rangle \geq \|w_i\| (\Delta^{i,j}(x) - 2c_1 \|y\|) > 0,$$

which contradicts $h \in \cup_{r \leq j} L^{i,r}(x + c_1 y)$. Therefore $\langle U_h w_i, x \rangle = \langle U_{g_1} w_i, x \rangle$ and thus $h \in L^{i,j}(x)$. But then

$$\begin{aligned} 0 &\geq \langle U_h w_i, x + c_2 y \rangle - \langle U_{g_1} w_i, x + c_2 y \rangle = \langle U_h w_i, c_2 y \rangle - \langle U_{g_1} w_i, c_2 y \rangle \\ &= c_2 (\langle U_h w_i, y \rangle - \langle U_{g_1} w_i, y \rangle) \end{aligned}$$

and

$$\begin{aligned} 0 &< \langle U_h w_i, x + c_1 y \rangle - \langle U_{g_1} w_i, x + c_1 y \rangle = \langle U_h w_i, c_1 y \rangle - \langle U_{g_1} w_i, c_1 y \rangle \\ &= c_1 (\langle U_h w_i, y \rangle - \langle U_{g_1} w_i, y \rangle). \end{aligned}$$

These contradict each other since $c_1, c_2 > 0$. This proves $L^{i,j}(x + c_2 y) \subset L^{i,j}(x + c_1 y)$; the reverse inclusion follows by exchanging the roles of c_1 and c_2 , since the hypothesis is symmetric in them. $\square$

The induction uses two kinds of perturbation of a partial sum $\sum_{r=1}^k z_r$: rescaling the components to $\sum_r a_r z_r$ with each a_r close to 1, and adding a small vector e . Lemma 2.5 shows that under both of them the active sets $L^{i,j}$ are unchanged, while the gaps $\Delta^{i,j}$ change by at most a fixed factor. This is what lets us carry a whole neighborhood of $\sum_r z_r$, rather than a single point, through the argument.

Lemma 2.5 (Perturbation of active sets and gaps) *Fix $w_i \in \mathcal{V} \setminus \{0\}$ and $j \in [N]$. For any $k \geq 1$, fix $z_1 \in \mathcal{V}$ of unit norm, $\|z_1\| = 1$, and choose arbitrary vectors $z_2, \dots, z_k$ that satisfy only the norm conditions*

$$\|z_{l+1}\| \leq \min\left(\frac{1}{4}\Delta^{i,j}\left(\sum_{r=1}^l z_r\right), \frac{1}{4}\|z_l\|\right) \quad \forall l \in [k-1].$$

Denote $\delta^{i,j} = \min\left(\frac{1}{8}, \frac{1}{16k}\Delta^{i,j}\left(\sum_{r=1}^k z_r\right)\right)$. For any scalars $a_1, \dots, a_k \in (1 - \delta^{i,j}, 1 + \delta^{i,j})$, the following hold:

1. *If $g_1, g_2 \in L^{i,j}(\sum_{r=1}^k z_r)$ then*

$$\langle U_{g_1} w_i, \sum_{r=1}^k a_r z_r \rangle = \langle U_{g_2} w_i, \sum_{r=1}^k a_r z_r \rangle.$$

2.

$$L^{i,j}\left(\sum_{r=1}^k a_r z_r\right) = L^{i,j}\left(\sum_{r=1}^k z_r\right). \quad (9)$$

3.

$$\frac{1}{4}\Delta^{i,j}\left(\sum_{r=1}^k a_r z_r\right) < \Delta^{i,j}\left(\sum_{r=1}^k z_r\right) < 4\Delta^{i,j}\left(\sum_{r=1}^k a_r z_r\right). \quad (10)$$

4. For every $e \in \mathcal{V}$ with $\|e\| < \frac{1}{16} \Delta^{i,j}(\sum_{r=1}^k z_r)$,

$$L^{i,j}(\sum_{r=1}^k a_r z_r + e) = L^{i,j}(\sum_{r=1}^k z_r + e). \quad (11)$$

Remark 2.6 Note that this lemma allows us to choose arbitrary directions for the vectors $z_2, \dots, z_k$. The conclusions of the lemma remain valid even if $z_2 = \dots = z_k = 0$. On the other hand, the norm conditions

$$\|z_{l+1}\| \leq \frac{1}{4} \|z_l\| \quad \text{and} \quad \|z_1\| = 1$$

ensure that the sum $\sum_{l=1}^k z_l$ can never be zero. Hence,

$$\Delta^{i,j}\left(\sum_{r=1}^l z_r\right) > 0 \quad \text{for all } l \in [k].$$

Proof We first record an estimate that will be used repeatedly. Since $\|z_1\| = 1$ and $\|z_{r+1}\| \leq \frac{1}{4} \|z_r\|$, we have $\sum_{r=1}^k \|z_r\| < \frac{4}{3}$, hence

$$2 \sum_{r=1}^k |1 - a_r| \|z_r\| \leq 2 \delta^{i,j} \sum_{r=1}^k \|z_r\| < \frac{8}{3} \delta^{i,j} \leq \frac{1}{6k} \Delta^{i,j}\left(\sum_{r=1}^k z_r\right) \leq \frac{1}{2} \Delta^{i,j}\left(\sum_{r=1}^k z_r\right). \quad (12)$$

In particular $\|\sum_{r=1}^k (a_r - 1)z_r\| < \frac{1}{2} \Delta^{i,j}(\sum_{r=1}^k z_r)$, so Lemma 2.3 applies to the perturbation $\sum_{r=1}^k (a_r - 1)z_r$ of $\sum_{r=1}^k z_r$; this is how Lemma 2.3 is invoked in part 2 below.

1. Note that by the norm conditions on z_r and Lemma 2.3, if $g_1, g_2 \in L^{i,j}(\sum_{r=1}^k z_r)$, then $\langle U_{g_1} w_i, \sum_{r=1}^l z_r \rangle = \langle U_{g_2} w_i, \sum_{r=1}^l z_r \rangle \forall l \in [k]$. Starting with $l = 1$ and proceeding recursively, we get $\langle U_{g_1} w_i, z_r \rangle = \langle U_{g_2} w_i, z_r \rangle \forall r \in [k]$. Therefore $\langle U_{g_1} w_i, a_r z_r \rangle = \langle U_{g_2} w_i, a_r z_r \rangle \forall r \in [k]$ and so $\langle U_{g_1} w_i, \sum_{r=1}^k a_r z_r \rangle = \langle U_{g_2} w_i, \sum_{r=1}^k a_r z_r \rangle$.
2. “ $\subset$ ”. From Lemma 2.3, we have that $L^{i,j}(\sum_{r=1}^k a_r z_r) \subset L^{i,j}(\sum_{r=1}^k z_r)$.
 “ $\supset$ ”. Let $g_1 \in L^{i,j}(\sum_{r=1}^k z_r)$. Take any $g_2 \in L^{i,j}(\sum_{r=1}^k a_r z_r)$. By Lemma 2.3, $g_2 \in L^{i,j}(\sum_{r=1}^k z_r)$, and from part 1 of this lemma,

$$\langle U_{g_1} w_i, \sum_{r=1}^k a_r z_r \rangle = \langle U_{g_2} w_i, \sum_{r=1}^k a_r z_r \rangle.$$

Therefore $g_1 \in L^{i,j}(\sum_{r=1}^k a_r z_r)$, which proves the other inclusion.

3. Case 1. Suppose $L^{i,j}(\sum_{r=1}^k z_r) = G$. By part 2 of this lemma, $L^{i,j}(\sum_{r=1}^k a_r z_r) = L^{i,j}(\sum_{r=1}^k z_r) = G$, so both $\Delta^{i,j}(\sum_{r=1}^k z_r)$ and $\Delta^{i,j}(\sum_{r=1}^k a_r z_r)$ are given by the

$L^{i,j} = G$ branch of the definition, namely $\frac{1}{\|w_i\|} \|\sum_{r=1}^k z_r\|$ and $\frac{1}{\|w_i\|} \|\sum_{r=1}^k a_r z_r\|$ respectively.

Moreover, since $|a_r - 1| < \delta^{i,j} \leq \frac{1}{8}$ for every $r \in [k]$, we have $a_1, \dots, a_k \in (7/8, 9/8)$. Therefore,

$$\Delta^{i,j}(\sum_{r=1}^k a_r z_r) = \frac{1}{\|w_i\|} \|\sum_{r=1}^k a_r z_r\| \leq \frac{9}{8\|w_i\|} \left(\sum_{r=1}^k \|z_r\| \right) < \frac{3}{2\|w_i\|} \|z_1\|$$

and

$$\Delta^{i,j}(\sum_{r=1}^k a_r z_r) = \frac{1}{\|w_i\|} \|\sum_{r=1}^k a_r z_r\| \geq \frac{1}{\|w_i\|} \left(\frac{7}{8} \|z_1\| - \frac{9}{8} \sum_{r=2}^k \|z_r\| \right) > \frac{1}{2\|w_i\|} \|z_1\|.$$

Similarly,

$$\Delta^{i,j}(\sum_{r=1}^k z_r) = \frac{1}{\|w_i\|} \|\sum_{r=1}^k z_r\| \leq \frac{1}{\|w_i\|} \left(\sum_{r=1}^k \|z_r\| \right) < \frac{4}{3\|w_i\|} \|z_1\|$$

and

$$\Delta^{i,j}(\sum_{r=1}^k z_r) = \frac{1}{\|w_i\|} \|\sum_{r=1}^k z_r\| \geq \frac{1}{\|w_i\|} \left(\|z_1\| - \sum_{r=2}^k \|z_r\| \right) > \frac{2}{3\|w_i\|} \|z_1\|.$$

So,

$$\begin{aligned} \frac{1}{4} \Delta^{i,j}(\sum_{r=1}^k a_r z_r) &< \frac{3}{8\|w_i\|} \|z_1\| < \frac{2}{3\|w_i\|} \|z_1\| < \Delta^{i,j}(\sum_{r=1}^k z_r) \\ &< \frac{4}{3\|w_i\|} \|z_1\| < \frac{2}{\|w_i\|} \|z_1\| < 4\Delta^{i,j}(\sum_{r=1}^k a_r z_r), \end{aligned}$$

which is exactly (10).

Case 2. Now assume that $L^{i,j}(\sum_{r=1}^k z_r) \neq G$. By part 2 of this lemma, $L^{i,j}(\sum_{r=1}^k a_r z_r) = L^{i,j}(\sum_{r=1}^k z_r) \neq G$.

For the lower bound, fix $g_1 \in G$ that achieves $\Delta^{i,j}(\sum_{r=1}^k a_r z_r)$, i.e.,

$$\frac{1}{\|w_i\|} |\langle U_{g_1} w_i, \sum_{r=1}^k a_r z_r \rangle - \Phi_{i,j}(\sum_{r=1}^k a_r z_r)| = \Delta^{i,j}(\sum_{r=1}^k a_r z_r),$$

and $g_2 \in L^{i,j}(\sum_{r=1}^k a_r z_r)$. In particular $g_1 \notin L^{i,j}(\sum_{r=1}^k a_r z_r) = L^{i,j}(\sum_{r=1}^k z_r)$ while $g_2 \in L^{i,j}(\sum_{r=1}^k z_r)$, hence

$$\frac{1}{\|w_i\|} \left| \langle U_{g_1} w_i, \sum_{r=1}^k z_r \rangle - \langle U_{g_2} w_i, \sum_{r=1}^k z_r \rangle \right| \geq \Delta^{i,j}(\sum_{r=1}^k z_r).$$

Then, using (12),

$$\begin{aligned}\Delta^{i,j}\left(\sum_{r=1}^k a_r z_r\right) &= \frac{1}{\|w_i\|} \left| \langle U_{g_1} w_i, \sum_{r=1}^k a_r z_r \rangle - \langle U_{g_2} w_i, \sum_{r=1}^k a_r z_r \rangle \right| \\ &\geq \frac{1}{\|w_i\|} \left| \langle U_{g_1} w_i, \sum_{r=1}^k z_r \rangle - \langle U_{g_2} w_i, \sum_{r=1}^k z_r \rangle \right| - 2 \sum_{r=1}^k |1 - a_r| \|z_r\| \\ &\geq \Delta^{i,j}\left(\sum_{r=1}^k z_r\right) - 2 \sum_{r=1}^k |1 - a_r| \|z_r\| > \frac{1}{2} \Delta^{i,j}\left(\sum_{r=1}^k z_r\right).\end{aligned}$$

