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

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

$$d_k(A, B) = \| \theta \|_2, \quad A^T B = U \text{diag}(\cos \theta) V^T$$

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

$$d_F(A, B) = \sin \| \theta \|_2^{-1} = 2^{-1} \| [\Pi(A) - \Pi(B)]_F \|_F$$

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

$$\mathbb{P}(k,d) = \{ P \in \mathbb{R}^{d \times d} : P^T P = P = P^T, \text{tr}(P) = k \}$$

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

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

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.

Problem Formulation

From a set of $p_i$-dimensional subspaces $X_i \in \mathbb{G}(k,d)$, we aim to learn a $k$-dimensional subspace $B \in \mathbb{G}(k,d)$ with $p_i \leq k$. For each $X_i$, we introduce an auxiliary local subspace $Z_i \in \mathbb{G}(k,d)$ 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:

$$\arg \min_B \sum_i d_F^2(B, Z_i) \text{ s.t. } Z_i \supseteq X_i$$

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

$$\arg \min_P \sum_i \frac{\| P - Q_i \|_F^2}{2} \text{ s.t. } Q_i X_i = X_i$$

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 $X_i$, as demonstrated by the following observation connecting AGI with standard PCA.

$$\min_{Q_i \in \mathbb{P}(k,d)} \frac{1}{2} \| P - Q_i \|_F^2 \text{ s.t. } Q_i X_i = X_i = X_i - P X_i$$

Inference and Subspace Completion

The solution subspace associated with $P$ 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 $P$ formed from the top $k$ dimensions of $P$, the completed subspace spans the columns of $X_i = [X_i \cdot X_i]$, where $X_i$ contains the top $m = p_i$ eigenvectors of $(1 - \alpha X_i^T P_i^T)$ as its columns.

Finally, a lower dimensional subspace associated with the projection matrix $M_i \in \mathbb{P}(m,k)$ can be found as:

$$M_i = \arg \min_{M_i \in \mathbb{P}(m,k)} \| X_i X_i^T - B M_i B^T \|_F^2 = \bar{X_i}^T \bar{P}_X$$

Extensions

AGI can also accommodate additional prior knowledge encoded as other constraints $c_i$ 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: $C^* = \{ Q : Q \in \mathbb{P}(k,d), Q X_i + E_i = X_i, E_i \| X_i \|_2 \leq \varepsilon \}$

Experimental Results

Convergence Performance of Averaged Projections

Nearest Neighbor Classification of Synthetic Data

Inference and Subspace Completion

Latent Space Visualization

Transfer Learning by Subspace Completion