## A Grassmannian minimum enclosing ball approach for common subspace extraction

Emilie Renard<sup>1</sup>, Kyle A. Gallivan<sup>2</sup>, and P.-A. Absil<sup>1</sup>

 ICTEAM Institute, Université catholique de Louvain Louvain-la-Neuve, 1348, Belgium {emilie.renard,pa.absil}@uclouvain.be
 Department of Mathematics, Florida State University Tallahassee FL 32306-4510, USA kgallivan@fsu.edu

**Abstract.** We study the problem of finding a subspace representative of multiple datasets by minimizing the maximal dissimilarity between this subspace and all the subspaces generated by those datasets. After arguing for the choice of the dissimilarity function, we derive some properties of the corresponding formulation. We propose an adaptation of an algorithm used for a similar problem on Riemannian manifolds. Experiments on synthetic data show that the subspace recovered by our algorithm is closer to the true common subspace than the solution obtained using an SVD.

Keywords: Common subspace extraction, total Grassmannian, minimal enclosing ball

## 1 Introduction

We address the problem of extracting common information from multiple datasets. In recent years data has become increasingly easy to generate and store for analysis to guide decision making, and it is not uncommon to have access to datasets representing similar but not exactly equivalent phenomena. A typical example can be found in bioinformatic, where datasets usually have a few tens to at most a few hundreds of samples for a few (tens of) thousands of features. However there usually exist various datasets measuring the same disease on different sets of patients, but corresponding to different studies and different experimental conditions that should be taken into account in further analysis. Considering all those similar datasets at once can be very useful to deal with the high number of features since statistical inferences require a large number of samples to be robust enough and generalizable to other data.

Beside the basic possibility to simply concatenate all the datasets  $X_1,...,X_m$ into a larger dataset  $X = [X_1 \ldots X_m]$  and apply usual methods such as principal components analysis on X, more specific approaches exist to extract common components present in the datasets. A method to factorize two datasets with a

common factor was proposed in [1] with a closed-form solution, and an extension to more than two datasets was proposed in [2]. However, such methods assume that the common dimension of the datasets is full-rank, which is not the case if we consider datasets with a higher number of variables than samples, such as gene expression datasets. The best known method is probably canonical correlation analysis (CCA) [3], which aims to find a linear combination of the initial features for both datasets maximizing the correlation between those two combinations. When dealing with two datasets only, an exact solution can be computed based on the covariance matrix. In order to find more than one pair of correlated combination of features, deflation is usually used: the same procedure (CCA) is repeated on the data from which the previous components were removed. Another well known method, partial least square regression [4], aims to find linear combinations of features for the two datasets such that the covariance between those two new representations is maximal. As in CCA a closed-form solution exists, and deflation can be used to compute the following components. Another variation is co-inertia analysis (CIA) and its extension multiple CIA [5] that maximizes a sum of weighted squared covariances between linear combination of the datasets features and a reference vector. Consensus principal component analysis is very similar to CIA, the main difference being in the deflation process [6]. Different extensions of those methods to more than two datasets have been proposed, with various criteria to optimize (see for example [7, 8] and references within): maximizing a sum on all pairs of datasets of covariances or correlations, possibly squared or in absolute value, and with different constraints. In such cases, a closed-form solution does not always exist.

A central question when using more than two datasets is the importance to give to those different (pairs of) datasets. Common approaches are to give all datasets the same importance or, as in [7], to consider if a pair of datasets is connected or not and to give to the corresponding term a weight of 1 or 0. If we are dealing with a set of datasets all very similar except one (for example, because measured using another technology), those kind of choices can lead to components representing very well all the similar datasets but being not representative at all of the last one. Here, we want to avoid this situation, and in order to take all  $X_i$  into account we propose to minimize the maximal dissimilarity d between the common component  $U \in \mathbb{R}^{p \times K}$  and all datasets  $X_i \in \mathbb{R}^{p \times n_i}$ :

$$U^* = \arg\min_{U} \max_{i} d(U, X_i).$$
(1)

This formulation can be viewed as looking for the center of the smallest-radius sphere enclosing all  $X_i$ , and can be linked to the minimum enclosing ball, 1-center problem or minimax optimization problem. However, since here U represents a subspace, we are really interested in the subspace generated by the columns of U. So we want to solve problem (1) such that  $d(U, X_i) = d(\mathcal{U}, \mathcal{X}_i)$  is a dissimilarity measure between  $\mathcal{U}$  and  $\mathcal{X}_i$ , the subspaces generated by the columns of U and X.