For the upper bound, we evaluate the perturbed quantity at a group element that achieves the minimum for the *unperturbed* vector. Fix $h_1 \notin L^{i,j}(\sum_{r=1}^k z_r)$ that achieves $\Delta^{i,j}(\sum_{r=1}^k z_r)$, i.e.,

$$\frac{1}{\|w_i\|} \left| \langle U_{h_1} w_i, \sum_{r=1}^k z_r \rangle - \Phi_{i,j}\left(\sum_{r=1}^k z_r\right) \right| = \Delta^{i,j}\left(\sum_{r=1}^k z_r\right),$$

and let $h_2 \in L^{i,j}(\sum_{r=1}^k z_r)$, so that $\Phi_{i,j}(\sum_{r=1}^k z_r) = \langle U_{h_2} w_i, \sum_{r=1}^k z_r \rangle$. By part 2 of this lemma, $h_1 \notin L^{i,j}(\sum_{r=1}^k a_r z_r)$ and $h_2 \in L^{i,j}(\sum_{r=1}^k a_r z_r)$, so that $\Phi_{i,j}(\sum_{r=1}^k a_r z_r) = \langle U_{h_2} w_i, \sum_{r=1}^k a_r z_r \rangle$. Therefore, by the definition of $\Delta^{i,j}$ and (12),

$$\begin{aligned}\Delta^{i,j}\left(\sum_{r=1}^k a_r z_r\right) &\leq \frac{1}{\|w_i\|} \left| \langle U_{h_1} w_i, \sum_{r=1}^k a_r z_r \rangle - \langle U_{h_2} w_i, \sum_{r=1}^k a_r z_r \rangle \right| \\ &\leq \frac{1}{\|w_i\|} \left| \langle U_{h_1} w_i, \sum_{r=1}^k z_r \rangle - \langle U_{h_2} w_i, \sum_{r=1}^k z_r \rangle \right| + 2 \sum_{r=1}^k |1 - a_r| \|z_r\| \\ &= \Delta^{i,j}\left(\sum_{r=1}^k z_r\right) + 2 \sum_{r=1}^k |1 - a_r| \|z_r\| \leq 2 \Delta^{i,j}\left(\sum_{r=1}^k z_r\right).\end{aligned}$$

Therefore,

$$\frac{1}{2} \Delta^{i,j}\left(\sum_{r=1}^k z_r\right) < \Delta^{i,j}\left(\sum_{r=1}^k a_r z_r\right) \leq 2 \Delta^{i,j}\left(\sum_{r=1}^k z_r\right),$$

which in particular implies (10).

4. By Lemma 2.3, $L^{i,j}(\sum_{r=1}^k z_r + e) \subset L^{i,j}(\sum_{r=1}^k z_r)$ and $L^{i,j}(\sum_{r=1}^k a_r z_r + e) \subset L^{i,j}(\sum_{r=1}^k a_r z_r)$; by part 2 these two sets coincide, so put $L := L^{i,j}(\sum_{r=1}^k z_r) = L^{i,j}(\sum_{r=1}^k a_r z_r)$. Both $L^{i,j}(\sum_{r=1}^k z_r + e)$ and $L^{i,j}(\sum_{r=1}^k a_r z_r + e)$ are thus contained in L , and it suffices to show they single out the same elements of L .

By part 1, for $g \in L$ the values $\Phi := \langle U_g w_i, \sum_{r=1}^k z_r \rangle$ and $\Phi' := \langle U_g w_i, \sum_{r=1}^k a_r z_r \rangle$ are independent of g . Hence for every $g \in L$,

$$\langle U_g w_i, \sum_{r=1}^k z_r + e \rangle - \langle U_g w_i, \sum_{r=1}^k a_r z_r + e \rangle = \Phi - \Phi',$$

a constant independent of g : the two coorbit functions, restricted to L , induce the same order on L .

The elements ranked strictly above L are also the same for the two sums. The two perturbations $\sum_{r=1}^k (a_r - 1)z_r$ and $\sum_{r=1}^k (1 - a_r)z_r$ have the same norm, and by the middle estimate of (12),

$$\left\| \sum_{r=1}^k (a_r - 1)z_r \right\| \leq \sum_{r=1}^k |1 - a_r| \|z_r\| < \frac{1}{12k} \Delta^{i,j} \left(\sum_{r=1}^k z_r \right) \leq \frac{1}{12} \Delta^{i,j} \left(\sum_{r=1}^k z_r \right),$$

while part 3 gives $\frac{1}{2} \Delta^{i,j} \left(\sum_{r=1}^k a_r z_r \right) > \frac{1}{8} \Delta^{i,j} \left(\sum_{r=1}^k z_r \right)$. Hence this perturbation has norm smaller than half the gap at *both* base points, so the inclusions of Lemma 2.3 apply at the base point $\sum_{r=1}^k z_r$ with perturbation $\sum_{r=1}^k (a_r - 1)z_r$ and, in the reverse direction, at the base point $\sum_{r=1}^k a_r z_r$ with perturbation $\sum_{r=1}^k (1 - a_r)z_r$, yielding

$$\{g \in G \mid \langle U_g w_i, \sum_{r=1}^k z_r \rangle > \Phi\} = \{g \in G \mid \langle U_g w_i, \sum_{r=1}^k a_r z_r \rangle > \Phi'\};$$

call its cardinality $M \leq j-1$. Since $\|e\| < \frac{1}{16} \Delta^{i,j} \left(\sum_{r=1}^k z_r \right) < \frac{1}{2} \Delta^{i,j} \left(\sum_{r=1}^k a_r z_r \right)$ (part 3), adding e keeps each such element strictly above every element of L , for both sums; the identical estimate keeps every element ranked strictly below L strictly below it. Hence position j is still attained inside L (note that $M \leq j-1$ and $j \leq M + |L|$, since at the unperturbed vectors the j -th sorted value is attained on L): $L^{i,j} \left(\sum_{r=1}^k z_r + e \right)$ consists of the elements of L attaining the $(j - M)$ -th largest value of $g \mapsto \langle U_g w_i, \sum_{r=1}^k z_r + e \rangle$ on L , and likewise for $\sum_{r=1}^k a_r z_r + e$. As these two functions differ on L only by the constant $\Phi - \Phi'$, they select the same elements. Therefore

$$L^{i,j} \left(\sum_{r=1}^k z_r + e \right) = L^{i,j} \left(\sum_{r=1}^k a_r z_r + e \right).$$

□

2.2.2 Positivity of the lower Lipschitz constant

We now prove that the lower Lipschitz bound must be positive if the embedding map $\hat{\Phi}_{w,S}$ is injective. The proof proceeds by contradiction.

The strategy is as follows: assume that the lower Lipschitz constant is zero.

- First, we find a unit-norm vector z_1 where the local lower Lipschitz constant vanishes.

- Next, we inductively construct mutually orthogonal vectors $z_2, \dots, z_d$ such that for each k the local lower Lipschitz constant vanishes at the sum $\sum_{r=1}^k z_r$, and the active sets $L^{i,j}$ of the sum are well separated (the “robotic arm” construction, Lemma 2.14).
- When $k = d$, the vectors $\{z_1, \dots, z_d\}$ form a basis of $\mathcal{V}$. From the vanishing limit, we obtain a unit vector $\tilde{z}$ satisfying $\langle U_g w_i, \tilde{z} \rangle = 0$ for every g in the relevant active sets. Choosing a small $\epsilon > 0$ and comparing $\sum_{r=1}^d z_r$ with $\sum_{r=1}^d z_r + \epsilon \tilde{z}$ yields two distinct points whose images under $\Phi_{\mathbf{w},S}$ coincide, contradicting the injectivity of $\hat{\Phi}_{\mathbf{w},S}$.

We begin by showing that if the lower Lipschitz bound is zero, then this bound can be achieved locally. We say that the *local lower Lipschitz constant of $\Phi_{\mathbf{w},S}$ vanishes at $z \in \mathcal{V}$* if there exist sequences $(x_n)_n, (y_n)_n$ in $\mathcal{V}$ with $x_n \approx y_n$ for all n , $\lim_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} y_n = z$, and

$$\lim_{n \rightarrow \infty} \frac{\|\Phi_{\mathbf{w},S}(x_n) - \Phi_{\mathbf{w},S}(y_n)\|^2}{\mathbf{d}([x_n], [y_n])^2} = 0.$$

Lemma 2.7 (Base case: Localizing the degeneracy) *Fix $\mathbf{w} = (w_1, \dots, w_p) \in \mathcal{V}^p$ and $S \subset [N] \times [p]$. Assume that $\hat{\Phi}_{\mathbf{w},S}$ is injective. If the lower Lipschitz constant of the map $\Phi_{\mathbf{w},S}$ is zero, then there exist sequences $(x_n)_n, (y_n)_n$ in $\mathcal{V}$ such that*

$$\lim_{n \rightarrow \infty} \frac{\|\Phi_{\mathbf{w},S}(x_n) - \Phi_{\mathbf{w},S}(y_n)\|^2}{\mathbf{d}([x_n], [y_n])^2} = 0$$

and which, additionally, satisfy the following relations:

1. (convergence) They share a common limit z_1 ,

$$\lim_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} y_n = z_1, \quad (13)$$

with $\|z_1\| = 1$;

2. (boundedness) For all n :

$$\|x_n\| = 1 \quad (14)$$

$$\|y_n\| \leq 1 \quad (15)$$

3. (alignment) For all n :

$$\|x_n - y_n\| = \min_{g \in G} \|x_n - U_g y_n\| \quad (16)$$

$$\|x_n - z_1\| = \min_{g \in G} \|x_n - U_g z_1\| \quad (17)$$

$$\|y_n - z_1\| = \min_{g \in G} \|y_n - U_g z_1\|. \quad (18)$$

Proof Because the lower Lipschitz bound of the map $\Phi_{\mathbf{w},S}$ is zero, we have that

$$\inf_{\substack{x, y \in \mathcal{V} \\ x \approx y}} \frac{\|\Phi_{\mathbf{w},S}(x) - \Phi_{\mathbf{w},S}(y)\|^2}{\mathbf{d}([x], [y])^2} = 0.$$

Thus, we can find sequences $(x_n)_n, (y_n)_n$ in $\mathcal{V}$ such that

$$\lim_{n \rightarrow \infty} \frac{\|\Phi_{\mathbf{w},S}(x_n) - \Phi_{\mathbf{w},S}(y_n)\|^2}{\mathbf{d}([x_n], [y_n])^2} = 0.$$

Now, notice that for all $t > 0$, we have $\Phi_{\mathbf{w},S}(tx) = t\Phi_{\mathbf{w},S}(x)$ and $\mathbf{d}([tx], [ty]) = t\mathbf{d}([x], [y])$. So, for every $t > 0$,

$$\frac{\|\Phi_{\mathbf{w},S}(x_n) - \Phi_{\mathbf{w},S}(y_n)\|^2}{\mathbf{d}([x_n], [y_n])^2} = \frac{\|\Phi_{\mathbf{w},S}(tx_n) - \Phi_{\mathbf{w},S}(ty_n)\|^2}{\mathbf{d}([tx_n], [ty_n])^2}.$$

By setting

$$t = \frac{1}{\max(\|x_n\|, \|y_n\|)},$$

we can assume without loss of generality that both x_n and y_n lie in the unit ball. Moreover, since the formulas are symmetric in the two sequences, we can further assume that one of them, say $(x_n)_n$, lies on the unit sphere. That is,

$$\|x_n\| = 1 \quad \text{and} \quad \|y_n\| \leq 1 \quad \text{for all } n \in \mathbb{N}.$$

By compactness, we can find a convergent subsequence $(x_{n_k})_k$ of $(x_n)_n$ with $x_{n_k} \rightarrow x_\infty$. Similarly, we can find a convergent subsequence $(y_{n_{k_l}})_l$ of $(y_{n_k})_k$ with $y_{n_{k_l}} \rightarrow y_\infty$. Clearly, $x_{n_{k_l}} \rightarrow x_\infty$. For ease of notation, we denote the sequences $(x_{n_{k_l}})_l$ and $(y_{n_{k_l}})_l$ by $(x_n)_n$ and $(y_n)_n$, respectively.

Next, suppose that $x_\infty \approx y_\infty$. Then,

$$\frac{\|\Phi_{\mathbf{w},S}(x_\infty) - \Phi_{\mathbf{w},S}(y_\infty)\|^2}{\mathbf{d}([x_\infty], [y_\infty])^2} = \lim_{n \rightarrow \infty} \frac{\|\Phi_{\mathbf{w},S}(x_n) - \Phi_{\mathbf{w},S}(y_n)\|^2}{\mathbf{d}([x_n], [y_n])^2} = 0,$$

and thus, $\Phi_{\mathbf{w},S}(x_\infty) = \Phi_{\mathbf{w},S}(y_\infty)$, which contradicts the injectivity assumption. Hence, $x_\infty \sim y_\infty$.

