Joint subspace intersections as a fitting problem

Christian Lageman and Knut Hüper

***Abstract*— In this paper we consider the task of estimating a joint intersection of several subspaces from perturbed measurements of the subspaces. We treat this problem as a fitting problem on a Grassmann manifold. A potential application in face recognition is discussed.**

I. INTRODUCTION

Computing the joint intersection of several subspaces $\cap_{j=1}^m W_j$ is a more or less straightforward task and can be solved e.g. by SVD based [5] or eigenvalue based approaches [9]. However, in most practical applications measurements of such subspaces will be corrupted by noise, making an application of these straightforward algorithms infeasible. Here, we consider this subspace intersection estimation problem for such perturbed measurements. We propose a formulation of this problem as a least squares fitting problem on a Grassmann manifold. Three different distance measures, namely the chordal distance, the geodesic distance (with respect to the normal or Euclidean metric) and the subspace distance, are used. For the subspace and the chordal distance, it is shown how the fitting problem can be reduced to a smooth optimization problem with a simple cost function. Such optimization problems can be efficiently tackled by optimization on manifolds approaches [6], [1]. Possible applications for these type of problems occur in computer vision and in machine learning as well. Specifically, we propose an application in face recognition.

II. THE SUBSPACE INTERSECTION PROBLEM

First, we consider the subspace intersection problem for spaces of the same dimension.

Let $W_1, \ldots, W_m$ be subspaces of $\mathbb{R}^n$ each of dimension k. We assume that the joint intersection of all these subspaces is a p-dimensional subspace $V \subset \mathbb{R}^n$. In this paper we consider the task to compute an estimate for this subspace V from perturbed measurements $\widetilde{W}_1, \ldots, \widetilde{W}_n$ of the W_i.

A. The Grassmann manifold

The set of all subspaces of dimension k of $\mathbb{R}^n$ has a natural structure as a smooth manifold, the so-called Grassmann manifold or Grassmannian $\mathrm{Grass}(n,k)$, see e.g. [6]. The Grassmann manifold can be identified with the smooth manifold of orthogonal rank k projectors in $\mathbb{R}^n$, i.e.

$$\mathrm{Grass}(n,k) = \{P \in \mathbb{R}^{n\times n} \mid P^T = P, P^2 = P, \operatorname{rank} P = k\},$$

by identifying a subspace with the orthogonal projection onto itself. Note that there are alternative representations of the Grassmann manifold which can be more suitable for the implementation of a specific algorithm [1]. However, such implementation related details are not the main concern of this work. Therefore, we exclusively use the projector representation here. For a subspace W we denote by P_W the orthogonal projector onto W.

The orthogonal group

$$O(n) = \{U \in \mathbb{R}^{n\times n} \mid UU^T = I_n\}$$

acts on $\mathrm{Grass}(n,k)$ in a transitive way via conjugation,

$$\begin{aligned} \mathrm{conj}\colon O(n) \times \mathrm{Grass}(n,k) &\to \mathrm{Grass}(n,k), \\ \mathrm{conj}(U, P_W) &= U^T P_W U, \end{aligned} \tag{1}$$

which means

$$\mathrm{conj}(U, \mathrm{conj}(T, P_W)) = \mathrm{conj}(UT, P_W)$$

for all $U, T \in O(n)$, $P_W \in \mathrm{Grass}(n,k)$. Conjugation $\mathrm{conj}(U, P_W)$ corresponds to mapping the subspace W onto its image $U(W)$ under $U \in O(n)$, i.e. $\mathrm{conj}(U, P_W) = P_{U(W)}$. We denote by

$$\mathrm{stab}_k(P_W) = \{U \in O(n) \mid U^T P_W U = P_W\}$$

the isotropy (stabilizer) subgroup of $O(n)$ stabilizing $P_W \in \mathrm{Grass}(n,k)$.

B. Differential geometry of the intersection problem

To study the geometry of the intersection problem, we first consider the unperturbed case. The subspaces $W_1, \ldots, W_m$ correspond to points P_{W_i} on the Grassmann manifold $\mathrm{Grass}(n,k)$. Since the W_i intersect in a subspace V, the P_{W_i} are all elements of the set

$$M_V = \{P_W \in \mathrm{Grass}(n,k) \mid V \subset W\}.$$

Proposition 1: For any p-dimensional subspace $V \subset \mathbb{R}^n$ the set M_V is an orbit under the action of $\mathrm{stab}_p(P_V)$ on $\mathrm{Grass}(n,k)$ by conjugation, i.e.

$$M_V = \{U^T P_W U \mid U \in \mathrm{stab}_P(P_V)\}$$

with P_W an arbitrary point on M_V. In particular, M_V is a smooth, embedded, compact submanifold of $\mathrm{Grass}(n,k)$. Computing the intersection V is therefore equivalent to computing the submanifold M_V.

