Approximate Grassmannian Intersections: Subspace-Valued Subspace Learning

Calvin Murdock (cmurdock@cs.cmu.edu), Fernando De la Torre (ftorre@cs.cmu.edu)
Carnegie Mellon University, Machine Learning Department

Introduction

Applications of subspace learning in computer vision are ubiquitous, ranging from **dimensionality reduction** to **denoising**. As geometric objects, subspaces have also been successfully used for **efficiently** representing **invariant** data such as **faces**. However, due to their **nonlinear geometric structure**, subspace-valued data are generally **incompatible** with most standard machine learning techniques.

To address this issue, we propose **Approximate Grassmannian Intersections (AGI)**, a novel geometric interpretation of subspace learning posed as finding the **approximate intersection** of constraint sets on the **Grassmann manifold**. AGI can naturally be applied to subspaces of **varying dimension** and enables new analyses by embedding them in a **shared** low-dimensional Euclidean space .

Grassmannian Geometry Preliminaries

The **Grassmannian** $\mathcal{G}_{k,d}$ is a manifold that parametrizes the space of all k-dimensional subspaces of $\mathbb{R}^d$. Each point in $\mathcal{G}_{k,d}$ corresponds to a single subspace that is **invariant** to a particular choice of basis:

$$\mathcal{G}_{k,d} = \{\mathrm{col}(\mathbf{B}) : \mathbf{B} \in \mathcal{V}_{k,d}\}, \quad \mathcal{V}_{k,d} = \{\mathbf{B} : \mathbf{B}^\mathsf{T}\mathbf{B} = \mathbf{I}\}$$

Since it is not a vector space, inner products between elements are not well-defined. However, as a **metric space**, $\mathcal{G}_{k,d}$ does allow for the computation of **distances**. The natural geodesic $d_\mathcal{G}$ can be expressed as the norm of the vector of **principal angles** between subspaces, which is found using a singular value decomposition:

$$d_\mathcal{G}(A,B) = \|\boldsymbol{\theta}\|_2, \quad \mathbf{A}^\mathsf{T}\mathbf{B} = \mathbf{U}\mathrm{diag}(\cos\boldsymbol{\theta})\mathbf{V}^\mathsf{T}$$

In this work, we use the projection F-norm distance d_P, which is similarly expressed in terms of the sine of the principal angles. While close to the geodesic distance for small angles, d_P can be computed **more efficiently** via bijective, isometric embeddings $\Pi(\cdot)$ that associate each subspace with its unique **projection** matrix:

$$d_P(A,B) = \|\sin\boldsymbol{\theta}\|_2 = 2^{-\frac{1}{2}}\|\Pi(A) - \Pi(B)\|_F$$

$$\Pi : \mathcal{G}_{k,d} \to \mathbb{R}^{d\times d}, \quad \Pi(B) = \mathbf{B}\mathbf{B}^\mathsf{T}$$

We represent subspaces as elements of the set of all projectors:

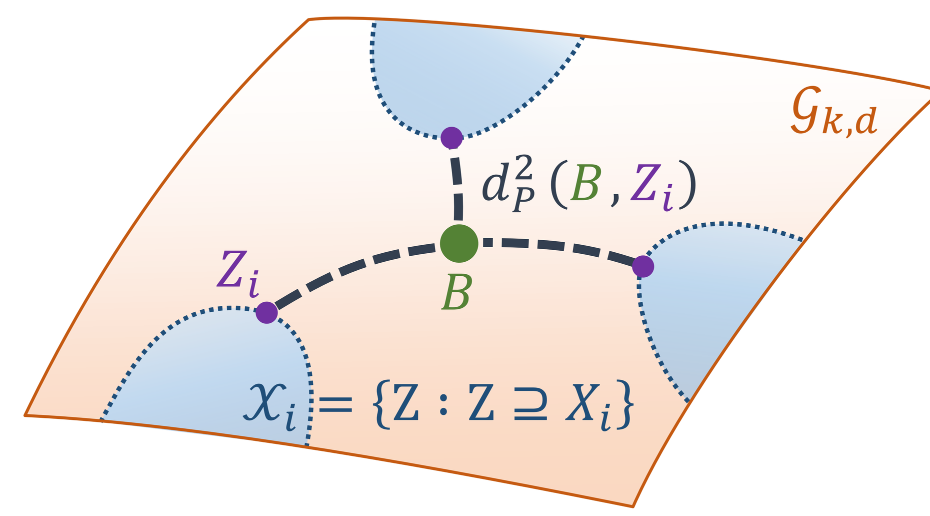
$$\mathcal{P}_{k,d} = \{\mathbf{P} \in \mathbb{R}^{d\times d} : \mathbf{P}^\mathsf{T} = \mathbf{P}, \mathbf{P}^2 = \mathbf{P}, \mathrm{tr}(\mathbf{P}) = k\}$$