The problem of finding the smallest enclosing ball of a finite point set  $\mathbb{X} = \{x_1, \ldots, x_m\}$  has been already thoroughly investigated in Euclidean space,

and an efficient approximation algorithm has been proposed in [9]. An adaptation of the algorithm presented in [9] to Riemannian geometry is proposed in [10] with a study of the convergence rate, and in [11] to compute Riemannian  $L_1$  and  $L_{\infty}$ center of mass of structure tensor images in order to denoise those images.

In this paper we assume that each point  $\mathcal{X}_i$  represents a subspace of dimension  $n_i$  in  $\mathbb{R}^p$ , that is  $\mathcal{X}_i$  belongs to the Grassmannian manifold  $\mathcal{G}(n_i, p)$  and so  $\mathbb{X} = \{\mathcal{X}_1, \ldots, \mathcal{X}_m\}$  is included in the total Grassmannian  $\cup \mathcal{G}(n_i, p)$ . The proposed approach to solve problem (1) is inspired by [10]. The main difference is that each data point  $\mathcal{X}_i$  belongs to a different Grassmannian  $\mathcal{G}(n_i, p)$ , which prevents us by using the usual Grassmannian distance. Instead we use an adaptation based on principal angles, which allows us to measure the dissimilarity between any pair of subspaces of different dimension, and to project  $\mathcal{G}(n_i, p)$  on  $\mathcal{G}(K, p)$  in order to return to a common manifold.

The paper is organized as followed. We first discuss the choice of the dissimilarity measure and the resulting problem in Section 2, then details of the proposed approach are presented in Section 3. Section 4 describes the results obtained on synthetic data, and we conclude in Section 5.

## 2 Problem formulation

Let  $X_i \in \mathbb{R}^{p \times n_i}$  be a matrix of p variables times  $n_i$  samples, for i = 1, ..., m. Our goal is to find a subspace  $\mathcal{U}$  of dimension K representative of all the subspaces  $\mathcal{X}_i$ , where  $\mathcal{X}_i$  is the subspace generated by the columns of  $X_i$ . In other words we are looking for a  $U \in \mathbb{R}^{p \times K}$  minimizing  $d(U, X_i)$  for all i, where  $d(U, X) = d(\mathcal{U}, \mathcal{X})$  is a dissimilarity measure between the span of U and the span of X.

## 2.1 Dissimilarity measure

Different dissimilarities are possible to quantify d(U, X), we detail some of them below. For K = 1, a possible choice to evaluate if a vector  $u \in \mathbb{R}^p$  is close to  $\mathcal{X}$ is the angle between u and its orthogonal projection on  $\mathcal{X}$ . A vector u is close to the subspace  $\mathcal{X}$  if the (positive) angle between them is small. If we define  $\phi$ as the angle between u and  $\mathcal{X}$  (in  $[-\frac{\pi}{2}, \frac{\pi}{2}]$ ), we have

$$u^{\top} \check{X} \check{X}^{\top} u = \cos^2 \phi$$

where ||u|| = 1 and  $\check{X}$  is an orthonormal basis of  $\mathcal{X}$ . The term  $u^{\top}\check{X}\check{X}^{\top}u$  can then be seen as a similarity measure evaluating how close u is to  $\mathcal{X}$ , with a value of 1 when u is in the subspace  $\mathcal{X}$ , and 0 when they are orthogonal. We can then define a dissimilarity:

$$d(u, X) = 1 - u^{\mathsf{T}} \check{X} \check{X}^{\mathsf{T}} u = \sin^2 \phi$$

with d(u, X) = 0 if and only if  $u \in \mathcal{X}$ .