Assume now that only the perturbed measurements $\widetilde{W}_i$ instead of the original W_i are available. In this case we can only try to compute an *estimate* $\widetilde{V}$ of V from the data. However, the $\widetilde{W}_i$ will in general not intersect in a p-dimensional subspace. Therefore, to get an estimate $\widetilde{V}$ we have to find a p-dimensional subspace which fits best to

The authors are with Institut für Mathematik, Julius-Maximilians-Universität Würzburg, Am Hubland, 97074 Würzburg, Germany {lageman,hueper}@mathematik.uni-wuerzburg.de

ISBN 978-963-311-370-7

our data. By Proposition 1 the p-dimensional subspaces of $\mathbb{R}^n$ correspond to embedded submanifolds of $\mathrm{Grass}(n,k)$. Therefore fitting the subspace to all the $\widetilde{W}_i$ corresponds to fitting a submanifold M_V to the points $P_{\widetilde{W}_i}$ on $\mathrm{Grass}(n,k)$. We now formulate this fitting problem as a least-squares fitting problem on $\mathrm{Grass}(n,k)$:

Problem 1: Let dist be a distance function on $\mathrm{Grass}(n,k)$. Find a p-dimensional subspace $V \subset \mathbb{R}^n$ such that

$$h(V) = \sum_{i=1}^{m} \mathrm{dist}^2(P_{\widetilde{W}_j}, M_V) \tag{2}$$

is minimized with

$$\mathrm{dist}(P_{\widetilde{W}_j}, M_V) = \min_{P_W \in M_V} \mathrm{dist}(P_{\widetilde{W}_j}, P_W) \tag{3}$$

denoting the distance between a data point and the set M_V.

The statement of Problem 1 leaves the question open which distance should be chosen. On $\mathrm{Grass}(n,k)$ there are several well known possible candidates. Before we proceed we recall the definition of principal angles between subspaces, see [5]. Let W be a $k-$dimensional and V a p-dimensional subspace of $\mathbb{R}^n$, respectively. The *principal angles*

$$\psi_1(W,V) \le \ldots \le \psi_p(W,V) \in [0,\pi/2]$$

between W and V and the associated *principal directions* $w_i \in W$, $v_i \in V$ are defined recursively by $\|v_i\| = \|w_i\| = 1$, $v_i^T v_j = 0$, $w_i^T w_j = 0$ for all $j < i$, and

$$\cos(\psi_i(W,V)) = \max_{\substack{w\in W, \|w\|=1 \\ w^T w_j = 0 \\ j=1,\ldots,i-1}} \max_{\substack{v\in V, \|v\|=1 \\ v^T v_j = 0 \\ j=1,\ldots,i-1}} v^T w = v_i^T w_i$$

with $\|\cdot\|$ the Euclidean norm on $\mathbb{R}^n$.

On $\mathrm{Grass}(n,k)$ we have the following distances:

1) *Chordal distance*, cf. [3],

$$\mathrm{dist}_c^2(P_W, P_Z) = \tfrac{1}{2}\|P_W - P_Z\|_F^2 = \sum_{i=1}^{k} \sin^2(\psi_i(W,Z))$$

with $\|A\|_F = \sqrt{\mathrm{tr}(A^T A)}$ the Frobenius norm.

2) *Geodesic distance*, cf. [3],

$$\mathrm{dist}_g^2(P_W, P_Z) = \sum_{i=1}^{k} \psi_i^2(W,Z).$$

This is the squared geodesic distance with respect to the Riemannian metric on $\mathrm{Grass}(n,k)$ induced by the Euclidean one on $\mathbb{R}^{n\times n}$.

3) *Subspace distance*, cf. [5],

$$\mathrm{dist}_s(P_W, P_Z) = \|P_W - P_Z\|_2 = \sin(\psi_k(W,Z)) \tag{4}$$

with $\|A\|_2 = \max_{v\in\mathbb{R}^n, \|v\|=1} \|Av\|$, $\|v\|$ the Euclidean norm of a vector in $\mathbb{R}^n$.

Note that neither the geodesic distance nor the subspace distance are globally defined smooth functions on $\mathrm{Grass}(n,k) \times \mathrm{Grass}(n,k)$. This can be easily seen from the definition via the principal angles.

C. Subspace span estimation

Equivalent to our subspace intersection problem is the following subspace *span* estimation problem. Let $W_1, \ldots, W_m$ be subspaces of $\mathbb{R}^n$ of dimension k. Assume that $\mathrm{span}\{W_1,\ldots,W_k\} = V$ with V a p-dimensional subspace of $\mathbb{R}^n$, i.e. the joint span of these subspaces is equal to V. Consider the task of estimating V from perturbed measurements $\widetilde{W}_j$ of the W_j. Obviously, simply computing the span of the $\widetilde{W}_j$ will yield a far too large space. The problem can be solved by using the simple identity $V^\perp = \bigcap_{j=1}^n W_j^\perp$. Thus to estimate V, we can just estimate $V^\perp$ from $\widetilde{W}_j^\perp$ by solving the subspace intersection problem discussed before.

III. Cost functions for the fitting problem

Even for the distances discussed above the representation of the cost function in Problem 1 is not suitable for computational purposes since it contains a minimization problem itself: an optimization algorithm to find the minimum of this cost function would have to operate on a greatly enlarged parameter space. However, for the chordal distance and the subspace distance, we can give a simplified representation of the cost which makes the problem more accessible for standard optimization approaches on manifolds. Our argument depends on several results on the principal angles.

First, we have the result that for given subspaces W, V with $\dim W \ge \dim V$, the orthogonal projection of V to W is included in the span of the principal directions w_j in W.

Lemma 1: Let W and V be subspaces of $\mathbb{R}^n$ with $\dim W = k$ and $\dim V = p$. Denote by $w_1,\ldots,w_p \in W$, $v_1,\ldots,v_p \in V$ the principal directions. Then $\pi_W(V) \subset \mathrm{span}\{w_1,\ldots,w_p\}$ with $\pi_W\colon \mathbb{R}^n \to \mathbb{R}^n$ the orthogonal projection onto W.

This result enables one to construct a $\widetilde{V}$ in M_V which acts as a solution of $\min_{W\in M_V} \mathrm{dist}(W_j, W)^2$ for a fixed j, either for the chordal or for the subspace distance function.

Proposition 2: Let W and V be subspaces of $\mathbb{R}^n$ with $\dim W = k$ and $\dim V = p$. Then there exists $P_{\widetilde{V}} \in M_V$ such that $\widetilde{V} = V \oplus Z$ with $Z \subset W$ and $V \perp Z$. In particular,

$$\psi_i(W,\widetilde{V}) = 0 \text{ for } i = 1,\ldots,k-p,$$
$$\psi_i(W,\widetilde{V}) = \psi_{i-k+p}(W,V) \text{ for } i = k-p+1,\ldots,k.$$

We want to point out that the $\widetilde{V}$ constructed in Proposition 2 is not extremal in the sense that any other subspace $\widehat{V}$ containing V will have larger or equal principal angles with respect to W.

Additionally, we need the intuitive result, that extending a subspace V to a larger dimensional one, say $\widetilde{V}$, will not necessarily result in a reduced largest principal angle between $\widetilde{V}$ and a fixed subspace W when compared to the largest principal angle between V and W.

Proposition 3: Let W and V be subspaces of $\mathbb{R}^n$ with $\dim W = k$ and $\dim V = p$. Assume that we are given $\widetilde{V} \subset \mathbb{R}^n$ with $\dim \widetilde{V} = k$ and $V \subset \widetilde{V}$, then $\psi_k(W,\widetilde{V}) \ge \psi_p(W,V)$.

With these results, we can give explicit formulas for $\operatorname{dist}(P_{\widetilde{W}_j}, M_V)$ for the chordal distance and the subspace distance.

Theorem 1: Let V be a p-dimensional and W a k-dimensional subspace of $\mathbb{R}^n$. Then we have

$$\operatorname{dist}_c^2(P_W, M_V) = p - \operatorname{tr}(P_W P_V),$$
$$\operatorname{dist}_s^2(P_W, M_V) = 1 - \|P_W P_V\|_2^2.$$

Proof: We only sketch the proof of this theorem. First consider $\operatorname{dist}_c^2(P_W, M_V)$. Without loss of generality we can assume that $V = \operatorname{span}\{e_1, \ldots, e_p\}$. We have

$$\begin{aligned}
&\operatorname{dist}_c^2(P_W, M_V)\\
&= \min_{\substack{\Theta_1 \in O(p)\\ \Theta_2 \in O(n-p)}} \left\| P_W - \begin{pmatrix} \Theta_1 & 0 \\ 0 & \Theta_2 \end{pmatrix} \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \Theta_1^T & 0 \\ 0 & \Theta_2^T \end{pmatrix} \right\|_F^2\\
&= k - \max_{\Theta \in O(n-p)} \operatorname{tr}\left(P_W \begin{pmatrix} I_p & 0 \\ 0 & \Theta \end{pmatrix} \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} I_p & 0 \\ 0 & \Theta^T \end{pmatrix} \right)\\
&= k - \operatorname{tr}(P_W P_V) -\\
&\quad \max_{\Theta \in O(n-p)} \operatorname{tr}\left(\begin{pmatrix} I_p & 0 \\ 0 & \Theta \end{pmatrix} \begin{pmatrix} 0_p & 0 & 0 \\ 0 & I_{k-p} & 0 \\ 0 & 0 & 0_{n-k} \end{pmatrix} \begin{pmatrix} I_p & 0 \\ 0 & \Theta^T \end{pmatrix} \right).
\end{aligned}$$

Since by Proposition 2 we can choose a $Z \in W \cap V^\perp$ with $\dim Z = k - p$, there is a $\Theta \in O(n-p)$ fulfilling

$$\operatorname{tr}\left(\begin{pmatrix} I_p & 0 \\ 0 & \Theta \end{pmatrix} \begin{pmatrix} 0_p & 0 & 0 \\ 0 & I_{k-p} & 0 \\ 0 & 0 & 0_{n-k} \end{pmatrix} \begin{pmatrix} I_p & 0 \\ 0 & \Theta^T \end{pmatrix} \right) = k - p.$$

Since for all $\Theta \in O(n-p)$ the difference $k-p$ is an upper bound for this expression, the subspace $\widetilde{V}$ given by Proposition 2 yields a global minimum $P_{\widetilde{V}}$ of $Q \mapsto \operatorname{dist}_c^2(P_W, Q)$ on M_V. In particular,

$$\operatorname{dist}_c^2(P_W, M_V) = k - \operatorname{tr}(P_W P_V) + k - p.$$

We now consider the subspace distance. For all $P_R \in M_V$ we have by Proposition 3 that

$$\operatorname{dist}_s(P_W, P_R) = \sin(\psi_k(W, R)) \geq \sin(\psi_p(W, V)).$$

On the other hand, Proposition 2 yields a $P_{\widetilde{V}} \in M_V$ such that

$$\operatorname{dist}_s(P_W, P_R) = \sin(\psi_k(W, R)) = \sin(\psi_p(W, V)).$$

Thus $P_{\widetilde{V}}$ is a global minimum of $Q \mapsto \operatorname{dist}_s^2(P_W, Q)$ on M_V and

$$\operatorname{dist}_s(P_W, M_V) = \sin(\psi_p(W, V)).$$

Let $Q_W \in \mathbb{R}^{n \times k}$, $Q_V \in \mathbb{R}^{n \times p}$ with $Q_W^T Q_W = I_k$, $Q_V^T Q_V = I_p$ and $P_W = Q_W Q_W^T$, $P_V = Q_V Q_V^T$. The cosines of the principal angles are given by the singular values of $Q_W^T Q_V$, cf. [5]. Thus

$$\begin{aligned}
1 - \sin^2(\psi_p(W, V)) &= \max_{v \in \mathbb{R}^p} v^T Q_V P_W Q_V v\\
&= \max_{v \in \mathbb{R}^p} v^T P_V P_W P_W P_V v\\
&= \|P_W P_V\|_2^2.
\end{aligned}$$

■

Given these results we can eliminate the implicit minimization over M_V for the fitting problem as follows.

Corollary 1: For all $P_{\widetilde{W}_j} \in \operatorname{Grass}(n,k)$, $P_V \in \operatorname{Grass}(n,p)$ we have

$$\sum_{j=1}^m \operatorname{dist}_c^2(P_{\widetilde{W}_j}, M_V) = mp - \sum_{j=1}^m \operatorname{tr}(P_{\widetilde{W}_j} P_V),$$
$$\sum_{j=1}^m \operatorname{dist}_s^2(P_{\widetilde{W}_j}, M_V) = m - \sum_{j=1}^m \|P_{\widetilde{W}_j} P_V\|_2^2.$$

This reformulation of the distance function allows us to use the following optimization problems on $\operatorname{Grass}(n,p)$ (respectively, on $\operatorname{St}(n,p) \times (S^{p-1})^m$) to compute an estimate for the intersecting subspace.

Corollary 2: In case of the chordal distance, solving Problem 1 is equivalent to finding a global maximum of

$$f_c\colon \operatorname{Grass}(n,p) \to \mathbb{R},$$
$$f_c(V) = \operatorname{tr}(Q P_V) \quad \text{with} \quad Q = \sum_{j=1}^m P_{\widetilde{W}_j}.$$

The function f_c is a standard cost function for computing the dominant subspace of the positive semidefinite matrix Q. One can compute this subspace via an eigenvalue decomposition of Q, i.e. for $Q = U \operatorname{diag}(\lambda_1, \ldots, \lambda_n) U^T$, $\lambda_1 \geq \ldots \geq \lambda_n \geq 0$, $U \in O(n)$, the optimal $P_V \in \operatorname{Grass}(n,p)$ is given by

$$P_V = U \begin{pmatrix} I_p & 0 \\ 0 & 0 \end{pmatrix} U^T. \tag{5}$$

Alternatively one could apply various optimization methods of $\operatorname{Grass}(n,p)$ to this cost function. This can yield a more efficient algorithm, since one does not have to compute a full eigenvalue decomposition of Q but can exploit the lower dimension of $\operatorname{Grass}(n,p)$. We refer the reader to [6], [1] for appropriate algorithms.

Problem 1 can be reduced for the subspace distance to a smooth optimization problem, as well. However, the dimension of the smoothed problem increases significantly.

Corollary 3: Denote by $\operatorname{St}(n,p)$ the compact Stiefel manifold

$$\operatorname{St}(n,p) = \{X \in \mathbb{R}^{n \times p} \mid X^T X = I_p\}.$$

Then any global maximum of

$$f_s\colon \operatorname{St}(n,p) \times (S^{p-1})^m \to \mathbb{R},$$
$$f_s(X, w_1, \ldots, w_m) = \sum_{j=1}^m w_j^T X^T P_{\widetilde{W}_j} X w_j$$

yields a solution to Problem 1 via $P_V = XX^T$. Vice versa, for a solution P_V of Problem 1 there exist $w_1, \ldots, w_m$ and $X \in \operatorname{St}(n,k)$ with $XX^T = P_V$ such that $(X, w_1, \ldots, w_m)$ is a global maximum of f_s.

IV. Application: Joint Principal Subspaces in Face Characterization

In this section we illustrate the relevance of the subspace intersection problem for applications by sketching an application in computer vision.

Kirby and Sirovich introduced the use of PCA and *eigenfaces* for characterization and representation of human faces [8], [7]. This method has been utilized e.g. in face recognition applications [10]. In recent years generalizations of PCA like e.g. ICA have been used for such applications, too, see e.g. [2], [4].

For the computation of eigenfaces, one is given a training set of images of the face of a single person. The images are represented by vectors $p_1, \dots, p_r$ in $\mathbb{R}^n$ for suitable n. Note that the values of the components of the p_j are not necessarily gray scales or even color representations of pixels – they could be also obtained from a Fourier transform or any other image representation. From the p_j one could compute an average image $\overline{p}$ as the mean of the p_j. Performing a PCA on the normalized images, i.e. $p_j - \overline{p}$, we obtain the principal components $u_1, \dots, u_t$. These are the co-called *eigenfaces* and give the directions where the data set p_j varies most. Since in general, $t \ll n$ they can be used for a compressed representation for additional images of the same person, by storing $\langle u_j, q \rangle$ $j = 1, \dots, t$ instead of the new image q. Furthermore, for another image of a possibly different face the distance to $\operatorname{span}\{u_1, \dots, u_t\}$ provides a criterion for face recognition algorithms [10].

Assume now that for m different persons we have computed the eigenfaces $u_{1m}, \dots, u_{tm}$ from suitable training sets. In general the $W_j := \operatorname{span}\{u_{1m}, \dots, u_{tm}\}$ will not coincide. But a common intersection of all W_j would give directions in which all training sets vary greatly. This joint principal subspace would be of interest for tasks like the identification of facial expressions, etc. In general, due to the influence of noise, we cannot expect the W_j to have a common intersection beyond what is dictated by dimensional constraints. However, we can use the approach from the previous section to obtain an estimate for the common intersection of the subspaces.

One might consider the idea to simply perform PCA over the union of all training sets. However this might not give the desired result, since it will capture the variation between images from the different training sets and this is not information one is usually interested in.

Note that this application is just one illustrative example and one can apply our method to similar problems in machine learning and computer vision. The detailed study of such applications is the subject of ongoing research.

V. Intersections of subspaces with different dimensions

In the previous application, there is no guarantee that each training set yields the same number of eigenfaces, i.e. that the dimensions of the W_j are the same. Hence, we now strive the more general problem that the dimensions of the subspaces $W_1, \dots, W_m$ are not the same. That means that we are given subspaces $W_1, \dots, W_m$ of $\mathbb{R}^n$ with dimensions $\dim W_1 = k_1, \dots, \dim W_m = k_m$ whose joint intersection is a p-dimensional subspace V. We want to estimate V from perturbed measurements $\widetilde{W}_i$ of the W_i with $\dim \widetilde{W}_i = \dim W_i = k_i$.

To tackle this problem, we consider an equivalent formulation of the Problem 1. Assume that the dimensions are again identical, i.e. $k_1 = \dots = k_m = k$. Let

$$N = \operatorname{Grass}(n,k) \times \dots \times \operatorname{Grass}(n,k) = \operatorname{Grass}(n,k)^m$$

the m-fold product of $\operatorname{Grass}(n,k)$ with itself. Given any distance dist on $\operatorname{Grass}(n,k)$ we can equip $\operatorname{Grass}(n,k)^m$ with distance defined via

$$\operatorname{dist}^2_N((P_1, \dots, P_m), (Q_1, \dots, Q_m)) = \sum_{j=1}^m \operatorname{dist}^2(P_j, Q_j).$$

The unperturbed points $W_1, \dots, W_m$ correspond to the single point $P = (P_{W_1}, \dots, P_{W_m})$ in $\operatorname{Grass}(n,k)^m$. Since the W_i have unique joint intersection V the point P is contained in the set

$$M_V^N := \{(P_{R_1}, \dots, P_{R_m}) \mid P_{R_j} \in \operatorname{Grass}(n,k), V \subset R_j\}.$$

In fact, this set is the m-fold product of M_V and therefore a compact submanifold of $\operatorname{Grass}(n,k)^m$. For perturbed measurements $\widetilde{W}_j$ Problem 1 is equivalent to fit M_V^N to a *single point* in $\operatorname{Grass}(n,k)^m$.

Proposition 4: Let $P = (\widetilde{W}_1, \dots, \widetilde{W}_m)$. Then a p-dimensional subspace $V \subset \mathbb{R}^n$ solves Problem 1 if and only if it minimizes

$$\operatorname{dist}^2_N(P, M_V^N) = \min_{Q \in M_V^N} \operatorname{dist}^2_N(P, Q). \tag{6}$$

Proposition 4 suggests a natural generalization of the intersection problem to subspaces $W_1, \dots, W_m$ of different dimension. Consider the product N of m Grassmann manifolds of subspaces with dimensions k_1, i.e.

$$N := \operatorname{Grass}(n,k_1) \times \dots \times \operatorname{Grass}(n,k_m).$$

We assume that on each $\operatorname{Grass}(n,k_j)$ we are given a distance dist_j. Analogous to the case of identical dimensions, we can define on N a distance via

$$\operatorname{dist}^2_N((P_1, \dots, P_m), (Q_1, \dots, Q_m)) = \sum_{j=1}^m \operatorname{dist}^2_j(P_j, Q_j).$$

Furthermore for any p-dimensional subspace $V \supset \mathbb{R}^n$ we define the set

$$M_V^N = \{(P_{R_1}, \dots, P_{R_m}) \mid V \subset R_j, P_{R_j} \in \operatorname{Grass}(n,k_j)\}.$$

Proposition 5: For any p-dimensional subspace $V \subset \mathbb{R}^n$ the set M_V^N is the product $M_V^1 \times \dots \times M_V^m$ with

$$M_V^j = \{P_W \mid P_W \in \operatorname{Grass}(n,k_j), V \subset W\}.$$

In particular, it is an orbit of the action

$$\psi((U_1, .., U_m), (P_1, .., P_m)) = (U_1 P_1 U_1^T, .., U_m P_m U_m^T)$$

of $\operatorname{stab}_p(V)^m$ on N and therefore an embedded, compact submanifold of N.

Proposition 4 justifies to use an extension of (6) as a generalization of the Problem 1 to a collection of subspaces with different dimensions. Hence, we consider the following fitting problem to compute an estimate for the intersection of the unperturbed W_j.

Problem 2: Find a p-dimensional subspace V of $\mathbb{R}^n$ such that

$$\operatorname{dist}_N^2(P, M_V^N) = \min_{Q \in M_V^N} \operatorname{dist}_N^2(P, Q)$$

is minimized.

As in the case for identical dimensions of the W_j, the cost functions for Problem 2 contain an implicit minimization of a smooth function themselves. A direct solution would require an optimization over a greatly enlarged state space. However, if the distances on the Grassmannians $\operatorname{Grass}(n, k_j)$ are all of chordal or subspace type, we can use Theorem 1 to reduce this optimization problem to an optimization problem on $\operatorname{Grass}(n, p)$ or similarly to one on $\operatorname{St}(n, p) \times (S^{p-1})^m$.

Theorem 2: If all dist_j are chordal distances on the respective $\operatorname{Grass}(n, k_j)$ then

$$\operatorname{dist}_N^2(P, M_V^N) = mp - \sum_{j=1}^{m} \operatorname{tr}(P_{\widetilde{W}_j} P_V).$$

If all dist_j are subspace distances on the respective $\operatorname{Grass}(n, k_j)$ then

$$\operatorname{dist}_N^2(P, M_V^N) = m - \sum_{j=1}^{m} \|P_{\widetilde{W}_j} P_V\|_2^2.$$

We see that Theorem 2 provides the same functions as Corollary 1. Thus we can use the same cost function as before to compute the estimate for V. We have the following corollary.

Corollary 4: If all dist_j are chordal distances then the solutions of Problem 2 are given by the global maxima of

$$f_c\colon \operatorname{Grass}(n, p) \to \mathbb{R},$$
$$f_c(V) = \operatorname{tr}(QP_V) \quad \text{with} \quad Q = \sum_{j=1}^{m} P_{\widetilde{W}_j}.$$

If all dist_j are subspace distances then all solutions of Problem 2 are given by the global maxima of

$$f_s\colon \operatorname{St}(n, p) \times (S^{p-1})^m \to \mathbb{R},$$
$$f_s(X, w_1, \ldots, w_m) = \sum_{j=1}^{m} w_j^T X^T P_{\widetilde{W}_j} X w_j$$

via $P_V = XX^T$.

Remark 1: One could also consider the task of estimating V from measurements $\widetilde{W}_i$ with $\dim \widetilde{W}_i \neq \dim W_i$. This would arise e.g. when in applications as the feature extraction problem discussed above the dimension of the principal subspaces cannot uniquely determined from the data points of a group. However, to tackle this more general problem is beyond the scope of this paper.

References

[1] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, New Jersey, 2008.

[2] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski. Face recognition by independent component analysis. *IEEE Transactions on Neural Networks*, 13(6):1450–1464, 2002.

[3] J. H. Conway, R. H. Hardin, and N. J. A. Sloane. Packing lines, planes, etc.: Packings in Grassmannian spaces. *Experimental Mathematics*, 5:139–159, 1996.

[4] B. A. Draper, K. Baek, M. S. Bartlett, and J. R. Beveridge. Regocnizing faces with PCA and ICA. *Computer Vision and Image Understanding*, 91:115–137, 2003.

[5] G. H. Golub and C. F. van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, 1996.

[6] U. Helmke and J. Moore. *Optimization and Dynamical Systems*. Springer, London, 1994.

[7] M. Kirby and L. Sirovich. Application of the Karhunen-Loève procedure for characterization of human faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):103–108, 1990.

[8] L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A*, 4:519–524, 1987.

[9] C. W. Therrien. Eigenvalue properties of projection operators and their application to the subspace method of feature extraction. *IEEE Transactions on Computers*, 24(9):944–948, 1975.

[10] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.