Since its elements all have fixed rank, this set is **non-convex**, so for theoretical guarantees we also consider its convex hull, the **Fantope**:

$$\mathcal{F}_{k,d} = \mathrm{conv}(\mathcal{P}_{k,d}) = \{\mathbf{Q} \in \mathbb{R}^{d\times d} : \mathbf{0} \preceq \mathbf{Q} \preceq \mathbf{I}, \mathrm{tr}(\mathbf{Q}) = k\}$$

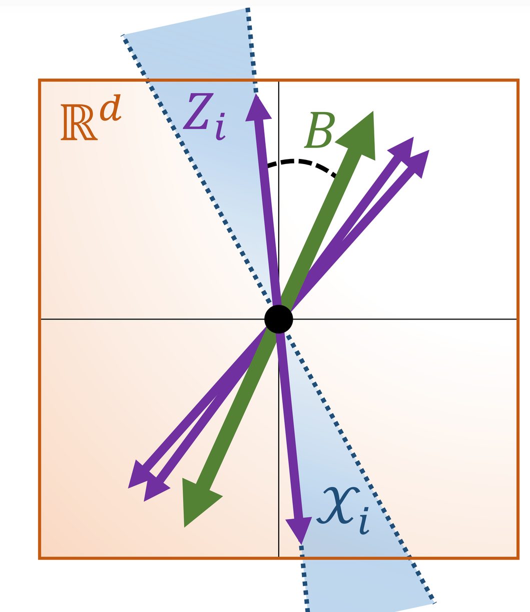
Motivation

Like standard subspace learning techniques, our goal is to learn a low-dimensional subspace that best **approximates** a given dataset. Instead of **reconstruction error**, we quantify this by the **average proximity** to the nearest subspaces containing the data. This **generalizes** previous approaches to support **subspace-valued data**.



Grassmannian View

⟺



Euclidean View

Problem Formulation

From a set of p_i-dimensional subspaces $X_i \in \mathcal{G}_{p_i,d}$, we aim to learn a k-dimensional subspace $B \in \mathcal{G}_{k,d}$ with $p_i \le k < d$. For each X_i, we introduce an **auxiliary local subspace** $Z_i \in \mathcal{G}_{k,d}$ that is constrained to contain it, i.e. $Z_i \supseteq X_i$. Our goal is to find the subspace B that is closest to them in terms of **average squared distance**:

$$\underset{B, Z_i \in \mathcal{G}_{k,d}}{\arg\min} \sum_{i=1}^{n} d_P^2(B, Z_i) \text{ s.t. } Z_i \supseteq X_i$$

Representing the learned subspaces as projection matrices, this is **equivalent** to the following objective, where $\mathbf{X}_i$ are orthonormal matrices with columns spanning X_i. Note that the subspace inclusion constraint can now be written as an **affine equality** constraint.

$$\underset{\mathbf{P}, \mathbf{Q}_i \in \mathcal{P}_{k,d}}{\arg\min} \sum_{i=1}^{n} \|\mathbf{P} - \mathbf{Q}_i\|_F^2 \text{ s.t. } \mathbf{Q}_i\mathbf{X}_i = \mathbf{X}_i$$

Global Optimality

Despite its **non-convexity**, this problem admits an efficient, **globally optimal** solution given by the top k left singular vectors of the matrix formed by concatenating all $\mathbf{X}_i$, as demonstrated by the following observation connecting AGI with standard PCA.

$$\min_{\mathbf{Q}_i \in \mathcal{P}_{k,d}} \left[\frac{1}{2}\|\mathbf{P} - \mathbf{Q}_i\|_F^2 \text{ s.t. } \mathbf{Q}_i\mathbf{X}_i = \mathbf{X}_i\right] = \|\mathbf{X}_i - \mathbf{P}\mathbf{X}_i\|_F^2$$