This can be extended to a more general  $U \in \mathbb{R}^{p \times K}$  with  $p \ge n \ge K \ge 1$ (with *n* the dimension of  $\mathcal{X}$ ) by summing the dissimilarities obtained for each element of an orthonormal basis  $\check{U}$  of  $\mathcal{U}$ :

$$d_a(U,X) = \sum_k 1 - \check{U}(:,k)^\top \check{X} \check{X}^\top \check{U}(:,k) = K - \operatorname{Tr}\left(\check{U}^\top \check{X} \check{X}^\top \check{U}\right) = \sum_k \sin^2 \phi_k(U,X)$$

with  $\cos \phi_k(U, X)$  the singular values of  $\check{U}^{\top} \check{X}$ . Note that this quantity does not depend on the  $\check{U}$  or  $\check{X}$  chosen.

Another possible dissimilarity is [12]:

$$d_b(U,X) = \frac{1}{\sqrt{2}} ||\check{U}\check{U}^\top - \check{X}\check{X}^\top||_F = \sqrt{\frac{K+n}{2} - \operatorname{Tr}(\check{U}^\top\check{X}\check{X}^\top\check{U})} = \sqrt{\frac{n-K}{2} + \sum_k \sin^2 \phi_k(U,X)}.$$

Similarly, we can consider the norm between  $\check{X}$  and its projection onto the common subspace  $\mathcal{U}$  (termed *chordal metric* in [13, Table 3]):

$$d_c(U,X) = ||(I - \check{U}\check{U}^{\top})\check{X}||_F = \sqrt{n - \operatorname{Tr}\left(\check{U}^{\top}\check{X}\check{X}^{\top}\check{U}\right)}$$
$$= \sqrt{n - K + \sum_k \sin^2 \phi_k(U,X)}$$

Another possibility is to consider the principal angles  $\phi_k$  between both subspaces:

$$d_d(U,X) = \sqrt{\sum_k \phi_k^2(U,X)}$$

See [13] for other possible dissimilarity measures.

Letting  $\sigma_k = \cos \phi_k(U, X)$  denote the *k*th singular value of  $\check{U}^{\top}\check{X}$ , we can compare the different dissimilarities in Table 1 (with  $n_u$  and  $n_x$  dimensions of subspaces  $\mathcal{U}$  and  $\mathcal{X}$ ). When using those dissimilarities in  $\min_U \max_i d(U, X_i)$ ,  $d_b$ and  $d_c$  will give more importance to datasets  $X_i$  with a higher  $n_i$ . All dissimilarities except  $d_d$  can be directly expressed in terms of  $\check{U}^{\top}XX^{\top}\check{U}$ . As  $d_a$  and  $d_d$ respect  $\mathcal{U} \subsetneq \mathcal{X} \Rightarrow d(U, X) = 0$ , they are not distances. Note that if  $n_x = n_u$ , we have  $\sqrt{d_a} = d_b = d_c$ .

In the context of (1), it is natural to require that d(U, X) = 0 when  $\mathcal{U} \subset \mathcal{X}$ or  $\mathcal{X} \subset \mathcal{U}$ . We opt for  $d_a$ , since it yields a simpler objective function than  $d_d$ . Hence, (1) becomes:

$$\min_{U \in \mathbb{R}^{p \times K}} \max_{i} K - \operatorname{Tr}(\check{U}^{\top}\check{X}_{i}\check{X}_{i}^{\top}\check{U}).$$

Since K is fixed and  $\check{U}$  verifies  $\check{U}^{\top}\check{U} = I_K$ , this is equivalent to

$$\max_{U^{\top}U=I} \min_{i} \operatorname{Tr}(U^{\top} \check{X}_{i} \check{X}_{i}^{\top} U).$$
(2)