Now, let us denote by g_∞ a group element such that $x_\infty = U_{g_\infty}y_\infty$. Observe that $\lim_{n \rightarrow \infty} \|x_n - U_{g_\infty}y_n\| = 0$. For each $n \in \mathbb{N}$ there exists at least one element $g_n \in G$ that achieves the quotient distance between x_n and $U_{g_\infty}y_n$, i.e., satisfying $\mathbf{d}([x_n], [U_{g_\infty}y_n]) = \|x_n - U_{g_n g_\infty}y_n\|$. But G is a finite group, so there must exist an element $g_0 \in G$ for which $g_n = g_0$ for infinitely many n . Let $(n_m)_m$ be the sequence of all such indices. We see that $\mathbf{d}([x_{n_m}], [U_{g_\infty}y_{n_m}]) = \|x_{n_m} - U_{g_0 g_\infty}y_{n_m}\|$ for all $m \in \mathbb{N}$.

Finally, for every $m \in \mathbb{N}$, let $g_m \in G$ be a group element that achieves the quotient distance between x_{n_m} and x_∞ , that is,

$$\mathbf{d}([x_{n_m}], [x_\infty]) = \|U_{g_m} x_{n_m} - x_\infty\|.$$

Denote $U_{g_m} x_{n_m}$ by x_m and $U_{g_m g_0 g_\infty} y_{n_m}$ by y_m ; for ease of notation we re-index these sequences by n and keep denoting them $(x_n)_n$ and $(y_n)_n$. So far we have obtained two sequences $(x_n)_n$ and $(y_n)_n$ that satisfy (13–17). Now let $h_n \in G$ denote a group element such that $\mathbf{d}([y_n], [z_1]) = \|y_n - U_{h_n} z_1\|$. Since G is finite, pass to a subsequence (again indexed by n) so that $h_n = h_0$ for some $h_0 \in G$. Therefore $\mathbf{d}([y_n], [z_1]) = \|y_n - U_{h_0} z_1\| \leq \|y_n - z_1\|$. But $\lim_{n \rightarrow \infty} y_n = z_1$. Thus, $U_{h_0} z_1 = z_1$. This shows (18), and the lemma is proved. $\square$

In what follows, we will denote by $H(z)$ the stabilizer group of z ; recall that

$$H(z) = \{g \in G \mid U_g z = z\}.$$

For a fixed nonzero vector z , we define the strictly positive number

$$\rho_0(z) = \begin{cases} \min_{g \in G \setminus H(z)} \|z - U_g z\|, & \text{if } H(z) \neq G \\ \|z\|, & \text{if } H(z) = G. \end{cases}$$

The following elementary observation shows that a sufficiently small perturbation cannot enlarge the stabilizer.

Lemma 2.8 (Small perturbations do not enlarge the stabilizer) *Let $z \in \mathcal{V} \setminus \{0\}$ and $e \in \mathcal{V}$ with $\|e\| < \frac{1}{2}\rho_0(z)$. Then, $H(z + e) \subseteq H(z)$.*

Proof If $H(z) = G$ then $\rho_0(z) = \|z\|$, so $\|e\| < \frac{1}{2}\|z\|$ and $z + e \neq 0$; the claim $H(z + e) \subseteq G = H(z)$ is trivial. Otherwise, let $g \in G \setminus H(z)$. Then,

$$\|U_g(z + e) - (z + e)\| \geq \|U_g z - z\| - \|U_g e - e\| \geq \rho_0(z) - 2\|e\| > 0,$$

so $g \notin H(z + e)$. Hence $H(z + e) \subseteq H(z)$. $\square$

Remark 2.9 Consequently, if $z_1, \dots, z_k$ satisfy $\|z_{l+1}\| < \frac{1}{2}\rho_0(\sum_{r=1}^l z_r)$ for all $l \in [k-1]$, then, applying Lemma 2.8 repeatedly with $z = \sum_{r=1}^l z_r$ and $e = z_{l+1}$,

$$H(\sum_{r=1}^k z_r) \subseteq H(\sum_{r=1}^{k-1} z_r) \subseteq \dots \subseteq H(z_1).$$

In particular, every $g \in H(\sum_{r=1}^k z_r)$ fixes each partial sum $\sum_{r=1}^l z_r$, and hence fixes each increment $z_l = \sum_{r=1}^l z_r - \sum_{r=1}^{l-1} z_r$: $U_g z_l = z_l$ for all $l \in [k]$.

Assume N_0 is large enough so that $\|x_n - z_1\| < \frac{1}{8}\rho_0(z_1)$ and $\|x_n - y_n\| < \frac{1}{8}\rho_0(z_1)$ for all $n > N_0$. Then,

$$\|y_n - z_1\| \leq \|y_n - x_n\| + \|x_n - z_1\| < \frac{\rho_0(z_1)}{4}.$$

Remark 2.10 In Fig. 3, we illustrate how the sequences $(x_n)_n$ and $(y_n)_n$ are constructed.

Lemma 2.11 (Alignment) *Let $z \in \mathcal{V} \setminus \{0\}$ and let $u, v \in \mathcal{V}$ with $\|u\|, \|v\| < \frac{1}{4}\rho_0(z)$. Let $x = z + u$ and $y = z + v$. Then, the following properties hold:*

1. $\mathbf{d}([x], [z]) = \|u\|$ and $\mathbf{d}([y], [z]) = \|v\|$, and
2. $\mathbf{d}([x], [y]) = \min_{g \in H(z)} \|u - U_g v\| = \min_{g \in H(z)} \|U_g u - v\|$.

Proof 1. If $u = 0$, then the claim is trivial. If $u \neq 0$, then $\mathbf{d}([x], [z]) = \min_{g \in G} \|x - U_g z\| = \min_{g \in G} \|z - U_g z + u\| \leq \|u\|$. On the other hand, suppose the minimum is achieved for some $g \in G$. If $g \in H(z)$, then $\mathbf{d}([x], [z]) = \|u\|$. If $g \notin H(z)$, then

$$\mathbf{d}([x], [z]) = \|z - U_g z + u\| \geq \|z - U_g z\| - \|u\| \geq \rho_0(z) - \|u\| > \|u\|,$$

since $\|u\| < \frac{1}{4}\rho_0(z)$; this contradicts $\mathbf{d}([x], [z]) \leq \|u\|$. Hence the minimum is achieved on $H(z)$ and $\mathbf{d}([x], [z]) = \|u\|$. The claim for y is identical.

2. First, for every $g \in H(z)$, we have $\|U_g x - y\| = \|U_g z - z + U_g u - v\| = \|U_g u - v\|$, and therefore $\mathbf{d}([x], [y]) \leq \min_{g \in H(z)} \|U_g u - v\|$. On the other hand, for $g \in G \setminus H(z)$ and any $h \in H(z)$,

$$\begin{aligned} \|U_g x - y\| &= \|U_g z - z + U_g u - v\| \\ &\geq \|U_g z - z\| - \|u\| - \|v\| \\ &\geq \rho_0(z) - 2\|u\| - 2\|v\| + \|U_h u - v\| \\ &\geq \|U_h u - v\|, \end{aligned}$$

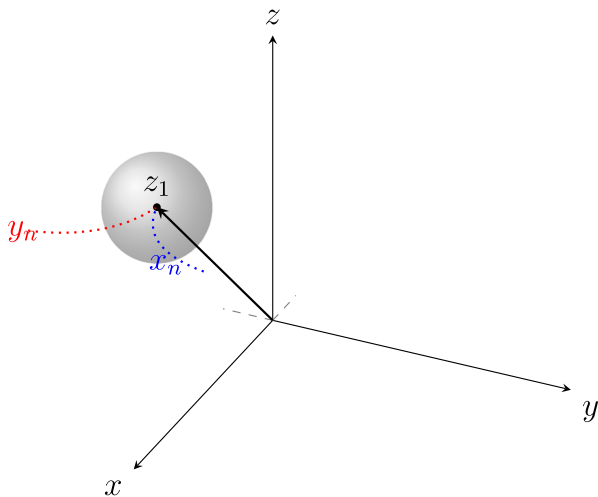


Fig. 3 Local analysis of sequences x_n and y_n converging to z_1

where we used $\|U_h u - v\| \leq \|u\| + \|v\|$ and $2\|u\| + 2\|v\| < \rho_0(z)$. Hence the minimum of $\|U_g x - y\|$ over $g \in G$ is attained on $H(z)$, and $\mathbf{d}([x], [y]) = \min_{g \in H(z)} \|U_g u - v\|$. The second equality in the statement follows by replacing g with g^{-1} , since $H(z)$ is a group and each U_g is orthogonal. $\square$

Remark 2.12 Applying Lemma 2.11 to two sequences $(x_k)_k$ and $(y_k)_k$ that satisfy (13–16) in Lemma 2.7, it follows that $\mathbf{d}([x_k], [z_1]) = \|x_k - z_1\|$ and $\mathbf{d}([y_k], [z_1]) = \|y_k - z_1\|$ for k large enough. Hence alignment must occur from some index on.

Lemma 2.13 *Let $(x_n)_n, (y_n)_n$ be two arbitrary sequences in $\mathcal{V}$. For each $i \in [p]$ and $j \in S_i$, we denote by $g_{1,i,j,n}$ a group element that achieves $\Phi_{i,j}(x_n)$ and by $g_{2,i,j,n}$ a group element that achieves $\Phi_{i,j}(y_n)$. That is, $\Phi_{i,j}(x_n) = \langle U_{g_{1,i,j,n}} w_i, x_n \rangle$ and $\Phi_{i,j}(y_n) = \langle U_{g_{2,i,j,n}} w_i, y_n \rangle$.*

We can find a sequence of natural numbers $(n_r)_r$ such that $g_{1,i,j,n_r} = g_{1,i,j}$ and $g_{2,i,j,n_r} = g_{2,i,j} \forall r \in \mathbb{N}, i \in [p], j \in S_i$.

Proof Fix an enumeration $(i_1, j_1), \dots, (i_m, j_m)$ of the index set S . Since G is finite, for the first pair (i_1, j_1) there is a subsequence $(x_{n_t})_t$ such that $g_{1,i_1,j_1,n_t} = g_{1,i_1,j_1}$ for every $t \in \mathbb{N}$. Similarly, for the second pair (i_2, j_2) , we can find a further subsequence, which we denote $(x_{n_l})_l$, such that $g_{1,i_2,j_2,n_l} = g_{1,i_2,j_2}, \forall l \in \mathbb{N}$. So, by induction, after $\sum_{i \in [p]} m_i = m$ steps, we construct a subsequence of $(x_n)_n$, which we denote $(x_{n_t})_t$, such that $g_{1,i,j,n_t} = g_{1,i,j}$ for every $i \in [p], j \in S_i$. Starting from the sequence $(y_{n_t})_t$, we repeat the same procedure, obtaining a subsequence $(y_{n_r})_r$ such that $g_{2,i,j,n_r} = g_{2,i,j}$ for every $r \in \mathbb{N}, i \in [p], j \in S_i$. Notice that the sequences $(x_{n_r})_r$ and $(y_{n_r})_r$, which from now on, we will call $(x_n)_n$ and $(y_n)_n$ for ease of notation, have the constant achievers claimed in the statement. $\square$

For the sequences $(x_n)_n, (y_n)_n$ and the vector z_1 defined above, let $u_n = x_n - z_1$ and $v_n = y_n - z_1$. Notice that

$$\begin{aligned} \|\Phi_{\mathbf{w},S}(x_n) - \Phi_{\mathbf{w},S}(y_n)\|^2 &= \sum_{i=1}^p \sum_{j \in S_i} |\langle U_{g_{1,i,j}} w_i, x_n \rangle - \langle U_{g_{2,i,j}} w_i, y_n \rangle|^2 \\ &= \sum_{i=1}^p \sum_{j \in S_i} |\langle U_{g_{1,i,j}} w_i - U_{g_{2,i,j}} w_i, z_1 \rangle \\ &\quad + \langle w_i, U_{g_{1,i,j}}^{-1} u_n - U_{g_{2,i,j}}^{-1} v_n \rangle|^2. \end{aligned}$$

This sequence converges to 0 as $n \rightarrow \infty$, while also $u_n, v_n \rightarrow 0$. So we conclude that for each $i \in [p]$ and $j \in S_i$, $\langle U_{g_{1,i,j}} w_i - U_{g_{2,i,j}} w_i, z_1 \rangle = 0$. Therefore

$$\|\Phi_{\mathbf{w},S}(x_n) - \Phi_{\mathbf{w},S}(y_n)\|^2 = \sum_{i=1}^p \sum_{j \in S_i} |\langle w_i, U_{g_{1,i,j}}^{-1} u_n - U_{g_{2,i,j}}^{-1} v_n \rangle|^2.$$

Thus, we have

$$\lim_{n \rightarrow \infty} \frac{\sum_{i=1}^p \sum_{j \in S_i} |\langle w_i, U_{g_{1,i,j}}^{-1} u_n - U_{g_{2,i,j}}^{-1} v_n \rangle|^2}{\|u_n - v_n\|^2} = 0, \quad (19)$$

where $\|u_n\|, \|v_n\| \rightarrow 0$, so for large enough n , we have $\|u_n\|, \|v_n\| \leq \frac{1}{4}\rho_0(z_1)$. Since $x_n - y_n = u_n - v_n$, the alignment property (16) of Lemma 2.7 shows that the denominator in (19) satisfies $\|u_n - v_n\| = \|x_n - y_n\| = \mathbf{d}([x_n], [y_n])$.