Inference and Subspace Completion

The solution subspace associated with $\mathbf{P}$ can be applied towards a variety of applications through **inference** of lower-dimensional **latent variables**. Analogous to matrix completion, we first employ **subspace completion** to infer missing dimensions of the data subspaces for consistency. Given a projection matrix $\hat{\mathbf{P}}$ formed from the top $m \le k$ dimensions of $\mathbf{P}$, the completed subspace spans the columns of $\hat{\mathbf{X}}_i = [\mathbf{X}_i, \bar{\mathbf{X}}_i]$ where $\bar{\mathbf{X}}_i$ contains the top $m - p_i$ eigenvectors of $(\mathbf{I} - \mathbf{X}_i\mathbf{X}_i^\mathsf{T})\hat{\mathbf{P}}$ as its columns. Finally, a lower dimensional subspace associated with the projection matrix $\mathbf{M}_i \in \mathcal{P}_{m,k}$ can be found as:

$$\mathbf{M}_i = \underset{\mathbf{M}_i \in \mathcal{P}_{m,k}}{\arg\min} \left\|\hat{\mathbf{X}}_i\hat{\mathbf{X}}_i^\mathsf{T} - \mathbf{B}\mathbf{M}_i\mathbf{B}^\mathsf{T}\right\|_F^2 = \hat{\mathbf{X}}_i^\mathsf{T}\mathbf{P}\hat{\mathbf{X}}_i$$

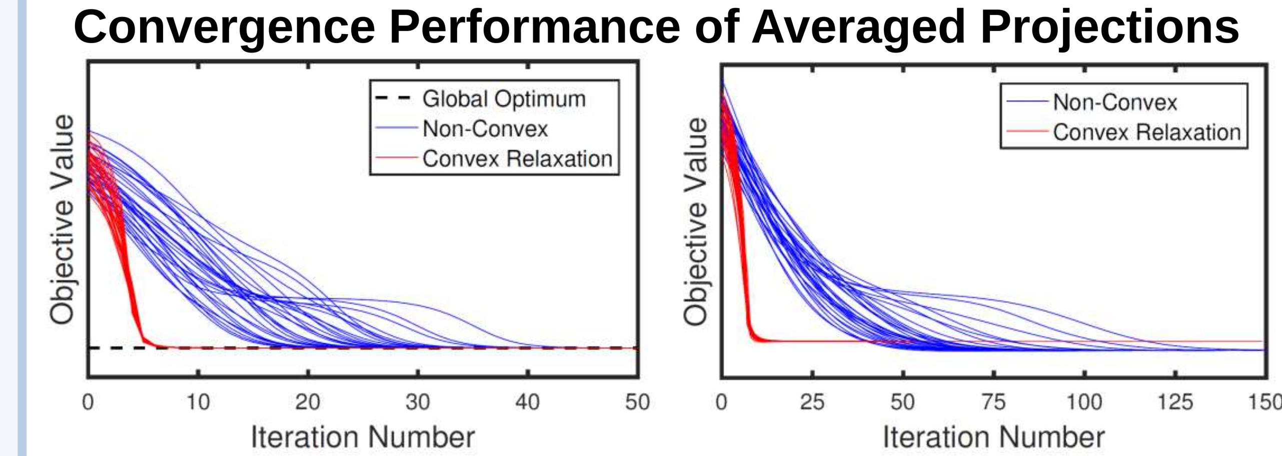
Extensions

AGI can also accommodate additional **prior knowledge** encoded as other **constraints** $\mathcal{C}_i$ on the local subspaces Z_i. In general, our optimization problem is equivalent to finding the **approximate intersection of sets** and can be solved using an iterative projection algorithm such as the method of **averaged projections**. Though empirically successful even with non-convex sets, this algorithm is theoretically **guaranteed** to find the **unique global minimum** if the sets are **convex**, motivating a relaxation using the Fantope $\mathcal{F}_{k,d}$.