Formulation	Distance	$\mathcal{U}\subset\mathcal{X}\Rightarrow d(U,X)=0$
$d_a(X,U) = \min(n_x, n_u) - \sum_k^{\min(n_x, n_u)} \cos^2(\phi_k)$	-	$\checkmark$
$d_b(X,U) = \sqrt{\frac{n_x + n_u}{2} - \sum_k^{\min(n_x, n_u)} \cos^2(\phi_k)}$	$\checkmark$	-
$d_c(X,U) = \sqrt{\max(n_x, n_u) - \sum_{k}^{\min(n_x, n_u)} \cos^2(\phi_k)}$	$\checkmark$	-
$d_d(X,U) = \sqrt{\sum_k^{\min(n_x,n_u)} \phi_k^2}$	-	$\checkmark$
Table 1. Summary of the dissimilarities		

Since  $\max_U \min_i f_i(U)$  is equivalent to  $\max_{U,\tau} \tau$  subject to  $\tau \leq f_i(U)$  for all *i*, (2) is equivalent to:

$$\max_{U,\tau} \tau$$
  
s.t.  $\tau - \sum_{k=1}^{K} u_k^{\top} \check{X}_i \check{X}_i^{\top} u_k \le 0 \qquad \forall i = 1, ..., m$  (3a)

$$u_j^{\top} u_j - 1 = 0 \qquad \qquad \forall j = 1, \dots, K \tag{3b}$$

$$u_j^{\top} u_k = 0$$
  $\forall k \neq j, \ j = 1, ..., K \ ; \ k = 1, ..., K$  (3c)

with  $u_i$  the *i*th column of U. Observe that (3) is an optimization problem with a linear objective function and quadratic (in)equality constraints.

#### **KKT** conditions $\mathbf{2.2}$

We derive the first order necessary conditions of optimality for problem (3). Associating Lagrange multipliers  $\gamma_i$ 's with constraints (3a),  $M_{jj}$ 's with constraints (3b) and  $M_{jk}$ 's with constraints (3c), the KKT conditions, see e.g., [14] can be written as:

$$\sum_{i} \gamma_i = 1 \tag{4a}$$

$$\left(\sum_{i} \gamma_{i} \check{X}_{i} \check{X}_{i}^{\top}\right) U = UM \tag{4b}$$

$$U^{\top}U = I \tag{4c}$$

$$\tau - \operatorname{Tr}\left(U^{\top}\check{X}_{i}\check{X}_{i}^{\top}U\right) \leq 0 \quad \forall i = 1, ..., m$$

$$\gamma_{i} \geq 0 \quad \forall i = 1, ..., m$$

$$(4d)$$

$$(4e)$$

$$\gamma_i \ge 0 \quad \forall i = 1, ..., m \tag{4e}$$

$$\gamma_i \left( \tau - \operatorname{Tr} \left( U^{\top} \check{X}_i \check{X}_i^{\top} U \right) \right) = 0 \quad \forall i = 1, ..., m$$
(4f)

The  $M_{ij}$ 's correspond to the Lagrange multipliers associated with constraints  $u_i^{\top}u_j = 0$  and the  $M_{ii}$ 's to  $u_i^{\top}u_i - 1 = 0$ , so M is symmetric. Therefore there exist a diagonal matrix D and an orthogonal matrix Q such that  $M = QDQ^{\top}$ . We have then  $(\sum_{i} \gamma_i \check{X}_i \check{X}_i^{\top}) UQ = UQD$  which means that UQ is a matrix of eigenvectors of  $\sum_{i} \gamma_i \check{X}_i \check{X}_i^{\dagger}$ . The  $\gamma_i$ 's can be interpreted as the importance given

#### $\mathbf{6}$ E. Renard et al.

to the corresponding subspaces, and are positive only for those subspaces that achieve the max of problem (3).

Let  $U_Y D_Y V_Y^{\top}$  be the singular value decomposition of

$$Y = \left[\sqrt{\gamma_1}\check{X}_1, \sqrt{\gamma_2}\check{X}_2, ..., \sqrt{\gamma_m}\check{X}_m\right] \in R^{p \times N}.$$

Observe that  $U_Y D_Y^2 U_Y^{\top}$  is then an eigendecomposition of  $YY^{\top}$ . A candidate solution of problem (3) would then be, for fixed  $\gamma_i$  respecting condition (4f):

$$M_{ij} = 0 \quad \forall i \neq j \qquad \qquad U = U_Y$$
$$M_{ii} = D_Y^2(i, i) \qquad \qquad \tau = \operatorname{Tr} \left( U^\top Y Y^\top U \right)$$

The last equality results from the combination of conditions (4a) and (4f):

$$\tau = \sum_{i} \gamma_{i} \tau = \sum_{i} \gamma_{i} \operatorname{Tr} \left( U^{\top} X_{i} X_{i}^{\top} U \right).$$

To maximize  $\tau$ , we should consider the K first singular values of Y. The difficulty is then to find  $\gamma_i$  such that condition (4f) is respected.

We can easily see that unless the optimal U belongs to all subspaces  $\mathcal{X}_i$ , more than one  $\gamma_i$  is nonzero. To see this, observe that if  $\gamma_i = 0$  for all  $i \neq j$ , constraint (4b) would imply that U belongs to subspace  $\mathcal{X}_j$ , which means that  $\operatorname{Tr}(U^{\top}\check{X}_{j}\check{X}_{j}^{\top}U) = K \text{ and } \tau = K \text{ by condition (4f). Since for all } i, k \text{ we have}$  $0 \leq u_k^{\top} \check{X}_i \check{X}_i^{\top} u_k \leq 1$  and  $\operatorname{Tr} \left( U^{\top} \check{X}_i \check{X}_i^{\top} U \right) \geq \tau = K$  by condition (4d), we have Tr  $(U^{\top}\check{X}_i\check{X}_i^{\top}U) = K$  for all *i*, and *U* belongs to all the other  $\mathcal{X}_i$ 's. As a result, any candidate solution should have at least two  $X_i$ 's realizing the optimum.

#### 3 Proposed approach

In [9], a fast and simple procedure is proposed to find an approximation of the minimum enclosing ball center of a finite-dimensional Euclidean space. The procedure is extended to arbitrary Riemannian manifolds in [10]:

- Initialize the candidate solution  $U^{(t)}$  with a point in the set Iteratively update as  $U^{(t+1)} = \text{Geodesic}\left(U^{(t)}, X_f^{(t)}, \frac{1}{t+1}\right)$ , where  $X_f^{(t)}$  is the farthest point to  $U^{(t)}$ , and Geodesic(p,q,t) represents the intermediate point m on the geodesic passing through p and q such that dist(p, m) = dist(p, q).

Since we are interested in finding the best subspace of dimension K in  $\mathbb{R}^p$ , our solution U belongs to the Grassmann manifold  $\mathcal{G}(K,p)$ . The main difference with [10] is that we are dealing with points representing subspaces of different dimensions  $n_i$  and therefore belonging to different manifolds  $\mathcal{G}(n_i, p)$ . The first consequence is that the usual Grassmaniann distance cannot be used to determine the farthest point  $X_f^{(t)}$ . Since we want to preserve  $d(U, X_i) = 0$  when  $\mathcal{U} \subset \mathcal{X}_i$ , we used a dissimilarity which is not a metric except if the two subspaces belongs to the same Grassmannian. The second consequence is that to

7

update the current iterate  $U^{(t)}$  using a geodesic,  $X_f^{(t)}$  should be first projected on  $\mathcal{G}(K,p)$ . The next proposition shows how, given  $\mathcal{X}_f \in \mathcal{G}(n_f,p)$  and  $\mathcal{U} \in \mathcal{G}(K,p)$  with  $n_f \geq K$ , we can compute  $\mathcal{Y}_f \in \mathcal{G}(K,p)$  included in  $\mathcal{X}_f$  that minimizes the distance to  $\mathcal{U}$ . We can then update U using the corresponding geodesic.

**Proposition 1** Let  $\mathcal{Y}$ ,  $\mathcal{U} \in \mathcal{G}(K, p)$  and  $\mathcal{X} \in \mathcal{G}(n, p)$  where  $n \geq K$ , with  $\check{X}$  and  $\check{U}$  orthonormal basis of  $\mathcal{X}$  and  $\mathcal{U}$ . Let  $A_1 D_1 B_1^T$  be an SVD of  $\check{U}^T \check{X}$ , then we have

$$\min_{\mathcal{Y}\subset\mathcal{X}} d_a(\mathcal{Y},\mathcal{U}) = d_a(\mathcal{X},\mathcal{U}) = d_a(\operatorname{Col}(XB_1),\mathcal{U}).$$

Those equalities hold also for  $d_d$ .

An adaptation is proposed in Algorithm 1, integrating results obtained from the KKT conditions analysis. We initialize using a K-truncated SVD of  $Y = [\tilde{X}_1, \tilde{X}_2, ..., \tilde{X}_m]$ , corresponding to the case where all the  $\gamma_i$ 's are equal (line 2), and stop when the two farthest subspaces have close dissimilarity values (line 18). As explained in Subsection 2.2, this is a necessary, but not sufficient, condition at optimality. The farthest  $X_i$  from current  $U^{(t)}$  is determined using the chosen dissimilarity based on the principal angles (lines 5 to 8). The associated orthonormal basis  $S_0$  and  $S_1$  of  $\mathcal{U}$  and  $\mathcal{X}_{imax}$  are computed (lines 9 to 11) to update  $U^{(t)}$  in the direction of  $X_{imax}$  with a step  $\frac{1}{t+1}$  along the Grassmannian geodesic [15] (lines 12 to 16).

# **Algorithm 1** Heuristic to extract a subspace minimizing the maximal dissimilarity with the $X_i$ .

```
Require: tol, \epsilon, X_1, ...,X_N
  1 : \ t \gets 1
  1. t \leftarrow 1

2: U^{(0)}D^{(0)}V^{(0)T} \leftarrow \text{SVD}([\check{X}_1...\check{X}_m], K)
  3: \operatorname{err}^{(0)} \leftarrow \operatorname{tol} + 1
  4: while \operatorname{err}^{(t)} \geq \operatorname{tol} \operatorname{do}

5: for all \check{X}_j do

6: d_j^{(t)} \leftarrow \operatorname{da}(U^{(t)}, \check{X}_j)

7: end for
                          i_{\max} \leftarrow \arg\max_{i} d_{i}^{(t)}U_{c}D_{c}V_{c}^{\top} \leftarrow \text{SVD}(\check{X}_{i_{\max}}^{\top}\check{U}^{(t)}, K)S_{0} \leftarrow U^{(t)}V_{c}
  8:
  9:
10:
                            S_1 \leftarrow \check{X}_{i_{\max}} U_c
11:
12:
                             \Theta \leftarrow \arccos \operatorname{diag} D_c
13:
                             \Gamma_{\alpha} \leftarrow \operatorname{diag} \cos \alpha \Theta
                            \Sigma_{\alpha} \leftarrow \operatorname{diag} \sin \alpha \Theta
14:
                         \begin{split} & \Sigma_{\alpha} \leftarrow \operatorname{trag \, sm \, cos} \\ & \delta \leftarrow \frac{1}{t+1} \\ & U^{(t+1)} \leftarrow S_0 \Gamma_{\delta} + (S_1 - S_0 \Gamma_1) \Sigma_1^{-1} \Sigma_{\delta} \\ & d_{\operatorname{sorted}} \leftarrow \operatorname{sort}_{\operatorname{decreasing}}(d^{(t)}) \\ & \operatorname{err}^{(t)} \leftarrow d_{\operatorname{sorted}}(1) - d_{\operatorname{sorted}}(2) \end{split}
15:
16:
17:
18:
19:
                             t \leftarrow t + 1
20: end while
```

## 4 Experiments

We generated synthetic data to represent a case where datasets are unevenly distributed in space and the minimax approach is justified. We first generated a common subspace  $U_c \in \mathbb{R}^{p \times K_c} \sim N(0, 1)$ . We then perturbed it to generate two different noisy versions  $U_j = U_c + N(0, s_j \mu_{U_c}), j \in \{1, 2\}$ , with  $\mu_{U_c} = mean(|U_c|)$ , from which we generated two groups of data. For each  $U_j, j \in \{1, 2\}$ , we generated different datasets  $X_i$ :

$$X_i = \begin{bmatrix} U_j \ A_i \end{bmatrix} \begin{bmatrix} V_i^\top \\ B_i^\top \end{bmatrix}$$

where  $B_i \in \mathbb{R}^{n_i \times K_i}$  is distributed  $\sim U_{[0,1]}$ , and  $A_i \in \mathbb{R}^{p \times K_i} \sim N(0,1)$ . Each column of matrices  $U_j$ ,  $A_i$  and  $B_i$  is normalized (using the  $L_2$  norm) to give the same importance to each component within the dataset. Each column  $V_i(:, j)$  of  $V_i \in \mathbb{R}^{n_i \times K_c}$  is distributed  $\sim U_{[0,\frac{3w_{ij}}{p}]}$ , where  $w_{ij}$  represents the importance of the common component j within dataset i. Finally, Gaussian noise  $\epsilon_i \sim N(0, \sigma_i * \mu_{X_i})$  is added to each datasets:  $X_i \leftarrow X_i + N(0, \sigma_i * \mu_{X_i})$  with  $\mu_{X_i} = mean(|X_i|)$ .

We generated datasets in two groups: the first, based on  $U_1$ , contains more datasets but with higher noise, while the second group, based on  $U_2$ , contains fewer less noisy datasets. The first group contains 17 datasets with  $s_1 = 1$ , while the second contains 3 datasets with  $s_2 = 0.1$ . We took  $K_c = 3$  common components and  $K_i = 5$  additional components, p = 1000 features and  $n_i \sim U_{[20\ 220]}$  samples for each dataset  $X_i$ . The weights  $w_{ij}$  were randomly generated as  $\sim U_{[0.05\ 0.5]}$  to 'hide' the common components in the datasets. The final added noise has  $\sigma_i = 0.1$ .

We compared our Grassmaniann Minimum Enclosing Ball approach GMEB<sub>da</sub> described in Algorithm 1 to a K-truncated SVD on  $X = [X_1...X_n]$  (SVD) and  $\check{X} = [\check{X}_1...\check{X}_n]$  (SVD<sub>o</sub>). Working with  $\check{X}$  instead of X improves the recovery of components that are (weakly) present in all  $X_i$ 's. For each subspace obtained, we computed its maximal dissimilarity to  $\check{X}_i$ , but also to the background truth  $U_c$  and the two noisy  $U_j$ . Mean results on 100 randomly generated datasets are shown on Figure 1, where we also give results when using dissimilarities  $d_b$ ,  $d_c$  or  $d_d$  in Algorithm 1.

When computing dissimilarities to the U's, we logically have  $\sqrt{d_a} = d_b = d_c$ since, in these cases,  $n_x$  and  $n_u$  of Table 1 are equivalent. Results obtained for  $d_b$  and  $d_c$  with  $\check{X}_i$  are similar for all methods, due to the influence of  $n_i$  in the dissimilarities. Since we have  $d_c(U, X_i) \in [\sqrt{n_i - K}, \sqrt{n_i}]$  and  $n_i > K$ , the results are mainly influenced by  $\max_i n_i$ . On the criterion minimized  $(d_a \text{ on } \check{X}_i)$ , the common subspace approach is the best one. As expected,  $SVD_o$  recovers very well the noisy components  $U_1$ , but the common subspace approach recovers better the  $U_2$ . The original  $U_c$  is then recovered better by the subspace approach than by  $SVD_o$ .



Fig. 1. Mean on 100 tests of maximal dissimilarity, for different dissimilarities and methods. Observe that methods  $\text{GMEB}_{da}$  and  $\text{GMEB}_{dd}$  perform best at recovering the ground truth  $U_c$ .

## 5 Conclusion

In this paper, we examined the problem of finding a subspace representative of multiple datasets by minimizing the maximal dissimilarity between this subspace and all the subspaces generated by those datasets. After arguing for a particular choice of dissimilarity measure, we derived some properties of the corresponding formulation. Based on those properties, we proposed an adaptation of an algorithm used for a similar problem on a Riemannian manifold. We then tested the proposed algorithm on synthetic data. Compared to SVD, the subspace recovered by our algorithm is closer to the true common subspace. Based on these promising results, the next step is to analyze properly the convergence of the proposed algorithm. Other approaches to solve the problem should also be investigated, for example based on the KKT conditions or on linearization.

Acknowledgments. Part of this work was performed while the second author was a visiting professor at Université catholique de Louvain.

## References

- Alter, O., Brown, P.O., Botstein, D.: Generalized singular value decomposition for comparative analysis of genome-scale expression data sets of two different organisms. Proceedings of the National Academy of Sciences 100(6) (2003) 3351–3356
- Ponnapalli, S.P., Saunders, M.A., Van Loan, C.F., Alter, O.: A higher-order generalized singular value decomposition for comparison of global mrna expression from multiple organisms. PloS one 6(12) (2011) e28072
- 3. Hotelling, H.: Relations between two sets of variates. Biometrika  ${\bf 28}(3/4)$  (1936) 321–377
- 4. Wold, H.: Partial least squares. Encyclopedia of statistical sciences  ${\bf 6}$  (1985) 581–591
- Meng, C., Kuster, B., Culhane, A.C., Gholami, A.M.: A multivariate approach to the integration of multi-omics datasets. BMC bioinformatics 15(1) (2014) 162
- Hanafi, M., Kohler, A., Qannari, E.M.: Connections between multiple co-inertia analysis and consensus principal component analysis. Chemometrics and intelligent laboratory systems 106(1) (2011) 37–40
- Tenenhaus, A., Tenenhaus, M.: Regularized generalized canonical correlation analysis. Psychometrika 76(2) (2011) 257–284
- Westerhuis, J.A., Kourti, T., MacGregor, J.F.: Analysis of multiblock and hierarchical pca and pls models. Journal of chemometrics 12(5) (1998) 301–321
- Badoiu, M., Clarkson, K.L.: Smaller core-sets for balls. In: Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms, Society for Industrial and Applied Mathematics (2003) 801–802
- Arnaudon, M., Nielsen, F.: On approximating the Riemannian 1-center. Computational Geometry 46(1) (2013) 93–104
- 11. Angulo, J.: Structure tensor image filtering using Riemannian  $L_1$  and  $L_{\infty}$  centerof-mass. Image Analysis & Stereology **33**(2) (2014) 95–105
- Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM journal on Matrix Analysis and Applications 20(2) (1998) 303–353
- Ye, K., Lim, L.H.: Schubert varieties and distances between subspaces of different dimensions. SIAM Journal on Matrix Analysis and Applications 37(3) (2016) 1176–1197
- 14. Nocedal, J., Wright, S.J.: Numerical optimization. Second edn. Springer Series in Operations Research and Financial Engineering. Springer, New York (2006)
- Gallivan, K.A., Srivastava, A., Liu, X., Van Dooren, P.: Efficient algorithms for inferences on grassmann manifolds. In: IEEE Workshop on Statistical Signal Processing. (2003) 315–318