The next lemma is the inductive step of the robotic-arm construction: it produces a vector z_{k+1} , orthogonal to $z_1, \dots, z_k$, along which the ratio still vanishes.

Lemma 2.14 (Induction step: the robotic arm) *Fix $p \in \mathbb{N}$, $\mathbf{w} \in \mathcal{V}^p$ and $S \subset [N] \times [p]$. Let $\Delta : \mathcal{V} \rightarrow \mathbb{R}$ be defined by $\Delta(x) = \min_{(i,j) \in [p] \times [N]} \Delta^{i,j}(x)$, where the map $\Delta^{i,j}$ is defined in (5). Fix nonzero vectors $z_1, \dots, z_k \in \mathcal{V}$ such that*

$$\|z_1\| = 1, \quad \langle z_i, z_j \rangle = 0 \quad \forall i, j \in [k], \quad i \neq j$$

and

$$\|z_{l+1}\| \leq \min \left(\frac{1}{4} \Delta \left(\sum_{r=1}^l z_r \right), \frac{1}{4} \|z_l\|, \frac{1}{4} \rho_0 \left(\sum_{r=1}^l z_r \right) \right) \quad \forall l \in [k-1].$$

Assume that the local lower Lipschitz constant of $\Phi_{\mathbf{w},S}$ vanishes at $z_1 + z_2 + \dots + z_k$, and that $\hat{\Phi}_{\mathbf{w},S}$ is injective. If $k < d$, then there exists a nonzero vector z_{k+1} such that:

- (i) $\langle z_{k+1}, z_j \rangle = 0 \quad \forall j \in [k]$;
- (ii) $\|z_{k+1}\| \leq \min \left(\frac{1}{4} \Delta \left(\sum_{r=1}^k z_r \right), \frac{1}{4} \|z_k\|, \frac{1}{4} \rho_0 \left(\sum_{r=1}^k z_r \right) \right)$; and
- (iii) The local lower Lipschitz constant vanishes at $z_1 + z_2 + \dots + z_{k+1}$, i.e., there are sequences of vectors $(x_n)_n, (y_n)_n$ such that

$$\lim_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} y_n = \sum_{r=1}^{k+1} z_r$$

and

$$\lim_{n \rightarrow \infty} \frac{\|\Phi_{\mathbf{w},S}(x_n) - \Phi_{\mathbf{w},S}(y_n)\|^2}{\mathbf{d}([x_n], [y_n])^2} = 0.$$

Proof Let $(x_n)_n, (y_n)_n$ be two sequences that both converge to $\sum_{r=1}^k z_r$ and achieve lower Lipschitz bound zero for the map $\Phi_{\mathbf{w},S}$. We align the sequences $(x_n)_n$ and $(y_n)_n$ by the same construction as in Lemma 2.7, now with base point $\sum_{r=1}^k z_r$ in place of the unit vector z_1 : for each n , we replace y_n by the orbit representative $U_{g_n} y_n$ closest to x_n , so that $\mathbf{d}([x_n], [y_n]) = \|x_n - y_n\|$, and we pass to a subsequence along which $g_n \equiv g_0$ is constant. This changes neither the ratio, since $\Phi_{\mathbf{w},S}$ is G -invariant and $\mathbf{d}$ depends only on orbits, nor the common limit: from $x_n, y_n \rightarrow \sum_{r=1}^k z_r$ and

$\|x_n - U_{g_0} y_n\| \rightarrow 0$, we get $U_{g_0} \sum_{r=1}^k z_r = \sum_{r=1}^k z_r$, i.e. $g_0 \in H(\sum_{r=1}^k z_r)$. We denote by $a_n = P_{E_k} x_n$ and $b_n = P_{E_k} y_n$ the orthogonal projections of the sequences $(x_n)_n$ and $(y_n)_n$ respectively, on the linear subspace $E_k = \text{span}\{z_1, \dots, z_k\}^\perp$.

Claim 1: First, we will show that $\exists M_0$ such that $\forall n \geq M_0$, $a_n \neq 0$ or $b_n \neq 0$. Assume otherwise; then there are two sequences of vectors $x_n = \sum_{r=1}^k c_{r,n} z_r$ and $y_n = \sum_{r=1}^k d_{r,n} z_r$, where $\lim_{n \rightarrow \infty} c_{r,n} = \lim_{n \rightarrow \infty} d_{r,n} = 1$, $\forall r \in [k]$ that achieve lower Lipschitz bound zero. Consequently, there is $M_0 \in \mathbb{N}$ such that

$$|1 - c_{r,n}|, |1 - d_{r,n}| < \min \left(\frac{1}{8}, \frac{1}{16k} \Delta \left(\sum_{r=1}^k z_r \right) \right) \leq \delta^{i,j}$$

for all $n \geq M_0$, $r \in [k]$ and $(i, j) \in S$, so part (2) of Lemma 2.5 applies and gives, for all $n \geq M_0$ and $(i, j) \in S$,

$$L^{i,j} \left(\sum_{r=1}^k c_{r,n} z_r \right) = L^{i,j} \left(\sum_{r=1}^k d_{r,n} z_r \right) = L^{i,j} \left(\sum_{r=1}^k z_r \right).$$

Note that $x_n - y_n = \sum_{r=1}^k (c_{r,n} - d_{r,n}) z_r \neq 0$, since $x_n \approx y_n$. Pass to a subsequence $(n_m)_m$ along which the unit vectors $\frac{\sum_{r=1}^k (c_{r,n} - d_{r,n}) z_r}{\|\sum_{r=1}^k (c_{r,n} - d_{r,n}) z_r\|}$ converge, and denote their limit, a unit vector, by $\tilde{z}$. Then, for $g_{i,j} \in L^{i,j}(\sum_{r=1}^k z_r)$,

$$\begin{aligned} 0 &= \lim_{n \rightarrow \infty} \frac{\|\Phi_{\mathbf{w},S}(x_n) - \Phi_{\mathbf{w},S}(y_n)\|^2}{\mathbf{d}([x_n], [y_n])^2} \\ &= \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^p \sum_{j \in S_i} |\langle U_{g_{i,j}} w_i, x_n - y_n \rangle|^2}{\mathbf{d}([x_n], [y_n])^2} \\ &\geq \lim_{m \rightarrow \infty} \frac{\sum_{i=1}^p \sum_{j \in S_i} |\langle U_{g_{i,j}} w_i, \sum_{r=1}^k (c_{r,n_m} - d_{r,n_m}) z_r \rangle|^2}{\|\sum_{r=1}^k (c_{r,n_m} - d_{r,n_m}) z_r\|^2} \\ &= \sum_{i=1}^p \sum_{j \in S_i} |\langle U_{g_{i,j}} w_i, \tilde{z} \rangle|^2. \end{aligned}$$

Since the choice of $g_{i,j} \in L^{i,j}(\sum_{r=1}^k z_r)$ was arbitrary (neither the subsequence $(n_m)_m$ nor $\tilde{z}$ depends on it), the computation above shows that $\langle U_g w_i, \tilde{z} \rangle = 0$ for every $g \in L^{i,j}(\sum_{r=1}^k z_r)$, $i \in [p]$, $j \in S_i$. We now choose $\epsilon > 0$ with $\sum_{r=1}^k z_r \approx \sum_{r=1}^k z_r + \epsilon \tilde{z}$: for each $g \in G$ the equation $\sum_{r=1}^k z_r + \epsilon \tilde{z} = U_g(\sum_{r=1}^k z_r)$ forces $\epsilon \tilde{z} = U_g(\sum_{r=1}^k z_r) - \sum_{r=1}^k z_r$, which, as $\tilde{z} \neq 0$, determines at most one value of ϵ ; since G is finite, at most N values are excluded, so we may also require $\epsilon \|\tilde{z}\| < \frac{1}{4} \Delta(\sum_{r=1}^k z_r)$. For every $(i, j) \in S$, Lemma 2.3 gives $L^{i,j}(\sum_{r=1}^k z_r + \epsilon \tilde{z}) \subset L^{i,j}(\sum_{r=1}^k z_r)$, so evaluating $\Phi_{i,j}$ at any g in the smaller set yields $\Phi_{i,j}(\sum_{r=1}^k z_r + \epsilon \tilde{z}) = \langle U_g w_i, \sum_{r=1}^k z_r \rangle + \epsilon \langle U_g w_i, \tilde{z} \rangle = \Phi_{i,j}(\sum_{r=1}^k z_r)$.

Hence

$$\Phi_{\mathbf{w},S}(\sum_{r=1}^k z_r) = \Phi_{\mathbf{w},S}(\sum_{r=1}^k z_r + \epsilon \tilde{z}),$$

which contradicts the injectivity property. This establishes Claim 1.

Now, we can assume that for all $n \geq M_0$, $a_n = P_{E_k} x_n \neq 0$ or $b_n = P_{E_k} y_n \neq 0$. If need be, pass to a subsequence and/or switch the definitions of x_n and y_n , so that $\|b_n\| \geq \|a_n\|$ for all n ; in particular, $b_n \neq 0$ for all $n \geq M_0$.

Let $c_{r,n}, d_{r,n}$ be the unique coefficients determined by $x_n = \sum_{r=1}^k c_{r,n} z_r + a_n$, $y_n = \sum_{r=1}^k d_{r,n} z_r + b_n$. Note $\lim_{n \rightarrow \infty} c_{r,n} = \lim_{n \rightarrow \infty} d_{r,n} = 1$.

Let $e_n = \sum_{r=1}^k (d_{r,n} - c_{r,n}) z_r + b_n$ and

$$s_n = \frac{\min \left(\|z_k\|, \Delta(\sum_{r=1}^k z_r), \rho_0(\sum_{r=1}^k z_r) \right)}{16\|e_n\|}.$$

Note $\|e_n\| \geq \|b_n\| \geq \|a_n\|$ for all n , and $e_n \neq 0$ for all $n \geq M_0$; discarding the finitely many terms with $n < M_0$, s_n is well defined.

Claim 2: The sequences $\tilde{x}_n = \sum_{r=1}^k z_r + s_n a_n$ and $\tilde{y}_n = \sum_{r=1}^k z_r + s_n e_n$ satisfy

$$\lim_{n \rightarrow \infty} \frac{\|\Phi_{\mathbf{w},S}(\tilde{x}_n) - \Phi_{\mathbf{w},S}(\tilde{y}_n)\|^2}{\mathbf{d}([\tilde{x}_n], [\tilde{y}_n])^2} = 0.$$

Note that $s_n \|e_n\| = \frac{1}{16} \min \left(\|z_k\|, \Delta(\sum_{r=1}^k z_r), \rho_0(\sum_{r=1}^k z_r) \right)$ for every n , and hence $\max(\|s_n a_n\|, \|s_n e_n\|) \leq \frac{1}{16}$. Pass to a subsequence so that both $(s_n a_n)_n$ and $(s_n e_n)_n$ converge, and let $\alpha = \lim_{n \rightarrow \infty} s_n a_n$ and $\beta = \lim_{n \rightarrow \infty} s_n e_n$. Notice that $\|\beta\| = \lim_{n \rightarrow \infty} s_n \|e_n\| > 0$, so $\beta \neq 0$.

The limits

$$\lim_{n \rightarrow \infty} c_{r,n} = \lim_{n \rightarrow \infty} d_{r,n} = 1, \quad \forall r \in [k] \text{ and } \lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} e_n = 0$$

imply that $\exists m_0 \in \mathbb{N}$ such that $\forall n \geq m_0$ and $\forall r \in [k]$

1. $|1 - c_{r,n}| < \min \left(\frac{1}{8}, \frac{1}{16k} \Delta(\sum_{r=1}^k z_r) \right)$
2. $|1 - d_{r,n}| < \min \left(\frac{1}{8}, \frac{1}{16k} \Delta(\sum_{r=1}^k z_r) \right)$
3. $|c_{r,n} - d_{r,n}| < \min \left(\frac{1}{8}, \frac{1}{16k} \Delta(\sum_{r=1}^k z_r) \right)$
4. $\|a_n\| < \frac{1}{16k} \Delta(\sum_{r=1}^k z_r)$
5. $\|e_n\| < \frac{1}{16k} \Delta(\sum_{r=1}^k z_r)$

Since $\Delta(\sum_{r=1}^k z_r) \leq \Delta^{i,j}(\sum_{r=1}^k z_r)$ for every (i, j) , conditions 1–3 place the scalars $c_{r,n}$, $d_{r,n}$ and $1 + d_{r,n} - c_{r,n}$ in the range $(1 - \delta^{i,j}, 1 + \delta^{i,j})$ required by

Lemma 2.5, while, by the definition of s_n ,

$$\max(\|s_n a_n\|, \|s_n e_n\|) \leq \frac{1}{16} \Delta \left(\sum_{r=1}^k z_r \right) \leq \frac{1}{16} \Delta^{i,j} \left(\sum_{r=1}^k z_r \right).$$

So, for any $(i, j) \in S$,

$$\begin{aligned} L^{i,j}(x_n) &= L^{i,j} \left(\sum_{r=1}^k c_{r,n} z_r + a_n \right) = L^{i,j} \left(\sum_{r=1}^k z_r + a_n \right) \\ &= L^{i,j} \left(\sum_{r=1}^k z_r + s_n a_n \right) = L^{i,j}(\tilde{x}_n) \subset L^{i,j} \left(\sum_{r=1}^k z_r \right) = L^{i,j} \left(\sum_{r=1}^k c_{r,n} z_r \right). \end{aligned}$$

Here, the second equality follows from Lemma 2.5 (part 4), the third equality from Lemma 2.4, the inclusion from Lemma 2.3, and the last equality from Lemma 2.5 (part 2).

Similarly,

$$\begin{aligned} L^{i,j}(y_n) &= L^{i,j} \left(\sum_{r=1}^k d_{r,n} z_r + b_n \right) = L^{i,j} \left(\sum_{r=1}^k (1 + d_{r,n} - c_{r,n}) z_r + b_n \right) \\ &= L^{i,j} \left(\sum_{r=1}^k z_r + e_n \right) = L^{i,j} \left(\sum_{r=1}^k z_r + s_n e_n \right) = L^{i,j}(\tilde{y}_n) \\ &\subset L^{i,j} \left(\sum_{r=1}^k z_r \right) = L^{i,j} \left(\sum_{r=1}^k c_{r,n} z_r \right). \end{aligned}$$

Here, the second equality follows from two applications of Lemma 2.5 (part 4), with $e = b_n$ and scalars $d_{r,n}$, respectively $1 + d_{r,n} - c_{r,n}$ (note $\|b_n\| \leq \|e_n\|$); the third equality is the identity $\sum_{r=1}^k (1 + d_{r,n} - c_{r,n}) z_r + b_n = \sum_{r=1}^k z_r + e_n$; the fourth follows from Lemma 2.4; the inclusion and the last equality are as before.

We now express both ratios through the perturbations a_n, e_n .

(i) *Numerators.* The elements $g_{1,i,j} \in L^{i,j}(x_n)$ and $g_{2,i,j} \in L^{i,j}(y_n)$ that achieve $\Phi_{i,j}(x_n)$ and $\Phi_{i,j}(y_n)$ satisfy $g_{1,i,j}, g_{2,i,j} \in L^{i,j}(\sum_{r=1}^k z_r)$ by the inclusions displayed above, and by Lemma 2.13 they may be taken independent of n along a subsequence. The same inclusions give $L^{i,j}(x_n) = L^{i,j}(\tilde{x}_n)$ and $L^{i,j}(y_n) = L^{i,j}(\tilde{y}_n)$, so these elements also achieve $\Phi_{i,j}(\tilde{x}_n)$ and $\Phi_{i,j}(\tilde{y}_n)$. For n large, $c_{r,n}, d_{r,n} \in (1 - \delta^{i,j}, 1 + \delta^{i,j})$, so part 1 of Lemma 2.5 yields both

$$\begin{aligned} \langle U_{g_{1,i,j}} w_i, \sum_{r=1}^k c_{r,n} z_r \rangle &= \langle U_{g_{2,i,j}} w_i, \sum_{r=1}^k c_{r,n} z_r \rangle \quad \text{and} \\ \langle U_{g_{1,i,j}} w_i, \sum_{r=1}^k z_r \rangle &= \langle U_{g_{2,i,j}} w_i, \sum_{r=1}^k z_r \rangle. \end{aligned}$$

Using $x_n = \sum_{r=1}^k c_{r,n} z_r + a_n$, $y_n = \sum_{r=1}^k d_{r,n} z_r + b_n$, $e_n = \sum_{r=1}^k (d_{r,n} - c_{r,n}) z_r + b_n$, the first identity gives

$$\Phi_{i,j}(x_n) - \Phi_{i,j}(y_n) = \langle w_i, U_{g_{1,i,j}}^{-1} a_n - U_{g_{2,i,j}}^{-1} e_n \rangle,$$

and, since $\tilde{x}_n = \sum_{r=1}^k z_r + s_n a_n$ and $\tilde{y}_n = \sum_{r=1}^k z_r + s_n e_n$, the second identity gives

$$\Phi_{i,j}(\tilde{x}_n) - \Phi_{i,j}(\tilde{y}_n) = s_n \langle w_i, U_{g_{1,i,j}}^{-1} a_n - U_{g_{2,i,j}}^{-1} e_n \rangle.$$

Writing $M_n := \sum_{i=1}^p \sum_{j \in S_i} |\langle w_i, U_{g_{1,i,j}}^{-1} a_n - U_{g_{2,i,j}}^{-1} e_n \rangle|^2$, we thus have

$$\|\Phi_{\mathbf{w},S}(x_n) - \Phi_{\mathbf{w},S}(y_n)\|^2 = M_n \text{ and } \|\Phi_{\mathbf{w},S}(\tilde{x}_n) - \Phi_{\mathbf{w},S}(\tilde{y}_n)\|^2 = s_n^2 M_n.$$

(ii) *Denominators.* Recall $a_n - e_n = x_n - y_n$, so $\tilde{x}_n - \tilde{y}_n = s_n(a_n - e_n)$. Since $(x_n)_n, (y_n)_n$ are aligned by the construction at the beginning of this proof, $\mathbf{d}([x_n], [y_n]) = \|x_n - y_n\| = \|a_n - e_n\|$ for all n ; hence the hypothesis reads

$$\lim_{n \rightarrow \infty} \frac{M_n}{\|a_n - e_n\|^2} = 0.$$

We use here that the stabilizer of $\sum_{r=1}^k z_r$ fixes each z_r (this argument follows the one in [11]). Indeed, the vectors z_r satisfy $\|z_{l+1}\| \leq \frac{1}{4} \rho_0(\sum_{r=1}^l z_r)$, so Remark 2.9 gives $U_g z_r = z_r$ for every $g \in H(\sum_{r=1}^k z_r)$ and every $r \in [k]$. Consequently, for any such g , expanding $x_n = \sum_{r=1}^k c_{r,n} z_r + a_n$ and $y_n = \sum_{r=1}^k d_{r,n} z_r + b_n$, and using $U_g z_r = z_r$ together with $e_n = \sum_{r=1}^k (d_{r,n} - c_{r,n}) z_r + b_n$,

$$\begin{aligned} a_n - U_g e_n &= a_n + \sum_{r=1}^k (c_{r,n} - d_{r,n}) z_r - U_g b_n \\ &= \left(\sum_{r=1}^k c_{r,n} z_r + a_n \right) - U_g \left(\sum_{r=1}^k d_{r,n} z_r + b_n \right) = x_n - U_g y_n. \end{aligned}$$

Hence, by the alignment of $(x_n)_n$ and $(y_n)_n$ arranged at the beginning of this proof,

$$\|a_n - U_g e_n\| = \|x_n - U_g y_n\| \geq \mathbf{d}([x_n], [y_n]) = \|a_n - e_n\|,$$

with equality when g is the identity element. Therefore the minimum over $H(\sum_{r=1}^k z_r)$ is attained at the identity:

$$\min_{g \in H(\sum_{r=1}^k z_r)} \|a_n - U_g e_n\| = \|a_n - e_n\|.$$

Finally, $s_n \|a_n\| \leq s_n \|e_n\| = \frac{1}{16} \min(\|z_k\|, \Delta(\sum_{r=1}^k z_r), \rho_0(\sum_{r=1}^k z_r)) < \frac{1}{4} \rho_0(\sum_{r=1}^k z_r)$, so Lemma 2.11 (with $z = \sum_{r=1}^k z_r$, $u = s_n a_n$, $v = s_n e_n$) yields

$$\mathbf{d}([\tilde{x}_n], [\tilde{y}_n]) = s_n \min_{g \in H(\sum_{r=1}^k z_r)} \|a_n - U_g e_n\| = s_n \|a_n - e_n\|.$$

Since $a_n - e_n = x_n - y_n$ and $x_n \approx y_n$, this quotient distance is positive: $\mathbf{d}([\tilde{x}_n], [\tilde{y}_n]) = s_n \|x_n - y_n\| > 0$, so $\tilde{x}_n \approx \tilde{y}_n$ for every n , as required by the definition of the vanishing local lower Lipschitz constant.

Combining (i) and (ii), the factor s_n^2 cancels and

$$\frac{\|\Phi_{\mathbf{w},S}(\tilde{x}_n) - \Phi_{\mathbf{w},S}(\tilde{y}_n)\|^2}{\mathbf{d}([\tilde{x}_n], [\tilde{y}_n])^2} = \frac{s_n^2 M_n}{s_n^2 \|a_n - e_n\|^2} = \frac{M_n}{\|a_n - e_n\|^2} \xrightarrow{n \rightarrow \infty} 0,$$

which proves Claim 2.

Since $\tilde{x}_n \rightarrow \sum_{r=1}^k z_r + \alpha$ and $\tilde{y}_n \rightarrow \sum_{r=1}^k z_r + \beta$, while, using that $\mathbf{d}([\tilde{x}_n], [\tilde{y}_n]) = \|s_n a_n - s_n e_n\| \leq \frac{1}{8}$ is bounded (by $\max(\|s_n a_n\|, \|s_n e_n\|) \leq \frac{1}{16}$) and $M_n / \|a_n - e_n\|^2 \rightarrow 0$,

$$\|\Phi_{\mathbf{w},S}(\tilde{x}_n) - \Phi_{\mathbf{w},S}(\tilde{y}_n)\|^2 = s_n^2 M_n = \mathbf{d}([\tilde{x}_n], [\tilde{y}_n])^2 \frac{M_n}{\|a_n - e_n\|^2} \xrightarrow{n \rightarrow \infty} 0,$$

continuity of $\Phi_{\mathbf{w},S}$ gives

$$\Phi_{\mathbf{w},S}(\sum_{r=1}^k z_r + \alpha) - \Phi_{\mathbf{w},S}(\sum_{r=1}^k z_r + \beta) = 0.$$

Since $\hat{\Phi}_{\mathbf{w},S}$ is injective,

$$\sum_{r=1}^k z_r + \alpha \sim \sum_{r=1}^k z_r + \beta.$$

Let $g_1 \in G$ denote a group element that realizes this equivalence, i.e.,

$$\sum_{r=1}^k z_r + \alpha = U_{g_1}(\sum_{r=1}^k z_r + \beta).$$

Note that $g_1 \in H(\sum_{r=1}^k z_r)$ because otherwise

$$\begin{aligned} 0 &= \|\sum_{r=1}^k z_r + \alpha - U_{g_1}(\sum_{r=1}^k z_r) - U_{g_1}\beta\| \geq \|\sum_{r=1}^k z_r - U_{g_1}(\sum_{r=1}^k z_r)\| - \|\alpha - U_{g_1}\beta\| \\ &\geq \rho_0(\sum_{r=1}^k z_r) - \|\alpha\| - \|\beta\| > 0, \end{aligned}$$

where the last inequality comes from the fact that $\|\alpha\| \leq \|\beta\| \leq \frac{1}{16} \rho_0(\sum_{r=1}^k z_r)$.

Additionally, $\alpha = U_{g_1}\beta$ because

$$0 = \left\| \sum_{r=1}^k z_r + \alpha - U_{g_1} \left(\sum_{r=1}^k z_r \right) - U_{g_1} \beta \right\| = \|\alpha - U_{g_1} \beta\|.$$

In particular $\|\alpha\| = \|\beta\| > 0$, so $\alpha \neq 0$.

Claim 3: The two vectors α and β are equal, $\alpha = \beta$.

Indeed, since $g_1 \in H(\sum_{r=1}^k z_r)$, part 2 of Lemma 2.11 and the identity $\mathbf{d}([\tilde{x}_n], [\tilde{y}_n]) = s_n \|a_n - e_n\|$ established in (ii) give, for all n ,

$$\|s_n a_n - s_n e_n\| = \mathbf{d}([\tilde{x}_n], [\tilde{y}_n]) \leq \|s_n a_n - s_n U_{g_1} e_n\|.$$

Letting $n \rightarrow \infty$ and using $\alpha = U_{g_1}\beta$,

$$\|\alpha - \beta\| \leq \lim_{n \rightarrow \infty} \|s_n a_n - s_n U_{g_1} e_n\| = \|\alpha - U_{g_1} \beta\| = 0.$$

We conclude that $\alpha = \beta \neq 0$.

Set $z_{k+1} = \alpha = \beta \neq 0$. We verify the three assertions. (i): each a_n lies in the closed subspace $E_k = \text{span}\{z_1, \dots, z_k\}^\perp$, hence so does $z_{k+1} = \lim_{n \rightarrow \infty} s_n a_n$. (ii): $\|z_{k+1}\| = \|\beta\| = \frac{1}{16} \min \left(\|z_k\|, \Delta(\sum_{r=1}^k z_r), \rho_0(\sum_{r=1}^k z_r) \right)$. (iii): the sequences $\tilde{x}_n$ and $\tilde{y}_n$ converge to $\sum_{r=1}^k z_r + z_{k+1} = \sum_{r=1}^{k+1} z_r$ and, by Claim 2, achieve the vanishing limit. $\square$

Remark 2.15 Our construction produces z_{k+1} with norm equal to

$$\frac{1}{16} \min \left(\|z_k\|, \Delta \left(\sum_{r=1}^k z_r \right), \rho_0 \left(\sum_{r=1}^k z_r \right) \right).$$

In particular $\|z_{k+1}\| \leq \frac{1}{16} \rho_0(\sum_{r=1}^k z_r)$, so the hypothesis $\|z_{l+1}\| \leq \frac{1}{4} \rho_0(\sum_{r=1}^l z_r)$ needed for Remark 2.9 is automatically maintained along the recursion.

Remark 2.16 (The condition $\|z_{l+1}\| \leq \frac{1}{4} \rho_0(\sum_{r \leq l} z_r)$ cannot be dropped) The step $\mathbf{d}([\tilde{x}_n], [\tilde{y}_n]) = s_n \|a_n - e_n\|$ in the proof above uses, through Remark 2.9, that $H(\sum_{r=1}^k z_r)$ fixes each z_r . Without a smallness condition of this type the identity fails: the quotient distance between the rescaled perturbations can be arbitrarily smaller than $s_n \|a_n - e_n\|$, and no lower Lipschitz information can be transferred. We illustrate this with an example.

Let $\mathcal{V} = \mathbb{R}^3$ with standard basis e_1, e_2, e_3 , fix $0 < \epsilon < 1$, and set $z_1 = e_1, z_2 = \epsilon e_2, Z = z_1 + z_2$. Let $G = \{I, R\}$ where R is the rotation by π about the axis $\mathbb{R}Z$; then R

acts orthogonally, $H(Z) = G$, but $U_R z_1 \neq z_1$. Let W be a unit vector orthogonal to Z in $\text{span}\{e_1, e_2\}$, so $U_R W = -W$ and $U_R e_3 = -e_3$. For $\alpha_n \downarrow 0$ put

$$q_n = \sqrt{\alpha_n} W, \quad a_n = \alpha_n e_3, \quad e_n = \alpha_n^2 W - \alpha_n e_3, \\ x_n = Z + q_n + a_n, \quad y_n = Z + q_n + e_n.$$

Then, $a_n = \alpha_n e_3$ and $b_n := -\alpha_n e_3$ both lie in $\text{span}\{z_1, z_2\}^\perp$, the in-plane coefficients converge to 1, and x_n, y_n are aligned as in (16) for n large, because $\|x_n - y_n\| = \alpha_n \sqrt{4 + \alpha_n^2}$ while $\|x_n - U_R y_n\| = 2\sqrt{\alpha_n} + \alpha_n^2$. Nevertheless $U_R z_1 \neq z_1$, and

$$a_n - U_R e_n = \alpha_n^2 W, \quad \text{so} \quad \frac{\min_{g \in H(Z)} \|a_n - U_g e_n\|}{\|a_n - e_n\|} \leq \frac{\alpha_n^2}{\alpha_n \sqrt{4 + \alpha_n^2}} \xrightarrow{n \rightarrow \infty} 0.$$

Here $2\|z_2\| = 2\epsilon > \rho_0(z_1) = 2\epsilon/\sqrt{1 + \epsilon^2}$, i.e., the condition $\|z_2\| \leq \frac{1}{4}\rho_0(z_1)$ fails. It is exactly this configuration, in which an element of the stabilizer of Z moves the plane $\text{span}\{z_1, \dots, z_k\}$, that the hypothesis excludes.

Now, we can complete the proof of Theorem 1.5.

Proof We argue by contradiction. The upper bound was established in Lemma 2.1, so it suffices to show the lower Lipschitz constant is strictly positive. Assume, to the contrary, that it is zero.

Then, by Lemma 2.7, there is a unit vector z_1 and sequences $(x_n)_n, (y_n)_n$ converging to z_1 that achieve lower Lipschitz bound zero. Starting from z_1 and these sequences, after $d - 1$ steps of the algorithmic construction of Lemma 2.14, we obtain d nonzero vectors $\{z_1, \dots, z_d\}$ and a pair of sequences $(\tilde{x}_n)_n, (\tilde{y}_n)_n$ such that

- (i) $\langle z_i, z_j \rangle = 0 \quad \forall i, j \in [d], i \neq j$;
- (ii) $\|z_{r+1}\| \leq \min(\frac{1}{4}\Delta(\sum_{l=1}^r z_l), \frac{1}{4}\|z_r\|, \frac{1}{4}\rho_0(\sum_{l=1}^r z_l)) \quad \forall r \in [d - 1]$; and
- (iii) $\lim_{n \rightarrow \infty} \tilde{x}_n = \lim_{n \rightarrow \infty} \tilde{y}_n = \sum_{r=1}^d z_r$ and

$$\lim_{n \rightarrow \infty} \frac{\|\Phi_{\mathbf{w}, S}(\tilde{x}_n) - \Phi_{\mathbf{w}, S}(\tilde{y}_n)\|^2}{\mathbf{d}([\tilde{x}_n], [\tilde{y}_n])^2} = 0.$$

Let $\tilde{x}_n = \sum_{r=1}^d l_{r,n} z_r$ and $\tilde{y}_n = \sum_{r=1}^d t_{r,n} z_r$. Notice that $\lim_{n \rightarrow \infty} l_{r,n} = \lim_{n \rightarrow \infty} t_{r,n} = 1, \quad \forall r \in [d]$.

Consequently, there is $M_0 \in \mathbb{N}$ such that

$$|1 - l_{r,n}|, |1 - t_{r,n}| < \min\left(\frac{1}{8}, \frac{1}{16d}\Delta\left(\sum_{r=1}^d z_r\right)\right) \leq \delta^{i,j}$$

for all $n \geq M_0, r \in [d]$ and $(i, j) \in S$, so part (2) of Lemma 2.5 applies and gives, for all $n \geq M_0$ and $(i, j) \in S$,

$$L^{i,j}\left(\sum_{r=1}^d l_{r,n} z_r\right) = L^{i,j}\left(\sum_{r=1}^d t_{r,n} z_r\right) = L^{i,j}\left(\sum_{r=1}^d z_r\right).$$

Fix $g_{i,j} \in L^{i,j}(\sum_{r=1}^d z_r)$. Since $z_1, \dots, z_d$ are d mutually orthogonal nonzero vectors in the d -dimensional space $\mathcal{V}$, they form a basis; hence $\tilde{x}_n - \tilde{y}_n = \sum_{r=1}^d (l_{r,n} - t_{r,n})z_r$, and $\mathbf{d}([\tilde{x}_n], [\tilde{y}_n]) \leq \|\tilde{x}_n - \tilde{y}_n\| = \|\sum_{r=1}^d (l_{r,n} - t_{r,n})z_r\|$. Note that $\tilde{x}_n - \tilde{y}_n \neq 0$, since $\tilde{x}_n \approx \tilde{y}_n$. Pass to a subsequence $(n_m)_m$ along which the unit vectors $\frac{\sum_{r=1}^d (l_{r,n} - t_{r,n})z_r}{\|\sum_{r=1}^d (l_{r,n} - t_{r,n})z_r\|}$ converge, and denote their limit, a unit vector, by $\tilde{z}$. Then, using the cancellation of the constant terms exactly as in Lemma 2.14,

$$\begin{aligned} 0 &= \lim_{n \rightarrow \infty} \frac{\|\Phi_{\mathbf{w},S}(\tilde{x}_n) - \Phi_{\mathbf{w},S}(\tilde{y}_n)\|^2}{\mathbf{d}([\tilde{x}_n], [\tilde{y}_n])^2} \\ &= \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^p \sum_{j \in S_i} |\langle U_{g_{i,j}} w_i, \tilde{x}_n - \tilde{y}_n \rangle|^2}{\mathbf{d}([\tilde{x}_n], [\tilde{y}_n])^2} \\ &\geq \lim_{m \rightarrow \infty} \frac{\sum_{i=1}^p \sum_{j \in S_i} |\langle U_{g_{i,j}} w_i, \sum_{r=1}^d (l_{r,n_m} - t_{r,n_m})z_r \rangle|^2}{\|\sum_{r=1}^d (l_{r,n_m} - t_{r,n_m})z_r\|^2} \\ &= \sum_{i=1}^p \sum_{j \in S_i} |\langle U_{g_{i,j}} w_i, \tilde{z} \rangle|^2. \end{aligned}$$

Since the choice of $g_{i,j} \in L^{i,j}(\sum_{r=1}^d z_r)$ was arbitrary (neither the subsequence $(n_m)_m$ nor $\tilde{z}$ depends on it), the computation above shows $\langle U_g w_i, \tilde{z} \rangle = 0$ for every $g \in L^{i,j}(\sum_{r=1}^d z_r)$, $i \in [p]$, $j \in S_i$. We now choose $\epsilon > 0$ with $\sum_{r=1}^d z_r \approx \sum_{r=1}^d z_r + \epsilon \tilde{z}$: for each $g \in G$ the equation $\sum_{r=1}^d z_r + \epsilon \tilde{z} = U_g(\sum_{r=1}^d z_r)$ forces $\epsilon \tilde{z} = U_g(\sum_{r=1}^d z_r) - \sum_{r=1}^d z_r$, which, as $\tilde{z} \neq 0$, determines at most one value of ϵ ; since G is finite, at most N values are excluded, so we may also require $\epsilon \|\tilde{z}\| < \frac{1}{4} \Delta(\sum_{r=1}^d z_r)$. For every $(i, j) \in S$, Lemma 2.3 gives $L^{i,j}(\sum_{r=1}^d z_r + \epsilon \tilde{z}) \subset L^{i,j}(\sum_{r=1}^d z_r)$, so evaluating $\Phi_{i,j}$ at any g in the smaller set yields $\Phi_{i,j}(\sum_{r=1}^d z_r + \epsilon \tilde{z}) = \langle U_g w_i, \sum_{r=1}^d z_r \rangle + \epsilon \langle U_g w_i, \tilde{z} \rangle = \Phi_{i,j}(\sum_{r=1}^d z_r)$. Hence

$$\Phi_{\mathbf{w},S}(\sum_{r=1}^d z_r) = \Phi_{\mathbf{w},S}(\sum_{r=1}^d z_r + \epsilon \tilde{z}),$$

contradicting the injectivity of $\hat{\Phi}_{\mathbf{w},S}$. This contradiction shows that the lower Lipschitz constant is strictly positive, and Theorem 1.5 is proved. $\square$

3 Dimension reduction using linear maps

In the previous section, we considered an embedding

$$\hat{\Phi}_{\mathbf{w},S} : \hat{\mathcal{V}} \rightarrow \mathbb{R}^m$$

that is injective on the quotient space $\hat{\mathcal{V}}$. The dimension m of $\mathbb{R}^m$ may be very large. In this section, we show that the nonlinear map $\hat{\Phi}_{\mathbf{w},S}$ can be further processed through

a linear map into a lower-dimensional space while preserving both injectivity and the bi-Lipschitz property.

The idea of using linear maps for dimensionality reduction has a long history. The well-known Johnson–Lindenstrauss lemma [24] provides a nearly isometric projection for finite metric spaces. More recently, [14] showed that a linear dimension-reduction map preserves the Lipschitz properties of translation-invariant polynomial embeddings.

The first result combining sorted coorbits with linear maps was established in [8] for the symmetric group $G = \Sigma_n$ acting by left multiplication on $n \times d$ real matrices, i.e., $\mathcal{V} = \mathbb{R}^{n \times d}$. In that setting, the dimension of the intermediate coorbit space was $n(1 + (d-1)n!)$, while the final target space was shown to have dimension $2nd = 2 \dim(\mathcal{V})$. The authors of [20] demonstrated that the intermediate embedding can instead be realized in dimension $(2nd+1)n$, with the final target space of dimension $2nd+1$.

A more refined analysis by Matthias Wellershoff [37] showed that $2nd+1$ can be further reduced to $2nd-d$. In the Ph.D. thesis [32], one of the authors explained that this reduction from $2nd$ to $2nd-d$ corresponds to the dimension of a certain joint eigenspace of the representation.

The precise statement of our dimension reduction result is provided in Theorem 1.7.

To prove Theorem 1.7, we first need to construct several preliminary objects.

Let $D : \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}^m$ be the nonlinear map $D(x, y) = \Phi_{\mathbf{w},S}(x) - \Phi_{\mathbf{w},S}(y)$. Its range E is defined by $E = \text{Ran}(D) := \{\Phi_{\mathbf{w},S}(x) - \Phi_{\mathbf{w},S}(y) \mid x, y \in \mathcal{V}\} = \text{Ran}(\Phi_{\mathbf{w},S}) - \text{Ran}(\Phi_{\mathbf{w},S})$.

Fix an enumeration $g_1, \dots, g_N$ of the group elements. Define $\lambda_{i,j} : \mathcal{V} \rightarrow \mathbb{R}$ by $\lambda_{i,j}(x) = \langle U_{g_i} w_j, x \rangle$, so that the first index $i \in [N]$ is a group index and the second index $j \in [p]$ is a window index. Notice that $\lambda_{i,j}$ is a linear map (unlike $\Phi_{i,j}$) and also that

$$\begin{aligned} & \Phi_{\mathbf{w},S}(x) - \Phi_{\mathbf{w},S}(y) \\ &= [\lambda_{v_1(1),1}(x) - \lambda_{v_{p+1}(1),1}(y), \dots, \lambda_{v_1(m_1),1}(x) - \lambda_{v_{p+1}(m_1),1}(y), \\ & \quad \dots, \lambda_{v_p(1),p}(x) - \lambda_{v_{2p}(1),p}(y), \dots, \lambda_{v_p(m_p),p}(x) - \lambda_{v_{2p}(m_p),p}(y)] \end{aligned}$$

for some permutations $v_1, \dots, v_{2p} \in \Sigma_N$ that may depend on x and y . Here $m_j = |S_j| = |\{k \in [N] \mid (k, j) \in S\}|$ so that $m_1 + \dots + m_p = m$.

Now, fix permutations $\pi_1, \dots, \pi_{2p} \in \Sigma_N$ and let $L_{\pi_1, \dots, \pi_{2p}} : \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}^m$ denote the linear map

$$\begin{aligned} L_{\pi_1, \dots, \pi_{2p}}(x, y) = & [\lambda_{\pi_1(1),1}(x) - \lambda_{\pi_{p+1}(1),1}(y), \dots, \lambda_{\pi_1(m_1),1}(x) - \lambda_{\pi_{p+1}(m_1),1}(y), \\ & \dots, \lambda_{\pi_p(1),p}(x) - \lambda_{\pi_{2p}(1),p}(y), \dots, \lambda_{\pi_p(m_p),p}(x) - \lambda_{\pi_{2p}(m_p),p}(y)] \end{aligned}$$

Define

$$F = \cup_{\pi_1, \dots, \pi_{2p} \in \Sigma_N} \text{Ran}(L_{\pi_1, \dots, \pi_{2p}}).$$

Notice that F is a finite union of linear subspaces and that $E \subset F$. For fixed $\pi_1, \dots, \pi_{2p}$, the map $(x, y) \mapsto L_{\pi_1, \dots, \pi_{2p}}(x, y)$ is linear in (x, y) , $L_{\pi_1, \dots, \pi_{2p}}(v, v) = 0$

for all $v \in V_G$, and from the rank-nullity theorem, we have

$$\dim(\text{Ran}(L_{\pi_1, \dots, \pi_{2p}})) \leq 2d - d_G.$$

We will use the following standard transversality fact from algebraic geometry (see, e.g., [23]), which we state in the form convenient for our setup. Let r, s, m be non-negative integers with $r + s \leq m$, and let $\{F_a \mid a \in [T]\}$ be a finite collection of linear subspaces of $\mathbb{R}^m$ of dimension at most s . Then, a generic r -dimensional linear subspace K of $\mathbb{R}^m$ satisfies

$$K \cap F_a = \{0\}, \quad \forall a \in [T]. \quad (20)$$

Here “generic” means holding on an open dense subset in the Zariski topology on the Grassmannian of r -planes. Indeed, the union $\cup_a F_a$ is an algebraic variety of dimension at most s in $\mathbb{R}^m$, i.e. a projective variety of dimension at most $s - 1$ in $\mathbb{P}^{m-1}$; a generic linear subvariety of $\mathbb{P}^{m-1}$ of dimension $r - 1$ misses it whenever $(s - 1) + (r - 1) < m - 1$, that is, $s + r \leq m$.

We now apply this fact to our setup.

Corollary 3.1 *Consider the $(N!)^{2p}$ linear maps $L_{\pi_1, \dots, \pi_{2p}} : \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}^m$ introduced before. Then, for a generic $\ell : \mathbb{R}^m \rightarrow \mathbb{R}^{2d-d_G}$,*

$$\ker(\ell) \cap \bigcup_{\pi_1, \dots, \pi_{2p} \in \Sigma_N} \text{Ran}(L_{\pi_1, \dots, \pi_{2p}}) = \{0\}.$$

Proof If $m \leq 2d - d_G$, then the conclusion is satisfied for any full-rank ℓ . Therefore assume $m > 2d - d_G$. A generic linear map $\ell : \mathbb{R}^m \rightarrow \mathbb{R}^{2d-d_G}$ is full-rank. Hence $\dim(\text{Ran}(\ell)) = 2d - d_G$, and thus $\dim(\ker(\ell)) = m - 2d + d_G$. Recall from above that each $\text{Ran}(L_{\pi_1, \dots, \pi_{2p}})$ has dimension at most $2d - d_G$. Applying the transversality fact (20) with $K = \ker(\ell)$, $r = m - 2d + d_G$, $s = 2d - d_G$, $T = (N!)^{2p}$, and $\{F_a\} = \{\text{Ran}(L_{\pi_1, \dots, \pi_{2p}})\}$ (so that $r + s = m$), for a generic ℓ , we obtain

$$\ker(\ell) \cap \text{Ran}(L_{\pi_1, \dots, \pi_{2p}}) = \{0\}$$

for every $\pi_1, \dots, \pi_{2p} \in \Sigma_N$. $\square$

Lemma 3.2 *Let $\{F_a\}_{a=1}^T$ be a finite collection of linear subspaces of $\mathbb{R}^m$, and $\ell : \mathbb{R}^m \rightarrow \mathbb{R}^s$ be a full-rank linear transformation with $m \geq s$. Let Q_a denote the orthogonal projection onto the linear space F_a and Q_ℓ denote the orthogonal projection onto $\ker \ell$. Let $c_{a,\ell} = (1 - \|Q_a Q_\ell\|^2)^{1/2}$, and $c_\ell = \min_{a \in [T]} c_{a,\ell}$. Here $\|Q_a Q_\ell\|$ denotes the operator norm of $Q_a Q_\ell$, i.e., its largest singular value. Set $F = \cup_{a=1}^T F_a$. Suppose that $\ker(\ell) \cap F = \{0\}$. Note that then $\|Q_a Q_\ell\| < 1$ for every $a \in [T]$, since $F_a \cap \ker(\ell) = \{0\}$ and both subspaces are finite dimensional; in particular $c_\ell > 0$. Then,*

$$\inf_{\substack{x \in F \\ \|x\|=1}} \|\ell(x)\| \geq c_\ell \sigma_s(\ell), \quad (21)$$

where $\sigma_s(\ell)$ is the smallest strictly positive singular value of ℓ (it is the s -th singular value).

Proof Notice that for each $a \in [T]$, the unit sphere of F_a is a compact set. Thus,

$$\inf_{\substack{x \in F \\ \|x\|=1}} \|\ell(x)\| = \min_{\substack{x \in F \\ \|x\|=1}} \|\ell(x)\| = \|\ell(y_\infty)\|$$

for some $a \in [T]$ and some $y_\infty \in F_a$ with $\|y_\infty\| = 1$. Let $y_\infty = \sum_{k=1}^m \gamma_k u_k$, where u_k are the normalized right singular vectors of ℓ sorted by singular values $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_s > \sigma_{s+1} = \dots = \sigma_m = 0$. Notice that $\sum_{k=1}^m \gamma_k^2 = 1$ and $\sum_{k=1}^s \gamma_k^2 = 1 - \|Q_\ell y_\infty\|^2 \geq 1 - \|Q_a Q_\ell\|^2 \geq c_\ell^2$. Thus,

$$\begin{aligned} \|\ell(y_\infty)\|^2 &= \left\| \sum_{k=1}^m \gamma_k \ell(u_k) \right\|^2 = \left\| \sum_{k=1}^s \gamma_k \ell(u_k) \right\|^2 \\ &= \sum_{k=1}^s \gamma_k^2 \sigma_k^2 \geq c_{a,\ell}^2 \sigma_s(\ell)^2 \geq c_\ell^2 \sigma_s(\ell)^2 \end{aligned}$$

which proves the lemma. $\square$

Proof of Theorem 1.7 If $m < 2d - d_G$, then a generic linear map $\ell : \mathbb{R}^m \rightarrow \mathbb{R}^{2d-d_G}$ is injective on all of $\mathbb{R}^m$, and both conclusions follow at once, with lower Lipschitz constant $\sigma_m(\ell)a$; so assume $m \geq 2d - d_G$.

Corollary 3.1 shows that a generic linear map $\ell : \mathbb{R}^m \rightarrow \mathbb{R}^{2d-d_G}$ satisfies $\ker(\ell) \cap \text{Ran}(D) = \{0\}$. Thus, if $x, y \in \mathcal{V}$ are such that $\Psi_{\mathbf{w},S,\ell}(x) = \Psi_{\mathbf{w},S,\ell}(y)$, then $\ell(D(x, y)) = 0$. Therefore $D(x, y) = 0$. Since $\hat{\Phi}_{\mathbf{w},S}$ is injective, it follows that $x \sim y$. Thus, $\hat{\Psi}_{\mathbf{w},S,\ell}$ is also injective.

From Theorem 1.5, we have that if the map $\hat{\Phi}_{\mathbf{w},S}$ is injective, then it is also bi-Lipschitz. Let $a \leq b$ denote its bi-Lipschitz constants.

The composition of two Lipschitz maps is Lipschitz, hence $\Psi_{\mathbf{w},S,\ell}$ is Lipschitz. Furthermore, an upper Lipschitz constant of $\Psi_{\mathbf{w},S,\ell}$ is $\|\ell\|b$, where $\|\ell\| = \sigma_1(\ell)$ is the largest singular value of ℓ .

Finally, from Corollary 3.1 and Lemma 3.2 with $s = 2d - d_G$, $T = (N!)^{2p}$, $F_a = \text{Ran}(L_{\pi_1, \dots, \pi_{2p}})$, we have that for a generic linear map ℓ , for all $x, y \in \mathcal{V}$,

$$\begin{aligned} \|\Psi_{\mathbf{w},S,\ell}(x) - \Psi_{\mathbf{w},S,\ell}(y)\| &= \|\ell(D(x, y))\| \geq \\ &c_\ell \sigma_{2d-d_G}(\ell) \|D(x, y)\| \geq c_\ell \sigma_{2d-d_G}(\ell) a \mathbf{d}([x], [y]) \end{aligned}$$

where a is the lower Lipschitz constant of $\hat{\Phi}_{\mathbf{w},S}$. Therefore the map $\hat{\Psi}_{\mathbf{w},S,\ell}$ is bi-Lipschitz with lower Lipschitz constant $c_\ell \sigma_{2d-d_G}(\ell)a$. $\square$

Remark 3.3 We proved that if $\hat{\Phi}_{\mathbf{w},S}$ is injective then for almost any linear map $\ell : \mathbb{R}^m \rightarrow \mathbb{R}^{2d-d_G}$, $\hat{\Psi}_{\mathbf{w},S,\ell}$ is bi-Lipschitz. This left open the question whether, for an arbitrary such nonlinear embedding $\Psi_{\mathbf{w},S,\ell}$, injectivity implies bi-Lipschitz stability.

We settle it positively in our joint paper with Matthias Wellershoff [11]. Notice that, in general, if the map $f : X \rightarrow Y$ is bi-Lipschitz and the linear map $\ell : Y \rightarrow \mathbb{R}^q$ is so that $\ell \circ f$ is injective, then $\ell \circ f$ may not be bi-Lipschitz. Example: $f : [-1, 1] \rightarrow \mathbb{R}^2$, $f(t) = (t, t^3)$, $\ell : \mathbb{R}^2 \rightarrow \mathbb{R}$, $\ell(x, y) = y$; here f is bi-Lipschitz on the bounded interval, while $\ell \circ f(t) = t^3$ is injective and fails the lower Lipschitz bound at $t = 0$. (On all of $\mathbb{R}$ the map $t \mapsto (t, t^3)$ is not even Lipschitz.)

4 Universality of the representation

In the previous sections, we constructed the embedding

$$\hat{\Phi}_{\mathbf{w},S} : \hat{\mathcal{V}} \rightarrow \mathbb{R}^m$$

of the quotient space $\hat{\mathcal{V}}$ into a Euclidean space. In this section, we prove Theorems 1.8 and 1.9, which show that whenever $\hat{\Phi}_{\mathbf{w},S}$ is injective, every map defined on $\hat{\mathcal{V}}$ factors through $\hat{\Phi}_{\mathbf{w},S}$: in Theorem 1.8 for Lipschitz maps into a real Hilbert space, and in Theorem 1.9 for continuous maps into a locally convex topological vector space.

The proof of Theorem 1.8 relies on the Kirszbraun extension theorem, in the Hilbert-space form due to Valentine [25, 34], which we restate here for the reader's convenience:

Theorem 4.1 (Kirszbraun–Valentine extension theorem [25, 34]) *Let $E \subset H_1$ be an arbitrary subset of a Hilbert space H_1 and $f : E \rightarrow H_2$ be a Lipschitz function to another Hilbert space H_2 . Then, there exists an extension $F : H_1 \rightarrow H_2$ of f to the entire space H_1 that has the same Lipschitz constant as the original function f .*

Proof of Theorem 1.8

1. Let $t : \Phi_{\mathbf{w},S}(\mathcal{V}) \rightarrow H$ be defined by $t(\Phi_{\mathbf{w},S}(x)) = F([x])$. This is well defined: if $\Phi_{\mathbf{w},S}(x) = \Phi_{\mathbf{w},S}(y)$ then $a \mathbf{d}([x], [y]) \leq 0$, so $[x] = [y]$. Denote $u = \Phi_{\mathbf{w},S}(x)$ and $v = \Phi_{\mathbf{w},S}(y)$. Then,

$$\begin{aligned} \|t(u) - t(v)\| &= \|F([x]) - F([y])\| \leq \text{Lip}(F) \mathbf{d}([x], [y]) \\ &\leq \frac{1}{a} \text{Lip}(F) \|u - v\|. \end{aligned}$$

By the Kirszbraun–Valentine extension theorem, there exists $T : \mathbb{R}^m \rightarrow H$ such that

- (a) $T|_{\Phi_{\mathbf{w},S}(\mathcal{V})} = t$
- (b) $\text{Lip}(T) = \text{Lip}(t)$

Therefore, $F([x]) = T(\Phi_{\mathbf{w},S}(x))$ for all $x \in \mathcal{V}$, and $\text{Lip}(T) \leq \frac{1}{a} \text{Lip}(F)$.

2. Part 2 is straightforward. Let $x, y \in \mathcal{V}$. Then,

$$\begin{aligned}\|F([x]) - F([y])\| &= \|T(\Phi_{\mathbf{w},S}(x)) - T(\Phi_{\mathbf{w},S}(y))\| \\ &\leq \text{Lip}(T) \|\Phi_{\mathbf{w},S}(x) - \Phi_{\mathbf{w},S}(y)\| \\ &\leq b \text{Lip}(T) \mathbf{d}([x], [y]),\end{aligned}$$

where the last inequality uses that b is the upper Lipschitz constant of $\hat{\Phi}_{\mathbf{w},S}$. Hence $\text{Lip}(F) \leq b \text{Lip}(T)$. $\square$

The second universality result, Theorem 1.9, applies to the class of continuous functions rather than Lipschitz functions.

The proof of Theorem 1.9 follows from the following extension of Tietze's theorem [19, 33].

Theorem 4.2 (Dugundji–Tietze [19]) *Let X be a metric space and A a closed subset of X . Let L be a locally convex topological vector space. Given a continuous map $f : A \rightarrow L$, there exists a continuous extension $F : X \rightarrow L$ such that $F(X)$ is a subset of the convex hull of $f(A)$.*

Proof of Theorem 1.9 Let $Y = \Phi_{\mathbf{w},S}(\mathcal{V}) \subset \mathbb{R}^m$. Note that Y is a closed set as $\hat{\Phi}_{\mathbf{w},S}$ is bi-Lipschitz and $\hat{\mathcal{V}}$ is complete. Let $t : Y \rightarrow L$ be defined by $t(\Phi_{\mathbf{w},S}(x)) = F([x])$, for all $x \in \mathcal{V}$. Note t is continuous since $\hat{\Phi}_{\mathbf{w},S}$ is bi-Lipschitz on $\hat{\mathcal{V}}$. By Theorem 4.2, applied with $X = \mathbb{R}^m$ and the closed subset $A = Y$, there exists a continuous extension $T : \mathbb{R}^m \rightarrow L$ of t that satisfies the convex hull property. $\square$

Author contribution Radu Balan: supervision, conceptualization, methodology, validation, writing. Efstratios Tsoukanis: methodology, validation, writing. Radu Balan and Efstratios Tsoukanis read and approved the final manuscript.

Funding Open access funding provided by SCEL, Statewide California Electronic Library Consortium. The first author has been supported in part by the National Science Foundation under grant DMS-2510216.

Data availability Data sharing is not applicable to this article, as no datasets were generated or analysed during the current study.

Declarations

Conflict of interest Both authors, Radu Balan and Efstratios Tsoukanis, declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Alaifari, R., Daubechies, I., Grohs, P., Yin, R.: Stable phase retrieval in infinite dimensions. *Found. Comput. Math.* **19**(4), 869–900 (2019)
2. Alaifari, R., Wellershoff, M.: Stability estimates for phase retrieval from discrete Gabor measurements. *J. Fourier Anal. Appl.* **27**(2), 6 (2021)
3. Amir, T., Bendory, T., Dym, N., Edidin, D.: The stability of generalized phase retrieval problem over compact groups (2025). arXiv preprint [arXiv:2505.04190](https://arxiv.org/abs/2505.04190)
4. Aslan, B., Platt, D., Sheard, D.: Group invariant machine learning by fundamental domain projections. In: *NeurIPS Workshop on Symmetry and Geometry in Neural Representations*, pp. 181–218. PMLR (2023)
5. Balan, R., Wang, Y.: Invertibility and robustness of phaseless reconstruction. *Appl. Comput. Harmon. Anal.* **38**(3), 469–488 (2015)
6. Balan, R.: Frames and phaseless reconstruction. In: Okoudjou, K.A. (ed.) *Finite Frame Theory: A Complete Introduction to Overcompleteness*. Proceedings of Symposia in Applied Mathematics, vol. 73, pp. 175–199. American Mathematical Society (2016)
7. Balan, R., Casazza, P., Edidin, D.: On signal reconstruction without phase. *Appl. Comput. Harmon. Anal.* **20**(3), 345–356 (2006)
8. Balan, R., Haghani, N., Singh, M.: Permutation-invariant representations with applications to graph deep learning. *Appl. Comput. Harmon. Anal.*, p. 101798 (2025)
9. Balan, R., Tsoukanis, E.: G-invariant representations using coorbits: injectivity properties (2023). arXiv preprint [arXiv:2310.16365](https://arxiv.org/abs/2310.16365)
10. Balan, R., Tsoukanis, E.: Relationships between the phase retrieval problem and permutation invariant embeddings. In: *2023 International Conference on Sampling Theory and Applications (SampTA)*, pp. 1–6 (2023)
11. Balan, R., Tsoukanis, E., Wellershoff, M.: Stability of sorting based embeddings (2024). arXiv preprint [arXiv:2410.05446](https://arxiv.org/abs/2410.05446)
12. Blum-Smith, B., Villar, S.: Equivariant maps from invariant functions (2022). arXiv preprint [arXiv:2209.14991](https://arxiv.org/abs/2209.14991)
13. Bronstein, M.M., Bruna, J., LeCun, Y., Szlam, A., Vandergheynst, P.: Geometric deep learning: going beyond Euclidean data. *IEEE Signal Process. Mag.* **34**(4), 18–42 (2017)
14. Cahill, J., Contreras, A., Contreras-Hip, A.A.: Complete set of translation invariant measurements with Lipschitz bounds. *Appl. Comput. Harmon. Anal.* **49**(2), 521–539 (2020)
15. Cahill, J., Iverson, J.W., Mixon, D.G.: Towards a bi-Lipschitz invariant theory. *Appl. Comput. Harmon. Anal.* (2024)
16. Cahill, J., Iverson, J.W., Mixon, D.G., Packer, D.: Group-invariant max filtering. *Found. Comput. Math.*, pp. 1–35 (2024)
17. Derksen, H.: Bi-Lipschitz quotient embedding for Euclidean group actions on data (2024). arXiv preprint [arXiv:2409.06829](https://arxiv.org/abs/2409.06829)
18. Dufresne, E.: Separating invariants and finite reflection groups. *Adv. Math.* **221**(6), 1979–1989 (2009)
19. Dugundji, J.: An extension of Tietze’s theorem. *Pac. J. Math.* **1**(3), 353–367 (1951)
20. Dym, N., Gortler, S.J.: Low dimensional invariant embeddings for universal geometric learning. *Found. Comput. Math.* (2024)
21. Eriksson-Bique, S.: Quantitative bi-Lipschitz embeddings of bounded-curvature manifolds and orbifolds. *Geom. Topol.* **22**(4), 1961–2026 (2018)
22. Freeman, D., Oikhberg, T., Pineau, B., Taylor, M.A.: Stable phase retrieval in function spaces. *Math. Ann.* **390**(1), 1–43 (2024)
23. Harris, J.: *Algebraic Geometry: A First Course*. Graduate Texts in Mathematics, vol. 133. Springer-Verlag, New York (1992)
24. Johnson, W.B., Lindenstrauss, J., Schechtman, G.: Extensions of Lipschitz maps into Banach spaces. *Israel J. Math.* **54**(2), 129–138 (1986)
25. Kirschbraun, M.: Über die zusammenziehende und Lipschitzsche Transformationen. *Fundam. Math.* **22**(1), 77–108 (1934)
26. Maron, H., Ben-Hamu, H., Shamir, N., Lipman, Y.: Invariant and equivariant graph networks. In: *International Conference on Learning Representations* (2019)
27. Mixon, D.G., Packer, D.: Max filtering with reflection groups. *Adv. Comput. Math.* **49**(6), 82 (2023)

28. Mixon, D.G., Qaddura, Y.: Injectivity, stability, and positive definiteness of max filtering. *Constr. Approx.*, pp. 1–38 (2025)
29. Puny, O., Atzmon, M., Smith, E.J., Mishra, I., Grover, A., Ben-Hamu, H., Lipman, Y.: Frame averaging for invariant and equivariant network design. In: *International Conference on Learning Representations* (2022)
30. Qaddura, Y., Mixon, D.G.: Stable coorbit embeddings of orbifold quotients (2024). arXiv preprint [arXiv:2403.14042v1](https://arxiv.org/abs/2403.14042v1)
31. Sannai, A., Takai, Y., Cordonnier, M.: Universal approximations of permutation invariant/equivariant functions by deep neural networks (2020)
32. Tsoukanis, E.: *G-Invariant Representations Using Coorbits*. PhD thesis, University of Maryland, College Park (2024)
33. Urysohn, P.: Über die Mächtigkeit der zusammenhängenden Mengen. *Math. Ann.* **94**(1), 262–295 (1925)
34. Valentine, F.A.: A Lipschitz condition preserving extension for a vector function. *Am. J. Math.* **67**(1), 83–93 (1945)
35. Villar, S., Hogg, D.W., Storey-Fisher, K., Yao, W., Blum-Smith, B.: Scalars are universal: equivariant machine learning, structured like classical physics. *Adv. Neural Inf. Process. Syst.* **34**, 28848–28863 (2021)
36. Villar, S., Yao, W., Hogg, D.W., Blum-Smith, B., Dumitrascu, B.: Dimensionless machine learning: imposing exact units equivariance. *J. Mach. Learn. Res.* **24**(109), 1–32 (2023)
37. Wellershoff, M.: Personal Communication (2023)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.