Robust AGI: $\mathcal{C}_i^r = \left\{\mathbf{Q} : \mathbf{Q}(\mathbf{X}_i + \mathbf{E}_i) = \mathbf{X}_i + \mathbf{E}_i, \|\mathbf{E}_i^\mathsf{T}\|_{2,1} \le \varepsilon\right\}$

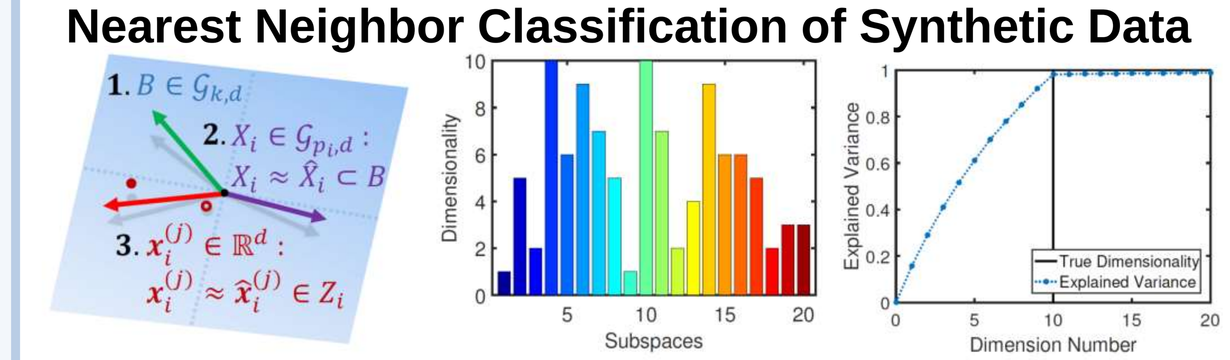
Experimental Results

Convergence Performance of Averaged Projections

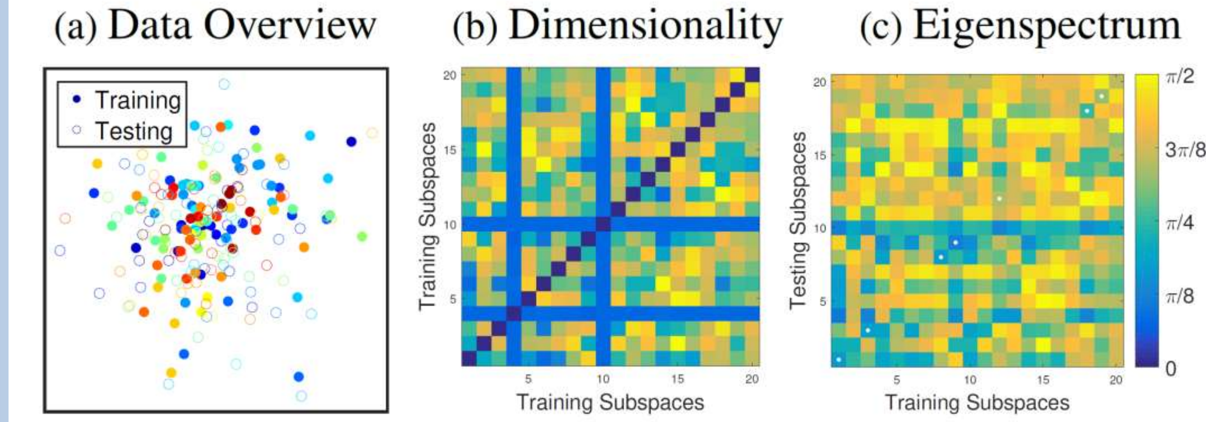


AGI — Robust AGI

Nearest Neighbor Classification of Synthetic Data



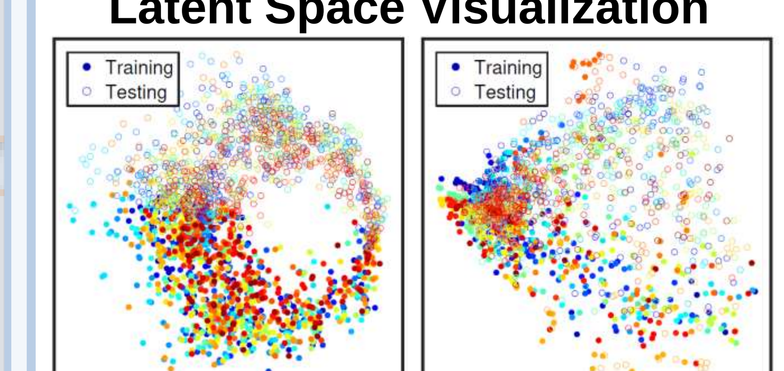
(a) Data Overview (b) Dimensionality (c) Eigenspectrum



(a) Embedding (b) Training Angles (c) Testing Angles

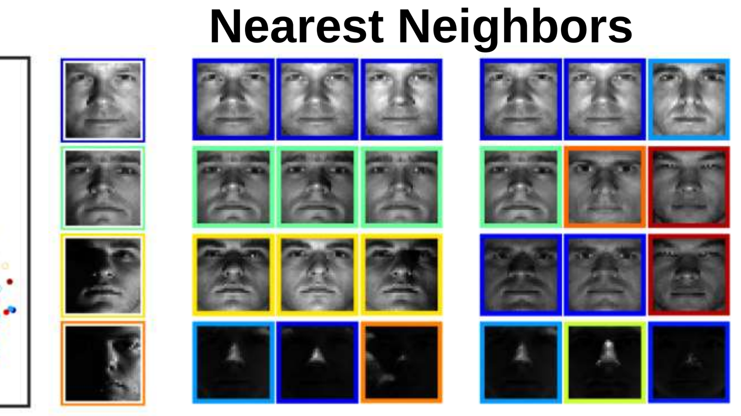
PCA	Subspace Angle	AGI (Vector)	AGI (Subspace)
14.6%	15.5%	96.1%	100%

Latent Space Visualization



AGI — PCA

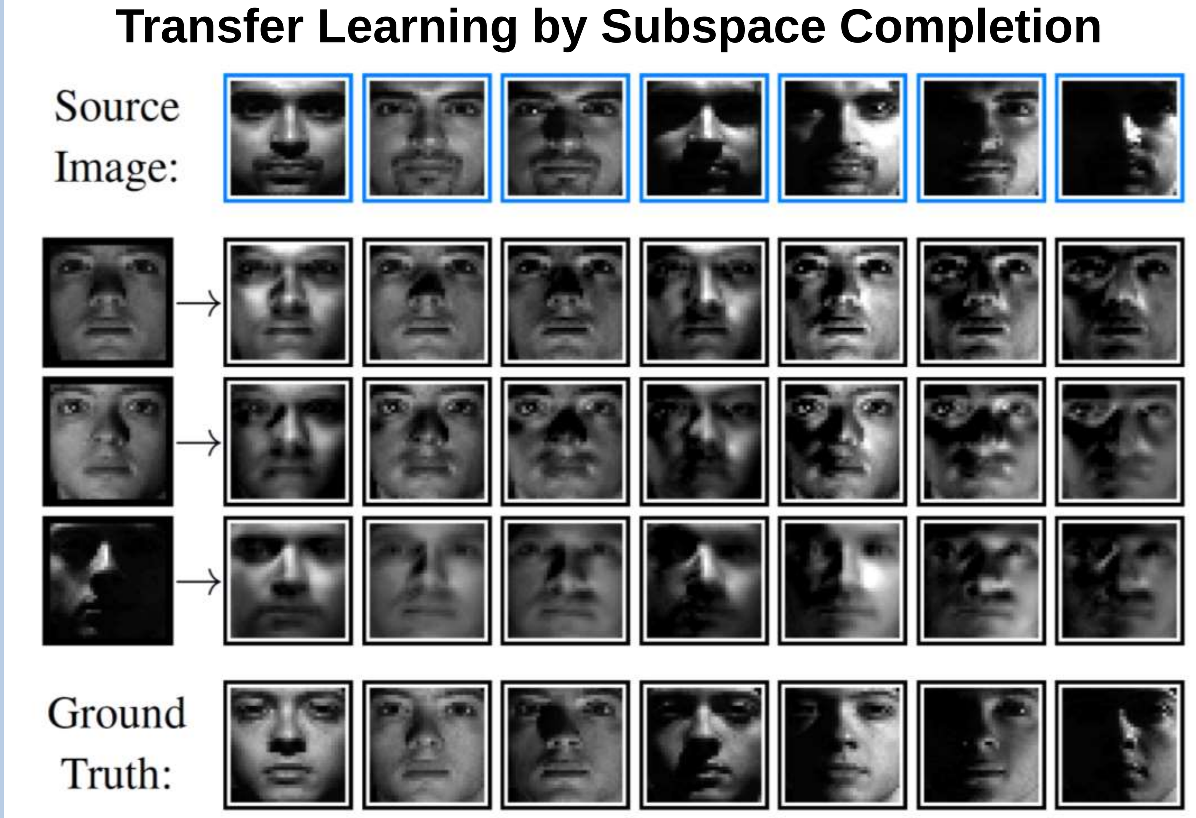
Nearest Neighbors



AGI — PCA

Transfer Learning by Subspace Completion

Source Image:



Ground Truth: