

---

# Second-order difference subspace

---

**Kazuhiro Fukui, Pedro H.V. Valois**

Department of Computer Science, University of Tsukuba, Japan

**Lincon Souza, Takumi Kobayashi**

National Institute of Advanced Industrial Science and Technology (AIST), Japan

## Abstract

Subspace representation is a fundamental technique in various fields of machine learning. Analyzing a geometrical relationship among multiple subspaces is essential for understanding subspace series' temporal and/or spatial dynamics. This paper proposes the second-order difference subspace, a higher-order extension of the first-order difference subspace between two subspaces that can analyze the geometrical difference between them. As a preliminary for that, we extend the definition of the first-order difference subspace to the more general setting that two subspaces with different dimensions have an intersection. We then define the second-order difference subspace by combining the concept of first-order difference subspace and principal component subspace (Karcher mean) between two subspaces, motivated by the second-order central difference method. We can understand that the first/second-order difference subspaces correspond to the velocity and acceleration of subspace dynamics from the viewpoint of a geodesic on a Grassmann manifold. We demonstrate the validity and naturalness of our second-order difference subspace by showing numerical results on two applications: temporal shape analysis of a 3D object and time series analysis of a biometric signal.

## 1 Introduction

Subspace representation appears as a critical component in various types of research in machine learning[1–10]. Depending on the application, subspace representations offer different advantages: a shape subspace allows invariance to rotation for skeleton data in tasks such as hand gesture and action recognition[11, 12]; in signal processing, a signal subspace generated from spectral analysis captures oscillatory components with complex frequency structure while providing invariance to changes in amplitude[9, 13–17]; and in natural language processing, a word subspace conveys the leading concepts in a piece of text[18, 19]; and subspaces can arise in many more applications where the data of interest has a low-dimensional structure represented by a high-dimensional feature space.

Analyzing temporal and/or spatial dynamics of the series of such subspaces is a central technique in machine learning. To construct such an analysis, we introduced the first-order difference subspace to represent the difference between two subspaces as a natural generalization of a difference vector between two vectors [20, 21]. The first-order difference subspace effectively works as valid feature extraction based on the first-order differential operator in various tasks such as shape analysis [8, 20, 21] and anomaly detection from the time series [22]. In this paper, we extend the concept of the first-order difference subspace to the second one and then define it corresponding to the second-order differential operator. By incorporating the second difference subspace, we significantly enhance our subspace analysis framework to advance the research based on subspace representation.

As a preparation for the definition of second-order difference, we first review the mechanism of the first-order difference subspace (DS)[20], defined in the simple case that two subspaces with the same dimension have no intersection. We then extend the definition to a general case in which two

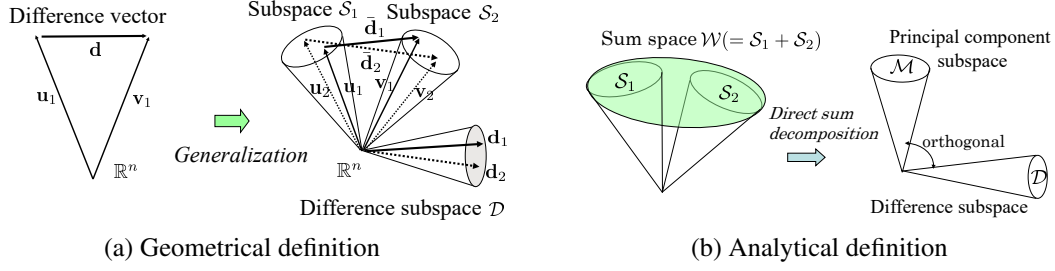


Figure 1: First-order difference subspace  $\mathcal{D}$  between subspaces  $\mathcal{S}_1$  and  $\mathcal{S}_2$  [20].

subspaces with different dimensions intersect. Our definition of the second-order difference subspace is motivated by the second-order central difference method for solving various differential equations. It works as the second-order differential operator corresponding to the first-order difference subspace.

Given three sequential subspaces,  $\mathcal{S}_1$ ,  $\mathcal{S}_2$  and  $\mathcal{S}_3$ , we naturally define the second-order DS as the first-order DS between the subspace  $\mathcal{S}_2$  and the principal component subspace (Karcher mean)  $\mathcal{M}$  between  $\mathcal{S}_1$  and  $\mathcal{S}_3$ . We can understand the physical meaning of this definition by considering a geodesic on a Grassmann manifold, where the first-order and second-order DSs correspond to the velocity and acceleration of a moving subspace, respectively.

We demonstrate the validity and naturalness of our second-order DS on two numerical experiments: temporal analysis of deforming 3D shape and time series analysis of biometric signals. In the first experiment, the 3D shape of a walking/jumping subject is represented by a series of three-dimensional subspaces. Each subspace models the shape at one time step, forming a series as the shape evolves through time [11, 12]. We show that the first/second-order DS can capture the walking action's expected velocity and acceleration. In the second one, a short duration signal is represented by a series of high dimensional subspaces known as signal subspace, each one representing an overlapping window of the signal[9]. The first/second-order DSs also show consistent behaviour in this scenario, although this setting is more complicated because there is a large intersection between sequential signal subspaces.

The contributions of this paper are summarized as follows:

- We extend the definition of the first-order difference subspace to a general case where two subspaces with different dimensions intersect.
- We propose the second-order difference subspace as a natural extension of the first-order difference subspace.
- We demonstrate our definition's validity and physical naturalness through numerical experiments.

The remainder of the paper is organized as follows. In Section 2, we review the definition of first-order DS in the simple case and extend it to a general case. In Section 3, we define the second-order DS. In Section 4, we discuss the geometry of second-order DS. In Section 5, we demonstrate the validity of the first/second-order difference subspaces. We conclude in Section 6.

## 2 First-order difference subspace

First-order difference subspace (DS)  $\mathcal{D}$  between two subspaces,  $\mathcal{S}_1$  and  $\mathcal{S}_2$ , is a generalization of the difference vector  $\mathbf{d}$  between two vectors,  $\mathbf{u}$  and  $\mathbf{v}$ , with unit length, as shown in Fig 1. The first-order DS  $\mathcal{D}(\mathcal{S}_1, \mathcal{S}_2)$  between  $\mathcal{S}_1$  and  $\mathcal{S}_2$  can be defined in both geometrical and analytical viewpoints [20].

The first-order DS is originally defined in a simple case in which there is no intersection subspace between two subspaces with the same dimension [20]. In this paper, we remove this limitation and provide the definition in the more general case that there is a  $r$ -dimensional intersection subspace between two subspaces with different dimensions.

## 2.1 Definitions in the simple case

**Geometrical definition:** Given  $d_1$ -dimensional subspaces  $\mathcal{S}_1$  and  $\mathcal{S}_2$  in  $n$ -dimensional vector space, assuming that the two subspaces have no intersection, we can calculate  $d_1$  canonical angles  $\{\theta_i\}_{i=1}^{d_1}$  between them [23, 24], in an ascending order,  $\theta_i \leq \theta_j \forall i < j$ . Let  $\bar{\mathbf{d}}_i = \mathbf{u}_i - \mathbf{v}_i \in \mathbb{R}^n$  be the difference vector between the canonical vectors  $\mathbf{u}_i \in \mathcal{S}_1$  and  $\mathbf{v}_i \in \mathcal{S}_2$ , which form the  $i$ th canonical angle,  $\theta_i$ , as shown in Fig.1(a). We can regard the normalized difference vectors  $\{\mathbf{d}_i = \frac{\bar{\mathbf{d}}_i}{\|\bar{\mathbf{d}}_i\|}\}_{i=1}^{d_1}$  as the orthonormal basis vectors of the difference subspace  $\mathcal{D}$ , as they are orthogonal to each other [20].

According to the geometrical definition, we can calculate the orthonormal basis of  $\mathcal{D}$  with low computational cost [25]. Let  $\Phi \in \mathbb{R}^{n \times d_1}$  and  $\Psi \in \mathbb{R}^{n \times d_1}$  be orthonormal basis of  $\mathcal{S}_1$  and  $\mathcal{S}_2$ . The canonical angles  $\{\theta_i\}_{i=1}^{d_1}$  can be calculated by performing the SVD as follows:

$$\Phi^\top \Psi = \mathbf{U} \Sigma \mathbf{V}^\top, \quad (1)$$

where  $\mathbf{U} \in \mathbb{R}^{d_1 \times d_1}$ ,  $\Sigma$  is a diagonal matrix  $\in \mathbb{R}^{d_1 \times d_1}$ ,  $\mathbf{V} \in \mathbb{R}^{d_1 \times d_1}$  and  $\text{diag}(\Sigma) = \{\cos \theta_1, \dots, \cos \theta_{d_1}\}$ .

The orthonormal basis  $\mathbf{D}$  of difference subspace  $\mathcal{D}$  is represented in matrix form as follows [25]:

$$\mathbf{D} = (\Phi \mathbf{U} - \Psi \mathbf{V}) \{2(\mathbf{I} - \Sigma)\}^{-\frac{1}{2}}, \quad (2)$$

where  $\Phi \mathbf{U} - \Psi \mathbf{V} = [\bar{\mathbf{d}}_1, \bar{\mathbf{d}}_2, \dots, \bar{\mathbf{d}}_{d_1}] \in \mathbb{R}^{n \times d_1}$  and  $\{2(\mathbf{I} - \Sigma)\}^{-\frac{1}{2}} = \text{diag}(\|\bar{\mathbf{d}}_1\|_2, \dots, \|\bar{\mathbf{d}}_{d_1}\|_2)^{-1} \in \mathbb{R}^{d_1 \times d_1}$ .

As a pair to  $\mathcal{D}$ , we defined the principal component subspace  $\mathcal{M}$  that is spanned by  $\{\mathbf{m}_i = \frac{\bar{\mathbf{m}}_i}{\|\bar{\mathbf{m}}_i\|}\}_{i=1}^{d_1}$ , where  $\bar{\mathbf{m}}_i = \mathbf{u}_i + \mathbf{v}_i \in \mathbb{R}^n$  [20]. The principal component subspace is equal to the Karcher mean between two subspaces  $\mathcal{S}_1$  and  $\mathcal{S}_2$  [25]. The orthonormal basis  $\mathbf{M}$  of the principal component subspace  $\mathcal{M}$  between two subspaces  $\mathcal{S}_1$  and  $\mathcal{S}_2$  is represented in matrix form as follows [25]:

$$\mathbf{M} = (\Phi \mathbf{U} + \Psi \mathbf{V}) \{2(\mathbf{I} + \Sigma)\}^{-\frac{1}{2}}, \quad (3)$$

where  $\Phi \mathbf{U} + \Psi \mathbf{V} = [\bar{\mathbf{m}}_1, \dots, \bar{\mathbf{m}}_{d_1}] \in \mathbb{R}^{n \times d_1}$  and  $\{2(\mathbf{I} + \Sigma)\}^{-\frac{1}{2}} = \text{diag}(\|\bar{\mathbf{m}}_1\|_2, \dots, \|\bar{\mathbf{m}}_{d_1}\|_2)^{-1} \in \mathbb{R}^{d_1 \times d_1}$ .

For the above derivation of  $\mathbf{D}$  and  $\mathbf{M}$ , see A.4 in the appendix.

**Analytical definition:** The difference subspace  $\mathcal{D}$  can also be analytically defined by using the orthogonal projection matrices of the two subspaces  $\mathcal{S}_1$  and  $\mathcal{S}_2$  [24].

The basis vectors of the difference subspace  $\mathcal{D}$  are calculated as the eigenvectors of the sum matrix  $\mathbf{G} = \mathbf{P}_1 + \mathbf{P}_2$  of the projection matrices as follows:

$$\mathbf{G} \mathbf{D} = \mathbf{D} \mathbf{\Lambda}, \quad (4)$$

where  $\mathbf{P}_1 = \Phi \Phi^\top$  and  $\mathbf{P}_2 = \Psi \Psi^\top$ .  $\mathbf{D}$  is the matrix arranging eigenvectors  $\{\mathbf{d}_i\}$  in columns and  $\text{diag}(\mathbf{\Lambda}) = [\lambda_1, \dots, \lambda_{d_1}]$ .

**Definition 2.1** (Difference subspace  $\mathcal{D}$ ).  $d_1$  eigenvectors  $\{\mathbf{d}'_i\}_{i=1}^{d_1}$  of  $\mathbf{G}$  corresponding to the positive eigenvalues smaller than one span the difference subspace  $\mathcal{D}(\mathcal{S}_1, \mathcal{S}_2)$  between  $\mathcal{S}_1$  and  $\mathcal{S}_2$ .

**Definition 2.2** (Principal component subspace (Karcher mean)  $\mathcal{M}$ ).  $d_1$  eigenvectors  $\{\mathbf{m}'_i\}_{i=1}^{d_1}$  of  $\mathbf{G}$  corresponding to eigenvalues larger than one span the principal component subspace  $\mathcal{M}(\mathcal{S}_1, \mathcal{S}_2)$  between  $\mathcal{S}_1$  and  $\mathcal{S}_2$ .

The above definitions correspond that the sum subspace  $\mathcal{W}$  of  $\mathcal{S}_1$  and  $\mathcal{S}_2$  is orthogonally decomposed into  $\mathcal{D}$  and  $\mathcal{M}$  as shown in Fig.1(b). This geometrical relationship means that the difference subspace  $\mathcal{D}$  is generated by subtracting  $\mathcal{M}$  that represents the mean of the two subspaces from the sum subspace  $\mathcal{W}$ ; in other words,  $\mathcal{D}$  can contain only the remaining difference component as its name suggests.

Note that the eigenvectors  $\{\mathbf{d}'_i\}$  and  $\{\mathbf{m}'_i\}$  are equal to  $\{\mathbf{d}_i\}$  and  $\{\mathbf{m}_i\}$  defined in the geometrical definition, respectively. For the details, see [20].

## 2.2 Definitions in general case

We consider more general case that  $d_1$ -dimensional subspace  $\mathcal{S}_1$  and  $d_2$ -dimensional subspace  $\mathcal{S}_2$  have a  $r$ -dimensional intersection subspace in  $n$ -dimensional vector space, for convenience sake,  $d_1 \leq d_2$ .

**Lemma 2.1.** The orthonormal basis  $\Phi \in \mathbb{R}^{n \times d_1}$  of  $\mathcal{S}_1$  and  $\Psi \in \mathbb{R}^{n \times d_2}$  of  $\mathcal{S}_2$  ( $d_1 \leq d_2$ ) can be orthogonally transformed to the following orthonormal bases  $\Phi^* \in \mathbb{R}^{n \times d_1}$  and  $\Psi^* \in \mathbb{R}^{n \times d_2}$ :

$$\Phi^* = [\gamma_1, \dots, \gamma_r, \mathbf{u}_{r+1}, \dots, \mathbf{u}_{d_1}], \quad (5)$$

$$\Psi^* = [\gamma_1, \dots, \gamma_r, \mathbf{v}_{r+1}, \dots, \mathbf{v}_{d_1}, \beta_{d_1+1}, \dots, \beta_{d_2}], \quad (6)$$

where  $\{\gamma_i\}$  represents the orthonormal basis of an intersection subspace  $\mathcal{I}$  between  $\mathcal{S}_1$  and  $\mathcal{S}_2$ ,  $\{\mathbf{u}_i\}$  and  $\{\mathbf{v}_i\}$  are a pair of the canonical vectors forming the  $i$ th nonzero canonical angle  $\theta_i$  between  $\mathcal{S}_1$  and  $\mathcal{S}_2$ .  $\{\beta_{d_1+1}, \dots, \beta_{d_2}\}$  are orthogonal to the remaining orthonormal basis,  $\{\gamma_i\}$ ,  $\{\mathbf{u}_i\}$  and  $\{\mathbf{v}_i\}$ .

For the proof, see Lemma A.1 in the appendix.

**Geometrical definition:** From the above Lemma, only  $d_1 - r$  nonzero canonical angles can be defined between them [23, 24]. The  $r$  remaining canonical angles corresponding to the intersection subspace are zero since the two canonical vectors are the same as  $\{\gamma_i\}_{i=1}^r$ . Accordingly,  $\mathcal{D}$  is defined as a subspace spanned by  $[\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_{d_1-r}]$ , where  $\mathbf{d}_i = \frac{\mathbf{v}_{i+r} - \mathbf{u}_{i+r}}{\|\mathbf{v}_{i+r} - \mathbf{u}_{i+r}\|}$ .  $\mathcal{M}$  is defined as a subspace spanned by  $[\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_{d_1-r}]$ , where  $\mathbf{m}_i = \frac{\mathbf{v}_{i+r} + \mathbf{u}_{i+r}}{\|\mathbf{v}_{i+r} + \mathbf{u}_{i+r}\|}$ . For Eqs.(2) and (3),  $\mathbf{U} \in \mathbb{R}^{d_1 \times (d_1-r)}$ ,  $\Sigma$  is a diagonal matrix  $\in \mathbb{R}^{(d_1-r) \times (d_1-r)}$ ,  $\mathbf{V} \in \mathbb{R}^{(d_1) \times (d_1-r)}$  and  $\text{diag}(\Sigma) = \{\cos \theta_1, \dots, \cos \theta_{d_1-r}\}$ .

**Analytical definition:** Given the sum matrix  $\mathbf{G} = \mathbf{P}_1 + \mathbf{P}_2$ , where  $\mathbf{P}_1 = \Phi \Phi^\top = \Phi^* \Phi^{*\top}$  and  $\mathbf{P}_2 = \Psi \Psi^\top = \Psi^* \Psi^{*\top}$  are the orthogonal projection matrix onto each subspace, respectively. Lemma A.1 leads us the definitions of difference subspace  $\mathcal{D}$  and principal component subspace  $\mathcal{M}$  in the general case. For the proof, see Lemma A.2 in the appendix

**Definition 2.3** (Difference subspace  $\mathcal{D}$ ).  $d_1 - r$  eigenvectors of  $\mathbf{P}_1 + \mathbf{P}_2$  corresponding to nonzero eigenvalues smaller than one span the difference subspace  $\mathcal{D}(\mathcal{S}_1, \mathcal{S}_2)$  between  $\mathcal{S}_1$  and  $\mathcal{S}_2$ .

**Definition 2.4** (Principal component subspace  $\mathcal{M}$ ).  $d_1$  eigenvectors of  $\mathbf{P}_1 + \mathbf{P}_2$  corresponding to eigenvalues larger than one span the principal component subspace  $\mathcal{M}(\mathcal{S}_1, \mathcal{S}_2)$  between  $\mathcal{S}_1$  and  $\mathcal{S}_2$ .

**Definition 2.5** (Intersection subspace  $\mathcal{I}$ ).  $r$  eigenvectors of  $\mathbf{P}_1 + \mathbf{P}_2$ , corresponding to the eigenvalues equal to 2, span the intersection subspace  $\mathcal{I}(\mathcal{S}_1, \mathcal{S}_2)$  between  $\mathcal{S}_1$  and  $\mathcal{S}_2$ .

A subspace  $\mathcal{Z}$  spanned by  $d_2 - d_1$  eigenvectors,  $\{\beta_i\}_{i=d_1}^{d_2}$  in Eq.(6), of  $\mathbf{P}_1 + \mathbf{P}_2$  corresponding to the eigenvalues equal to one are not included in either of  $\mathcal{M}$  and  $\mathcal{D}$ . The sum subspace  $\mathcal{W}$  is orthogonally decomposed into three subspaces, principal component subspace  $\mathcal{M}$ , difference subspace  $\mathcal{D}$  and the subspace  $\mathcal{Z}$  as shown in Fig. 2. The intersection subspace  $\mathcal{I}$  is included in the principal component subspace  $\mathcal{M}$ .

The eigenvectors corresponding to eigenvalues near to one and zero are unstable in the above definitions. Thus, in practice, we use only the eigenvectors corresponding to the eigenvalues smaller than  $1.0 - \delta$  and larger than a given small positive value  $\delta$ , where  $\delta$  is set in for example,  $1e-3 \sim 1e-6$ .

## 2.3 Magnitude of difference subspace

The difference subspace (DS) has information solely about *direction*. Thus, we introduce the magnitude of DS to measure the distance between two subspaces like length of different vector  $\|\mathbf{u} - \mathbf{v}\|_2^2$ .

**Definition 2.6** (Magnitude of difference subspace). The magnitude of DS  $\mathcal{D}$  between two subspaces,  $\mathcal{S}_1$  and  $\mathcal{S}_2$ , is formed by

$$\text{Mag}(\mathcal{D}(\mathcal{S}_1, \mathcal{S}_2)) \stackrel{\text{def}}{=} \text{Trace}(2(\mathbf{I} - \Sigma)) = \sum_{i=1}^{d_1} \|\bar{\mathbf{d}}_i\|_2^2,$$

where  $\Sigma$  is obtained in Eq. (1) and  $\bar{\mathbf{d}}_i$  is the difference vector,  $\mathbf{u}_i - \mathbf{v}_i$  as shown in Fig.1(a). This attribute reflects the hypervolume of a "gap" between the two subspaces.

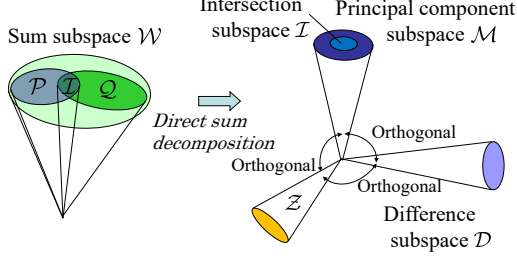


Figure 2: Orthogonal decomposition of sum subspace  $\mathcal{W}$  into three subspaces.

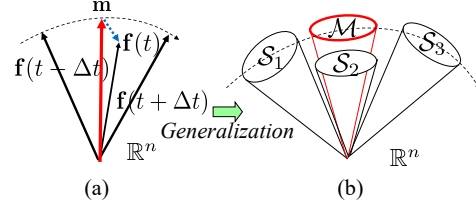


Figure 3: Three sequential vectors/subspaces in  $n$ -dimensional vector space.

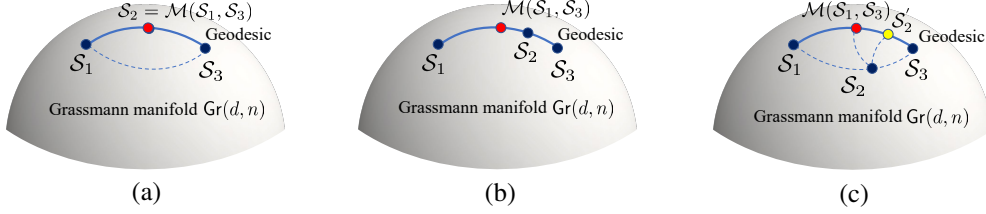


Figure 4: Temporal sequential subspaces,  $S_1$ ,  $S_2$  and  $S_3$  on Grassmann manifold  $\text{Gr}(d, n)$ . (a) Zero second-order DS  $\mathcal{D}^2$  in the case that two subspaces  $S_2$  and  $\mathcal{M}$  coincide completely. (b) Second-order DS  $\mathcal{D}^2$  in the case that  $S_2$  is slightly shifted from  $\mathcal{M}$  along the geodesic. (c) Second-order DS  $\mathcal{D}^2$  in that  $S_2$  is not on the geodesic.

### 3 Second-order difference subspace

We extend the concept of the first-order DS  $\mathcal{D}$  to the second-order DS  $\mathcal{D}^2$ .

#### 3.1 Definition based on second-order central difference

Our extension is motivated by the second-order central difference method for solving various differential equations. Given three sequential vectors  $\mathbf{f}(t - \Delta t)$ ,  $\mathbf{f}(t)$  and  $\mathbf{f}(t + \Delta t)$  in  $n$ -dimensional vector space as shown in Fig.3(a), we can write the second-order central difference as follows:

$$\begin{aligned} \frac{d^2 \mathbf{f}}{dt^2} &\approx \left( \frac{\mathbf{f}(t + \Delta t) - \mathbf{f}(t)}{\Delta t} - \frac{\mathbf{f}(t) - \mathbf{f}(t - \Delta t)}{\Delta t} \right) / \Delta t \\ &= \frac{\mathbf{f}(t + \Delta t) - 2\mathbf{f}(t) + \mathbf{f}(t - \Delta t)}{\Delta t^2} = \frac{2}{\Delta t^2} \left( \frac{\mathbf{f}(t + \Delta t) + \mathbf{f}(t - \Delta t)}{2} - \mathbf{f}(t) \right), \end{aligned} \quad (7)$$

where we assume that  $\Delta t=1$  and the vectors  $\mathbf{f}(t - \Delta t)$ ,  $\mathbf{f}(t)$  and  $\mathbf{f}(t + \Delta t)$  are normalized in length.

Next, given three sequential subspaces  $S_1$ ,  $S_2$  and  $S_3$  in  $n$ -dimensional vector space as shown in Fig.3(b), we correspond the mean vector,  $\mathbf{m} = \frac{\mathbf{f}(t-1) + \mathbf{f}(t+1)}{2}$ , to the Karcher mean (principal component subspace),  $\mathcal{M}(S_1, S_3)$ , representing the mean subspace of  $S_1$  and  $S_3$ , while  $\mathbf{f}(t)$  corresponds to  $S_2$  as well. Then, we correspond the difference vector,  $\mathbf{m} - \mathbf{f}(t)$ , to the difference subspace between  $\mathcal{M}(S_1, S_3)$  and  $S_2$ . Remember we have already generalized the difference vector  $\mathbf{d}$  between two normalized vectors to the difference subspace  $\mathcal{D}$  between two subspaces as described in Fig.1(a). Thus, this correspondence is natural when  $\mathbf{f}(t - 1)$  and  $\mathbf{f}(t + 1)$  are close enough such that  $\|\mathbf{m}\| \approx 1.0$ .

Accordingly, we define the second-order difference subspace  $\mathcal{D}^2$  of  $S_1$ ,  $S_2$  and  $S_3$  as the difference subspace between  $S_2$  and  $\mathcal{M}(S_1, S_3)$ .

**Definition 3.1** (Second-order difference subspace).

$$\mathcal{D}^2(S_1, S_2, S_3) = \mathcal{D}(S_2, \mathcal{M}(S_1, S_3)).$$

Further, we define the magnitude  $\text{Mag}(\mathcal{D}^2)$  of second-order DS according to Definition 2.6.

**Definition 3.2** (Magnitude of second-order difference subspace).

$$\text{Mag}(\mathcal{D}^2(\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3)) = \text{Mag}(\mathcal{D}(\mathcal{S}_2, \mathcal{M}(\mathcal{S}_1, \mathcal{S}_3))). \quad (8)$$

## 4 Mechanism of second-order DS

### 4.1 Representation based on Grassman manifold

Assuming three subspaces  $\mathcal{S}_1, \mathcal{S}_2$  and  $\mathcal{S}_3$  of identical dimension  $d$ , we can understand the geometrical mechanism of the proposed definition more clearly by considering the Grassmann manifold  $\text{Gr}(d, n)$ , where each  $d$ -dimensional subspace in  $n$ -dimensional vector space is represented by a point on  $\text{Gr}(d, n)$  [26].

The geometry of three sequential subspaces in the vector space shown in Fig.3(b) can be visualized as Fig.4(c) on the Grassmann manifold. A local neighborhood of  $n$ -dimensional vectors in  $n$ -dimensional vector space can be identified with Grassmann manifold  $\text{Gr}(1, n)$ . In this sense, our definition of second-order DS can be regarded as a natural generalization of the second-order central difference on  $\text{Gr}(1, n)$  to that on  $\text{Gr}(d, n)$ .

### 4.2 Orthogonal decomposition of magnitude

We introduce a geodesic path  $l(\mathcal{S}_i, \mathcal{S}_j)$  through  $\mathcal{S}_i$  and  $\mathcal{S}_j$ , which is the shortest path between points  $\mathcal{S}_i$  and  $\mathcal{S}_j$  on  $\text{Gr}(d, n)$ .

**Lemma 4.1.** Sum subspace  $\mathcal{W}(\mathcal{S}_i, \mathcal{S}_j)$  is equal to sum subspace  $\mathcal{W}(\{\mathcal{S}^*\})$  where  $\{\mathcal{S}^*\}$  is a set of all the subspaces on the geodesic  $l$  through the points  $\mathcal{S}_i$  and  $\mathcal{S}_j$  on  $\text{Gr}(d, n)$  [25]. The dimensions of  $\mathcal{S}_i, \mathcal{S}_j$  and  $\mathcal{W}(\mathcal{S}_i, \mathcal{S}_j)$  are  $d, d$  and  $2d$ , respectively.

Proof: see Lemma A.3 in the appendix.

The lemma shows that the  $2d$ -dimensional sum subspace  $\mathcal{W}(\mathcal{S}_1, \mathcal{S}_3)$  contains any  $d$ -dimensional subspaces on the geodesic path  $l(\mathcal{S}_1, \mathcal{S}_3)$ . Motivated by this characteristic, we introduce the subspace projection  $\omega(\mathcal{S}_2)$  of  $\mathcal{S}_2$  onto  $\mathcal{W}(\mathcal{S}_1, \mathcal{S}_3)$  for obtaining its projected point (subspace) on  $l(\mathcal{S}_1, \mathcal{S}_3)$ .

**Definition 4.1** (Subspace projection). The subspace projection  $\omega(\mathcal{S})$  of  $d_1$ -dimensional subspace  $\mathcal{S}$  onto  $d_2(\geq d_1)$ -dimensional subspace  $\mathcal{W}$  is defined as follows:

$$\omega(\mathcal{S}) = \underset{\mathcal{S}' \in \mathcal{W}}{\text{argmin}} \rho(\mathcal{S}, \mathcal{S}'), \quad (9)$$

where  $\rho(\mathcal{S}, \mathcal{S}')$  indicates the geodesic distance between  $\mathcal{S}$  and  $\mathcal{S}'$ .

**Lemma 4.2.** The subspace projection  $\omega(\mathcal{S})$  of  $d_1$ -dimensional subspace  $\mathcal{S}$  onto  $d_2(\geq d_1)$ -dimensional subspace  $\mathcal{W}$  can be calculated by using the singular value decomposition (SVD) as follows:

$$\text{SVD} : \mathbf{W}^\top \mathbf{S} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top, \quad (10)$$

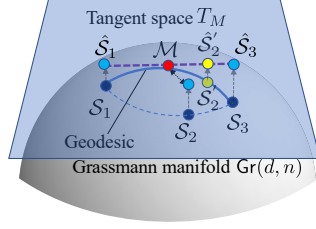
$$\omega(\mathcal{S}) : \mathcal{S} \rightarrow \mathcal{S}' \Rightarrow \mathbf{S} \rightarrow \mathbf{S}' = \mathbf{W} \mathbf{U}, \quad (11)$$

where  $\mathbf{W} \in \mathbb{R}^{n \times d_2}$  and  $\mathbf{S} \in \mathbb{R}^{n \times d_1}$  are the orthonormal basis matrices of  $\mathcal{W}$  and  $\mathcal{S}$ , respectively.

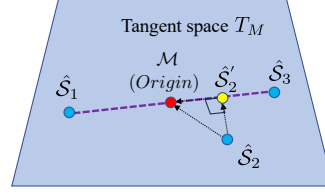
For the proof, see A.5 in the appendix.

On the basis of the subspace projection of  $d$ -dimensional  $\mathcal{S}_2$  onto  $2d$ -dimensional  $\mathcal{W}(\mathcal{S}_1, \mathcal{S}_3)$ , the magnitude of second-order DS  $\text{Mag}(\mathcal{D}^2)$  can be represented as the sum of two components:  $\text{Mag}(\mathcal{D}(\mathcal{S}_2, \mathcal{M}(\mathcal{S}_1, \mathcal{S}_3))) \simeq \text{Mag}(\mathcal{D}(\mathcal{S}_2, \mathcal{W}(\mathcal{S}_1, \mathcal{S}_3))) + \text{Mag}(\mathcal{D}(\omega(\mathcal{S}_2), \mathcal{M}(\mathcal{S}_1, \mathcal{S}_3)))$ , where one component is in the direction orthogonal to the geodesic and another is along the geodesic as shown in Fig.4(c).

It may seem strange that the subspace projection can link the two different operations in Euclidean and non-Euclidean spaces (Grassman manifold). We will provide proof of the validity of the subspace projection and the orthogonal decomposition in the future, although a numerical simulation supports them.



(a) Logarithmic mapping of the related subspaces on the tangent space  $T_M$  of  $Gr(d, n)$ .



(b) Orthogonal decomposition on the tangent space  $T_M$ .

Figure 5: Geometrical relationship of the related subspaces on the  $Gr(d, n)$  and the tangent space  $T_M$  of  $Gr(d, n)$ .

We analyze how the second-order DS works based on the decomposition in the following three cases. Fig.4(a) shows the case that  $\mathcal{S}_2$  and  $\mathcal{M}$  coincide completely. In this case, the magnitude of the second-order DS is zero, representing no acceleration. Fig.4(b) shows the case that  $\mathcal{S}_2$  shifts from  $\mathcal{M}$  along the geodesic on  $Gr(d, n)$ ;  $\mathcal{M}, \mathcal{S}_2 \in \mathcal{W}(\mathcal{S}_1, \mathcal{S}_3)$  in vector space  $\mathbb{R}^n$ . In this case, the magnitude of the second-order DS corresponds to acceleration only along the geodesic, exhibiting  $\text{Mag}(\mathcal{D}(\mathcal{S}_2, \mathcal{W}(\mathcal{S}_1, \mathcal{S}_3))) > 0$  while  $\text{Mag}(\mathcal{D}(\omega(\mathcal{S}_2), \mathcal{W}(\mathcal{S}_1, \mathcal{S}_3))) = 0$ . Fig.4(c) shows the case that  $\mathcal{S}_2$  departs from the geodesic;  $\mathcal{M} \in \mathcal{W}(\mathcal{S}_1, \mathcal{S}_3)$  and  $\mathcal{S}_2 \notin \mathcal{W}(\mathcal{S}_1, \mathcal{S}_3)$  in  $\mathbb{R}^n$ , where the magnitude of second-order DS consists of accelerations in two directions.

We can also introduce the tangent space  $T_M$  of  $Gr(d, n)$  with origin at  $\mathcal{M}$  to understand the mechanism of the orthogonal decomposition.  $T_M$  is useful to analyze the Grassman manifold since it is a vector space with  $\mathcal{M}$  as its origin [27]. Fig. 5(a) shows the points  $\hat{\mathcal{S}}_1$ ,  $\hat{\mathcal{S}}_2$ ,  $\hat{\mathcal{S}}_3$  and  $\hat{\mathcal{S}}'_2$  in the logarithm mapping [26] of  $\mathcal{S}_1$ ,  $\mathcal{S}_2$ ,  $\mathcal{S}_3$  and  $\mathcal{S}'_2$  onto the tangent space  $T_M$ , respectively. In Fig. 5(b), we can also see that the distance between  $\hat{\mathcal{S}}_2$  and  $\mathcal{M}$  can be decomposed into two orthogonal components: the distance between  $\hat{\mathcal{S}}_2$  and  $\hat{\mathcal{S}}'_2$  and that between  $\hat{\mathcal{S}}'_2$  and  $\mathcal{M}$ , where  $\hat{\mathcal{S}}'_2$  indicates the orthogonal projection of  $\hat{\mathcal{S}}_2$  onto the geodesic line in the tangent space.

## 5 Numerical experiments

In this section, we provide a brief numerical study on two applications to demonstrate the validity and naturalness of the definitions of our first/second-order DSs. The first application is a temporal analysis of deforming 3D shapes using a three-dimensional shape subspace, representing the changing pose of a human. This experimental setting is comparatively simple as there is no intersection between the sequential shape subspaces and the dimension of subspaces is only three. The second application is a time sequence analysis of biometric signals using signal subspaces generated by applying the singular spectrum analysis (SSA) to an input signal. This setting is more complicated than the first one in that each signal subspace has a higher dimension than three, and there is an intersection between sequential signal subspaces.

### 5.1 Temporal analysis of deforming 3D shape

In this experiment, we calculate the magnitudes of the first/second-order DSs from the temporal sequence of shape subspaces. In this setting, a 3D pose of a human at time  $t$  is represented by a three-dimensional subspace  $\mathcal{S}_t$  in high-dimensional vector space [12, 11]. A shape subspace can be generated from a set of 3D dot points extracted from a walking or jumping human. The significant merit of the shape subspace-based representation is that it is invariant to any affine transformation against the whole set of 3D points. In other words, the shape subspace is invariant to viewpoints.

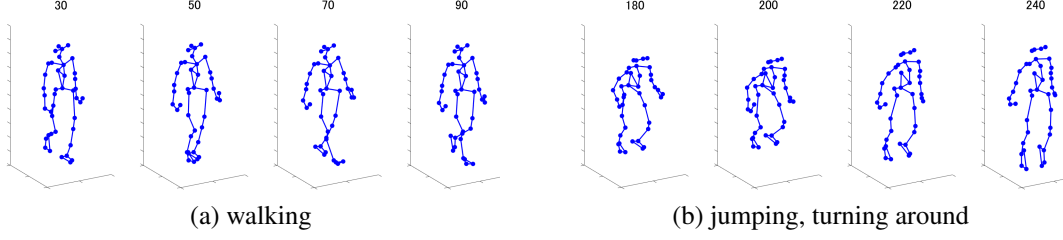


Figure 6: Sequential sets of 3D dots of walking and jumping data. The number of dots is 41. Each number indicates the frame number.

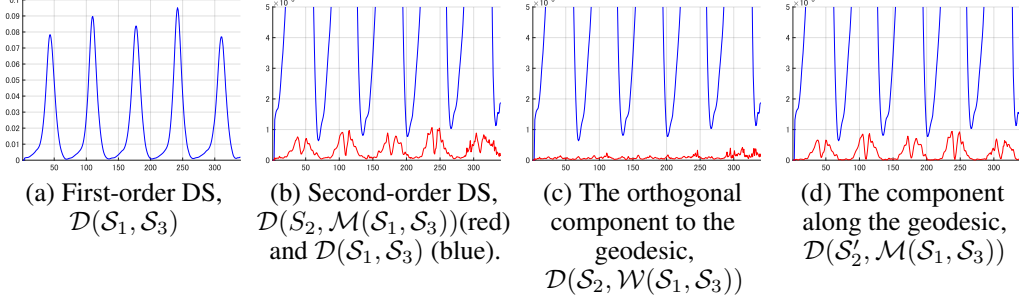


Figure 7: The magnitudes of first/second-order DSs on 3D dots data of walking.

Given  $p$  3D points of  $\{x_i, y_i, z_i\}_{i=1}^p$  of a 3D object (whole human body) at frame time  $t$ , we first generate the following matrix  $\mathbf{V}$ :

$$\mathbf{V} = \begin{pmatrix} x_1 - m_x & y_1 - m_y & z_1 - m_z \\ x_2 - m_x & y_2 - m_y & z_2 - m_z \\ \vdots & \vdots & \vdots \\ x_p - m_x & y_p - m_y & z_p - m_z \end{pmatrix},$$

where  $(m_x, m_y, m_z)$  indicates the center of gravity of all  $p$  points. We then apply the Gram-Schmidt orthogonalization to the column vectors of  $\mathbf{V}$ . The orthogonalized column vectors  $\hat{\mathbf{V}}$  is the basis of shape subspace  $S_t$ . Finally, a task of temporal analysis of 3D pose is converted to measuring the variations of first/second-order DSs between sequential shape subspaces in  $p$ -dimensional space.

## 5.2 Experimental setting

We used two sequential sets of 3D dot data from the CMU Graphics Lab Motion Capture Database [28]. One represents "walking action". Another represents "jumping, turning around". Figs.6(a) and (b) show example sets of 3D dot data of walking and jumping. The number of 3D dot data is 41. Thus, three-dimensional shape subspaces are in 41-dimensional vector space. The total number of walking and jumping frames are 343 and 2701, respectively. We used 3D shape subspaces that extracted every four frames, as the change between successive frames is very small.

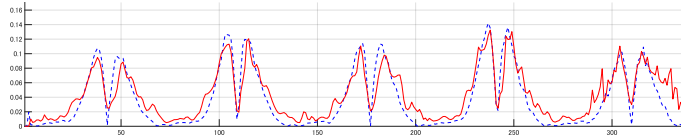


Figure 8: Comparison of the second-order DS output with the absolute value of derivative of the first-order DS's output. The dashed line indicates the derivative of the first-order DS. Both output sequence vectors are normalized.

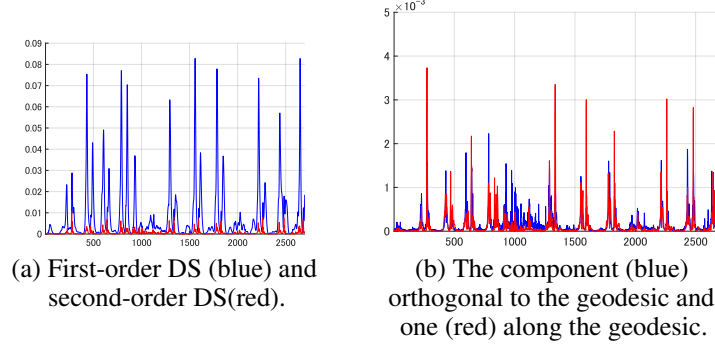


Figure 9: The magnitudes of first/second-order DSs on 3D dots data of jumping, turn around.

### 5.3 Experimental results

Figure 7(a) shows the magnitude of the first-order DS for walking data. We can see that it can capture the periodic variation in velocity during walking. Fig.7(b) shows the comparison between the magnitudes of the first-order DS (blue line) and the second-order DS (red line). The acceleration matches well with the change in velocity. Figs.7(c) and (d) show the two orthogonal components of the magnitude of the second-order DS. The component of  $\mathcal{D}'(\mathcal{S}_2', \mathcal{M}(\mathcal{S}_1, \mathcal{S}_3))$  is dominant in the walking data, while the component of  $\mathcal{D}(\mathcal{S}_2, \mathcal{W}(\mathcal{S}_1, \mathcal{S}_3))$  is small. This difference implies that the primary acceleration occurs along the geodesic. It is known that a movement along a geodesic can be smooth, satisfying the energy minimum principle. In this sense, we may consider walking as a smooth action.

To further confirm the naturalness of the definition of our second-order DS, we compared the absolute value of the derivative of the first-order DS's output with the second-order DS's. Both sequence vectors were normalized. We can see that they have almost the same variation pattern shown in Fig.8. In fact, the normalized correlation coefficient is a high value of 0.948.

Figure 9(a) shows the magnitude of the first-order DS for jumping data. We can see that it captures the periodic variation in the velocity of jumping. Fig.9(b) shows the magnitudes of the first-order DS (blue line) and the second-order DS (red line). Figs.9(c) and (d) show two orthogonal components of the second-order DS's magnitude. The two orthogonal components,  $\mathcal{D}(\omega(\mathcal{S}_2), \mathcal{M}(\mathcal{S}_1, \mathcal{S}_3))$  and  $\mathcal{D}(\mathcal{S}_2, \mathcal{W}(\mathcal{S}_1, \mathcal{S}_3))$  contribute equally to the whole acceleration unlike the result of walking. This result suggests that jumping, unlike walking, may not be regarded as a smooth action that satisfies the energy minimum principle.

### 5.4 Time series analysis of biometric signal

We demonstrate the validity of first/second-order DSs in analyzing one-channel signal [15, 13, 14] based on the framework of the singular spectrum analysis (SSA)[9]. SSA-based method for signal analysis relies on a low dimensional subspace, called signal subspace, generated in one of the steps of SSA as shown in Fig.10. The main advantage of using signal subspace is that it can represent the essential temporal structure of signal data compactly, hence largely reducing the computational cost.

The process flow of the SSA-based method consists of the following four steps, as shown in Fig.10. First, the entire time series data is divided into two parts: past and present time series. Second, two signal subspaces are generated by applying the SSA to the past and present time-series data. Next, the magnitude of the first /second-order DS between the present and past subspaces is measured as the degree of anomaly change. Finally, a specific anomaly change is detected when the anomaly score is larger than a given threshold value.

#### Generation of signal subspace

The signal subspace  $\mathcal{S}_t$  corresponding to a time-series data  $h(t)$  is generated by analyzing the trajectory matrix calculated from the time-series data in the process of the SSA as shown in Fig. 10. Given one-dimensional time series data  $h(t)$ , the corresponding trajectory matrix  $\mathbf{H}_t \in \mathbb{R}^{w \times M}$  is

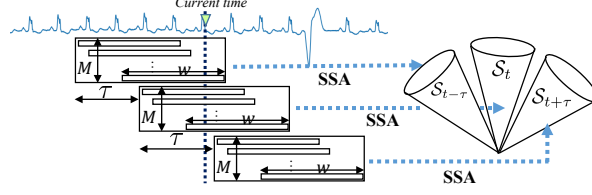


Figure 10: A framework for analyzing signal using SSA. It measures the temporal variations of the first-order DS between  $\mathcal{S}_{t-\tau}$  and  $\mathcal{S}_{t+\tau}$  and the second-order DS of  $\mathcal{S}_{t-\tau}$ ,  $\mathcal{S}_t$  and  $\mathcal{S}_{t+\tau}$ .

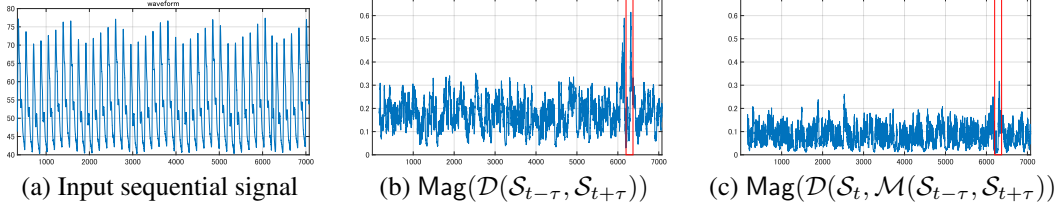


Figure 11: The input signal and its magnitudes of first/second-order DSs.

defined as follows:

$$\mathbf{H}_t = \begin{bmatrix} h(t-w-M+2) & \cdots & h(t-w+1) \\ \vdots & \ddots & \vdots \\ h(t-M+1) & \cdots & h(t) \end{bmatrix}, \quad (12)$$

where  $w$  is the width of a sliding window and  $M$  is the number of the sliding windows as shown in Fig. 10.

To obtain the principal components of  $h(t)$ , SSA solves the eigenvalue problem of  $\mathbf{H}_t \mathbf{H}_t^\top \Phi = \Phi \Sigma$ , where  $\Phi$  is the matrix arranging eigenvectors in columns and  $\Sigma$  is the matrix containing eigenvalues,  $\lambda_1, \dots, \lambda_w$ , in the diagonal elements. The  $d_s$ -dimensional signal subspace  $\mathcal{S}_t$  of the input time series data  $h(t)$  is spanned by the eigenvectors  $\{\phi_i\}_{i=1}^{d_s}$  corresponding to the  $d_s$  largest eigenvalues in  $w$ -dimensional vector space.

## 5.5 Experimental setting

For this experiment, we used the following data from the UCR Time Series Data Mining Archive [29]: 141\_UCR\_Anomaly\_InternalBleeding5\_4000\_6200\_6370, which is shown in Fig.11(a). The parameters,  $M$  and  $w$ , were set to 220 and 100, respectively. The dimension of signal subspaces was 40. Time lag  $\tau$  was set to 16 and  $\delta$  in Sec.2.2 was set to  $1e-4$ .

## 5.6 Experimental results

Figs.11(b) and (c) show the magnitudes of the first/second-order DSs. We can see that both the first and second-order DSs respond to the abnormal term (red rectangle) as expected. Figs.11(d) and (e) show  $\text{Mag}(\mathcal{D}(\mathcal{S}_t, \mathcal{W}(\mathcal{S}_{t-\tau}, \mathcal{S}_{t+\tau})))$  and  $\text{Mag}(\mathcal{D}(\hat{\mathcal{S}}_t, \mathcal{M}(\mathcal{S}_{t-\tau}, \mathcal{S}_{t+\tau})))$ , respectively. The main acceleration occurs along the geodesic, while the acceleration in the direction orthogonal to the geodesic is small, as in the first experiment using walking data. Therefore, we can mention that this biometrical signal (cardiac electrogram) changes smoothly, satisfying the energy minimum principle.

## 6 Conclusions

We proposed the novel concept of second-order difference subspace as an extension of first-order difference subspace. As preliminary to its definition, we defined the first-order difference subspace in a more general setting in which there is an intersection subspace between two subspaces with different dimensions. Given three sequential subspaces, we defined second-order difference subspace as the difference subspace between the principal component subspace between the first and third

subspaces and the middle subspace. The numerical experiments on 3D shape analysis using 3D shape subspace and time series analysis of biomedical signals clearly demonstrated the validity and naturalness of the definition of the second-order DS.

## **Acknowledgments**

We would like to thank Eamonn Keogh and all the other people who have contributed to the UCR Time Series Data Mining Archive. We would like to thank the CMU Graphics Lab Motion Capture Database. The data used in this research was obtained from [mocap.cs.cmu.edu](http://mocap.cs.cmu.edu). The database was created with funding from NSF EIA-0196217.

## References

- [1] Satoshi Watanabe and Nikhil Pakvasa. Subspace method of pattern recognition. In *1st international conference on pattern recognition (ICPR)*, pages 25–32, 1973.
- [2] Keinosuke Fukunaga. *Introduction to statistical pattern recognition*. Academic Press, 1972.
- [3] Erkki Oja. *Subspace methods of pattern recognition*. Research Studies Press, 1983.
- [4] Lance Parsons, Ehtesham Haque, and Huan Liu. Subspace clustering for high dimensional data: a review. *ACM SIGKDD Explorations Newsletter*, 6(1):90–105, 2004.
- [5] Tohru Katayama. *Subspace methods for system identification*. Springer, 2005.
- [6] Osamu Yamaguchi, Kazuhiro Fukui, and Ken-ichi Maeda. Face recognition using temporal image sequence. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 318–323, 1998.
- [7] Tae-Kyun Kim, Josef Kittler, and Roberto Cipolla. Discriminative learning and recognition of image set classes using canonical correlations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29:1005–1018, 2007.
- [8] Kazuhiro Fukui and Osamu Yamaguchi. Face recognition using multi-viewpoint patterns for robot vision. In *11th International Symposium of Robotics Research (ISRR2003)*, pages 192–201, 2003.
- [9] Nina Golyandina and Anatoly Zhigljavsky. *Singular Spectrum Analysis for Time Series*. Springer Berlin Heidelberg, 2020.
- [10] Kazuhiro Fukui. *Subspace Methods*, pages 777–781. Springer, 2014.
- [11] Yosuke Igarashi and Kazuhiro Fukui. 3d object recognition based on canonical angles between shape subspaces. In *10th Asian Conference on Computer Vision, Part IV*, volume 6495 of *Lecture Notes in Computer Science*, pages 580–591, 2010.
- [12] Takao Yoshinuma, Hideitsu Hino, and Kazuhiro Fukui. Personal authentication based on 3d configuration of micro-feature points on facial surface. In *Image and Video Technology - 7th Pacific-Rim Symposium (PSIVT) 2015*, volume 9431 of *Lecture Notes in Computer Science*, pages 433–446, 2015.
- [13] Lincon S. Souza, Bernardo B. Gatto, and Kazuhiro Fukui. Grassmann singular spectrum analysis for bioacoustics classification. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2018*, pages 256–260, 2018.
- [14] Souza, Lincon S. and Gatto, Bernardo B. and Fukui, Kazuhiro. Classification of bioacoustic signals with tangent singular spectrum analysis. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2019*, pages 351–355, 2019.
- [15] Bernardo B. Gatto, Juan G. Colonna, Eulanda M. dos Santos, and Eduardo F. Nakamura. Mutual singular spectrum analysis for bioacoustics classification. In *IEEE International Workshop on Machine Learning for Signal Processing*, pages 1–6, 2017.
- [16] Lincon S. Souza, Takumi Kobayashi, Yasunori Nishimori, Yasuko Sugase-Miyamoto, Kenji Kawano, Shotaro Akaho, and Narihisa Matsumoto. Local distance correlation embedding for time-series analysis on riemannian manifolds. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2024*, pages 5025–5029, 2024.
- [17] Maha Mahyub, Lincon S. Souza, Bojan Batalo, and Kazuhiro Fukui. Signal latent subspace: A new representation for environmental sound classification. *Applied Acoustics*, 225:110181, 2024.
- [18] Erica K. Shimomoto, Lincon Sales de Souza, Bernardo B. Gatto, and Kazuhiro Fukui. Text classification based on word subspace with term-frequency. In *2018 International Joint Conference on Neural Networks (IJCNN) 2018*, pages 1–8, 2018.

- [19] Erica K. Shimomoto, François Portet, and Kazuhiro Fukui. Text classification based on the word subspace representation. *Pattern Anal. Appl.*, 24(3):1075–1093, 2021.
- [20] Kazuhiro Fukui and Atsuto Maki. Difference subspace and its generalization for subspace-based methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(11):2164–2177, 2015.
- [21] Kazuhiro Fukui, Naoya Sogi, Takumi Kobayashi, Jing-Hao Xue, and Atsuto Maki. Discriminant feature extraction by generalized difference subspace. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2):1618–1635, 2023.
- [22] Takumi Kanai, Naoya Sogi, Atsuto Maki, and Kazuhiro Fukui. Time-series anomaly detection based on difference subspace between signal subspaces. *CoRR*, abs/2303.17802, 2023.
- [23] H. Hotelling. Relation between two sets of variables. *Biometrika*, 28:322–377, 1936.
- [24] S. N. Afriat. Orthogonal and oblique projectors and the characteristics of pairs of vector spaces. *Mathematical Proc. Cambridge Philosophical Society*, 53(4):800–816, 1957.
- [25] Takumi Kobayashi, Lincon Souza, and Kazuhiro Fukui. Domain-sum feature transformation for multi-target domain adaptation. In *34th British Machine Vision Conference (BMVC) 2023*, pages 20–24, 2023.
- [26] Thomas Bendokat, Ralf Zimmermann, and P-A Absil. A grassmann manifold handbook: Basic geometry and computational aspects. *Advances in Computational Mathematics*, 50(1):1–51, 2024.
- [27] P.-A. Absil, R. Mahony, and R. Sepulchre. Riemannian Geometry of Grassmann Manifolds with a View on Algorithmic Computation. *Acta Applicandae Mathematicae*, 80(2):199–220, 2004.
- [28] CMU Graphics Lab Motion Capture Database. <http://mocap.cs.cmu.edu/>.
- [29] Eamonn Keogh, Jessica Lin, and Ada Fu. Hot sax: Efficiently finding the most unusual time series subsequence. In *IEEE International Conference on Data Mining*, pages 226–233, 2005.

## A Appendix

**Lemma A.1.** The orthonormal basis  $\Phi \in \mathbb{R}^{n \times d_1}$  of  $\mathcal{S}_1$  and  $\Psi \in \mathbb{R}^{n \times d_2}$  of  $\mathcal{S}_2$  ( $d_1 \leq d_2$ ) can be orthogonally transformed to the following orthonormal bases  $\Phi^* \in \mathbb{R}^{n \times d_1}$  and  $\Psi^* \in \mathbb{R}^{n \times d_2}$ :

$$\Phi^* = [\gamma_1, \dots, \gamma_r, \mathbf{u}_{r+1}, \dots, \mathbf{u}_{d_1}], \quad (13)$$

$$\Psi^* = [\gamma_1, \dots, \gamma_r, \mathbf{v}_{r+1}, \dots, \mathbf{v}_{d_1}, \beta_{d_1+1}, \dots, \beta_{d_2}], \quad (14)$$

where  $\{\gamma_i\}$  represents the orthonormal basis of an intersection subspace  $\mathcal{I}$  between  $\mathcal{S}_1$  and  $\mathcal{S}_2$ ,  $\{\mathbf{u}_i\}$  and  $\{\mathbf{v}_i\}$  are a pair of the canonical vectors forming the  $i$ th nonzero canonical angle  $\theta_i$  between  $\mathcal{S}_1$  and  $\mathcal{S}_2$ .  $\{\beta_{d_1+1}, \dots, \beta_{d_2}\}$  are orthogonal to the remaining orthonormal basis,  $\{\gamma_i\}$ ,  $\{\mathbf{u}_i\}$  and  $\{\mathbf{v}_i\}$ .

*Proof.* We can orthogonally decompose  $\mathcal{S}_1$  into two subspaces: the intersection  $\mathcal{I}$  with  $\mathcal{S}_2$  and non-intersection  $\mathcal{A}_1$  as  $\mathcal{S}_1 = \mathcal{I} \oplus \mathcal{A}_1$ , where  $\mathcal{I} = \text{span}(\Gamma)$  and  $\Gamma = [\gamma_1, \dots, \gamma_r]$ . Thus, an orthonormal basis of  $\mathcal{S}_1$  can be written in the form of a block matrix  $\Phi = [\Gamma \quad \mathbf{A}_1]$  where  $\mathbf{A}_1 \in \mathbb{R}^{n \times (d_1-r)}$  is the orthonormal basis of subspace  $\mathcal{A}_1$ .

Analogously, we can orthogonally decompose  $\mathcal{S}_2$  into three subspaces as  $\mathcal{S}_2 = \mathcal{I} \oplus \mathcal{A}_2 \oplus \mathcal{A}_3$  where  $\mathcal{A}_1$  and  $\mathcal{A}_2$  have the same dimension of  $d_1 - r$ . Hence, an orthonormal basis of  $\mathcal{S}_2$  can be written in the form of a block matrix  $\Psi = [\Gamma \quad \mathbf{A}_2 \quad \mathbf{A}_3]$  where  $\mathbf{A}_2 \in \mathbb{R}^{n \times (d_1-r)}$  and  $\mathbf{A}_3 \in \mathbb{R}^{n \times (d_2-d_1)}$ .

We can focus on only the relationship among  $\mathcal{A}_1$ ,  $\mathcal{A}_2$  and  $\mathcal{A}_3$ , since these subspaces are orthogonal to  $\mathcal{I}$ . Then, we apply SVD to  $\mathbf{A}_1^\top [\mathbf{A}_2 \quad \mathbf{A}_3]$ :

$$\mathbf{A}_1^\top [\mathbf{A}_2 \quad \mathbf{A}_3] = \mathbf{U} [\Sigma \quad \mathbf{0}] [\mathbf{V} \quad \mathbf{V}_\perp]^\top, \quad (15)$$

where  $\mathbf{U} \in \mathbb{R}^{(d_1-r) \times (d_1-r)}$ ,  $\Sigma \in \mathbb{R}^{(d_1-r) \times (d_1-r)}$ ,  $\mathbf{0} \in \mathbb{R}^{(d_1-r) \times (d_2-d_1)}$ ,  $\mathbf{V} \in \mathbb{R}^{(d_2-r) \times (d_1-r)}$  and  $\mathbf{V}_\perp \in \mathbb{R}^{(d_2-r) \times (d_2-d_1)}$ . Note that  $\mathbf{U}^\top \mathbf{U} = \mathbf{V}^\top \mathbf{V} = \mathbf{V}_\perp^\top \mathbf{V}_\perp = \mathbf{I}$  and  $\mathbf{V}^\top \mathbf{V}_\perp = \mathbf{0}$ .

Then, we generate new orthonormal basis as follows:

$$\mathbf{B}_1 = \mathbf{A}_1 \mathbf{U}, \quad (16)$$

$$\mathbf{B}_2 = [\mathbf{A}_2 \quad \mathbf{A}_3] \mathbf{V}, \quad (17)$$

$$\mathbf{Z} = [\mathbf{A}_2 \quad \mathbf{A}_3] \mathbf{V}_\perp, \quad (18)$$

where  $\mathbf{B}_1 \in \mathbb{R}^{n \times (d_1-r)}$  and  $\mathbf{B}_2 \in \mathbb{R}^{n \times (d_1-r)}$  are the pairs of the canonical vectors forming the canonical angles  $\{\theta_i\}_{i=r+1}^{d_1}$ , as  $\mathbf{B}_1^\top \mathbf{B}_2 = \Sigma = \text{diag}(\Theta)$ . The size of  $\mathbf{Z}$  is  $n$  by  $d_2 - d_1$ .

It follows from  $\Gamma^\top \mathbf{A}_1 = \Gamma^\top \mathbf{A}_2 = \Gamma^\top \mathbf{A}_3 = \mathbf{0}$  that  $\Gamma^\top \mathbf{B}_1 = \Gamma^\top \mathbf{B}_2 = \Gamma^\top \mathbf{Z} = \mathbf{0}$ .

Further, we show that  $\mathbf{B}_2^\top \mathbf{Z} = \mathbf{0}$  as follows:

$$\mathbf{B}_2^\top \mathbf{Z} = \mathbf{V}^\top \begin{bmatrix} \mathbf{A}_2^\top \\ \mathbf{A}_3^\top \end{bmatrix} [\mathbf{A}_2 \quad \mathbf{A}_3] \mathbf{V}_\perp, \quad (19)$$

$$= \mathbf{V}^\top \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \mathbf{V}_\perp, \quad (20)$$

$$= \mathbf{V}^\top \mathbf{V}_\perp, \quad (21)$$

$$= \mathbf{0}. \quad (22)$$

Consequently, we obtain the orthonormal basis  $\Phi^*$  and  $\Psi^*$  as follows:

$$\Phi^* = [\Gamma \quad \mathbf{B}_1], \quad (23)$$

$$\Psi^* = [\Gamma \quad \mathbf{B}_2 \quad \mathbf{Z}], \quad (24)$$

where  $\mathbf{B}_1 = [\mathbf{u}_{r+1}, \dots, \mathbf{u}_{d_1}]$ ,  $\mathbf{B}_2 = [\mathbf{v}_{r+1}, \dots, \mathbf{v}_{d_1}]$  and  $\mathbf{Z} = [\beta_{d_1+1}, \dots, \beta_{d_2}]$ .

Next, we show that  $\mathbf{B}_1^\top \mathbf{Z} = \mathbf{0}$  as follows:

From Eq.(15), we obtain

$$\mathbf{U}^\top \mathbf{A}_1^\top [\mathbf{A}_2 \quad \mathbf{A}_3] [\mathbf{V} \quad \mathbf{V}_\perp] = [\Sigma \quad \mathbf{0}]. \quad (25)$$

Hence,

$$\mathbf{U}^\top \mathbf{A}_1^\top [\mathbf{A}_2 \quad \mathbf{A}_3] \mathbf{V}_\perp = \mathbf{0}. \quad (26)$$

$\mathbf{B}_1^\top \mathbf{Z}$  is written as follows:

$$\mathbf{B}_1^\top \mathbf{Z} = \mathbf{U}^\top \mathbf{A}_1^\top [\mathbf{A}_2 \quad \mathbf{A}_3] \mathbf{V}_\perp. \quad (27)$$

$$(28)$$

Therefore,

$$\mathbf{B}_1^\top \mathbf{Z} = \mathbf{U}^\top \mathbf{A}_1^\top [\mathbf{A}_2 \quad \mathbf{A}_3] \mathbf{V}_\perp = \mathbf{0}. \quad (29)$$

Finally, we conclude that  $\mathbf{\Gamma} \mathbf{Z}^\top = \mathbf{B}_1 \mathbf{Z}^\top = \mathbf{B}_2 \mathbf{Z}^\top = \mathbf{0}$ , implying that the three subspaces, corresponding to  $\mathbf{B}_1$ ,  $\mathbf{B}_2$  and  $\mathbf{Z}$ , are orthogonal to each other.  $\square$

**Lemma A.2.** Given two subspaces  $\mathcal{S}_1$  and  $\mathcal{S}_2$  such that they have an intersection. Let  $\Phi^*$  and  $\Psi^*$  be the orthonormal basis for each subspace, as defined in the previous lemma, and let  $\mathbf{P}_1 = \Phi \Phi^\top = \Phi^* \Phi^{*\top}$  and  $\mathbf{P}_2 = \Psi \Psi^\top = \Psi^* \Psi^{*\top}$  be the orthogonal projection matrix onto each subspace. Let the sum matrix  $\mathbf{G} = \mathbf{P}_1 + \mathbf{P}_2$ . The principal component subspace  $\mathcal{M}$  is spanned by the eigenvectors of  $\mathbf{G}$  corresponding to the  $d_1$  larger eigenvalues than 1. The difference subspace  $\mathcal{D}$  is spanned by the eigenvectors of  $\mathbf{G}$  corresponding to  $d_1 - r$  non-negative eigenvalues smaller than 1.

*Proof.* We again consider the non-intersection subspaces  $\mathcal{A}_1$  and  $\mathcal{A}_2$  and calculate the sum of their projection matrices as:

$$\mathbf{J} = \mathbf{B}_1 \mathbf{B}_1^\top + \mathbf{B}_2 \mathbf{B}_2^\top, \quad (30)$$

$$\mathbf{J} \mathbf{H} = \mathbf{H} \mathbf{S}. \quad (31)$$

By the definitions of  $\mathbf{B}_1$  and  $\mathbf{B}_2$ , we can derive the projection matrices of  $\mathcal{S}_1$  and  $\mathcal{S}_2$ , and then  $\mathbf{G}$ :

$$\Phi^* \Phi^{*\top} = \mathbf{\Gamma} \mathbf{\Gamma}^\top + \mathbf{B}_1 \mathbf{B}_1^\top, \quad (32)$$

$$\Psi^* \Psi^{*\top} = \mathbf{\Gamma} \mathbf{\Gamma}^\top + \mathbf{B}_2 \mathbf{B}_2^\top + \mathbf{Z} \mathbf{Z}^\top, \quad (33)$$

$$\mathbf{G} = 2\mathbf{\Gamma} \mathbf{\Gamma}^\top + \mathbf{J} + \mathbf{Z} \mathbf{Z}^\top, \quad (34)$$

where  $\mathbf{\Gamma} \mathbf{\Gamma}^\top$ ,  $\mathbf{Z} \mathbf{Z}^\top$  and  $\mathbf{J}$  are orthogonal to each other. Thus, the SVD of  $\mathbf{G}$  is of the following form:

$$\mathbf{G} = [\mathbf{\Gamma} \quad \mathbf{Z} \quad \mathbf{H}] \begin{bmatrix} 2\mathbf{I} & & \\ & \mathbf{I} & \\ & & \mathbf{S} \end{bmatrix} [\mathbf{\Gamma} \quad \mathbf{Z} \quad \mathbf{H}]^\top. \quad (35)$$

The intersection  $\mathcal{I}$  is spanned by the leading eigenvectors with value 2. The subspace  $\mathcal{Z}$  is spanned by the eigenvectors with value 1. The remainder are then the same eigenvalues of  $\mathbf{J}$ . Since  $\mathbf{J}$  is the sum of the two non-intersected subspaces, they follow the simple case definition of principal/difference subspaces, such that the principal component subspace  $\mathcal{M}$  is spanned by the eigenvectors corresponding to the  $d_1$  larger eigenvalues than 1 and the difference subspace  $\mathcal{D}$  is spanned by the  $d_1 - r$  non-negative eigenvalues smaller than 1.  $\square$

**Lemma A.3.** Sum subspace  $\mathcal{W}(\mathcal{S}_1, \mathcal{S}_2)$  is equal to  $\mathcal{W}(\{\mathcal{S}^*\})$  where  $\{\mathcal{S}^*\}$  is a set of all the subspaces on the geodesic  $l$  between two points,  $\mathcal{S}_1$  and  $\mathcal{S}_2$  on  $\text{Gr}(d, n)$ .

*Proof.* The projection matrix  $\mathbf{G}$  onto  $\mathcal{W}(\{\mathcal{S}^*\})$  can be written as the integral of all projection matrices of subspaces  $\mathcal{S}^*$  along the geodesic:

$$\mathbf{G} = \int_0^1 \hat{\mathbf{I}}(t) \hat{\mathbf{I}}(t)^\top dt. \quad (36)$$

In addition, the projection matrix  $\mathbf{H}$  onto  $\mathcal{W}(\mathcal{S}_1, \mathcal{S}_2)$  can be written simply as

$$\mathbf{H} = \Phi \Phi^\top + \Psi \Psi^\top. \quad (37)$$

As shown in [25], the projections matrices can be expressed as the following eigendecompositions:

$$\mathbf{G} = [\mathbf{M}, \mathbf{D}] \begin{bmatrix} \mathbf{\Lambda}_+ & \\ & \mathbf{\Lambda}_- \end{bmatrix} [\mathbf{M}, \mathbf{D}]^\top, \quad (38)$$

$$\mathbf{H} = [\mathbf{M}, \mathbf{D}] \begin{bmatrix} \mathbf{S}_+ & \\ & \mathbf{S}_- \end{bmatrix} [\mathbf{M}, \mathbf{D}]^\top, \quad (39)$$

where  $\mathbf{M}, \mathbf{D}$  are orthogonal bases for the principal and difference and  $\mathbf{\Lambda}_\pm, \mathbf{S}_\pm$  eigenvalue matrices, written as

$$\mathbf{M} = \text{colnorm}(\Phi\mathbf{U} + \Psi\mathbf{V}) \in \mathbb{R}^{n \times d}, \quad (40)$$

$$\mathbf{D} = \text{colnorm}(\Phi\mathbf{U} - \Psi\mathbf{V}) \in \mathbb{R}^{n \times d}, \quad (41)$$

$$\mathbf{\Lambda}_\pm = \text{diag}(\{1 \pm \text{sinc } \theta_k\}_{k=1}^d), \quad (42)$$

$$\mathbf{S}_\pm = \text{diag}(\{1 \pm \cos \theta_k\}_{k=1}^d). \quad (43)$$

Since their eigenvectors are exactly the same, it must be that their span is the same, proving our claim.  $\square$

**Lemma A.4** (The orthonormal basis for  $\mathcal{D}$  and  $\mathcal{M}$ ). Give two subspaces  $d_1$ -dimensional  $\mathcal{S}_1$  and  $\mathcal{S}_2$  such that there is no intersection between them. Let  $\Phi$  and  $\Psi$  be the orthonormal basis of  $\mathcal{S}_1$  and  $\mathcal{S}_2$ , respectively. The svd of  $\Phi^\top \Psi$  is written as  $\mathbf{U}\Sigma\mathbf{V}^\top$ . The orthonormal basis  $\mathbf{D}$  and  $\mathbf{M}$  of difference subspace  $\mathcal{D}$  and principal component subspace  $\mathcal{M}$  between the two subspaces  $\mathcal{S}_1$  and  $\mathcal{S}_2$  are written in matrix form as follows:

$$\mathbf{D} = (\Phi\mathbf{U} - \Psi\mathbf{V})\{2(\mathbf{I} - \Sigma)\}^{-\frac{1}{2}}, \quad (44)$$

$$\mathbf{M} = (\Phi\mathbf{U} + \Psi\mathbf{V})\{2(\mathbf{I} + \Sigma)\}^{-\frac{1}{2}}. \quad (45)$$

*Proof.* Each difference vector between a pair of canonical vectors can be written as follows:

$$[\mathbf{u}_1 - \mathbf{v}_1, \mathbf{u}_2 - \mathbf{v}_2, \dots, \mathbf{u}_{d_1} - \mathbf{v}_{d_1}] = [\bar{\mathbf{d}}_1, \bar{\mathbf{d}}_2, \dots, \bar{\mathbf{d}}_{d_1}], \quad (46)$$

$$= \Phi\mathbf{U} - \Psi\mathbf{V} \in \mathbb{R}^{n \times d_1}. \quad (47)$$

Similarly, each principal component vector between a pair of canonical vectors can be written as follows:

$$[\mathbf{u}_1 + \mathbf{v}_1, \mathbf{u}_2 + \mathbf{v}_2, \dots, \mathbf{u}_{d_1} + \mathbf{v}_{d_1}] = [\bar{\mathbf{m}}_1, \bar{\mathbf{m}}_2, \dots, \bar{\mathbf{m}}_{d_1}], \quad (48)$$

$$= \Phi\mathbf{U} + \Psi\mathbf{V} \in \mathbb{R}^{n \times d_1}. \quad (49)$$

Using the cosine theorem, we obtain  $\|\bar{\mathbf{d}}_i\|_2^2$  and  $\|\bar{\mathbf{m}}_i\|_2^2$  as follows:

$$\|\bar{\mathbf{d}}_i\|_2^2 = \|\mathbf{u}_i\|_2^2 + \|\mathbf{v}_i\|_2^2 - 2\|\mathbf{u}_i\|_2\|\mathbf{v}_i\|_2 \cos \theta_i, \quad (50)$$

$$= 1 + 1 - 2 \cos \theta_i, \quad (51)$$

$$= 2(1 - \cos \theta_i). \quad (52)$$

$$\|\bar{\mathbf{m}}_i\|_2^2 = \|\mathbf{u}_i\|_2^2 + \|\mathbf{v}_i\|_2^2 - 2\|\mathbf{u}_i\|_2\|\mathbf{v}_i\|_2 \cos(\pi - \theta_i), \quad (53)$$

$$= 1 + 1 + 2 \cos \theta_i, \quad (54)$$

$$= 2(1 + \cos \theta_i). \quad (55)$$

Hence,

$$\text{diag}(\|\bar{\mathbf{d}}_1\|_2, \|\bar{\mathbf{d}}_2\|_2, \dots, \|\bar{\mathbf{d}}_{d_1}\|_2) = (2(\mathbf{I}_{d_1 \times d_1} - \cos(\Theta)))^{\frac{1}{2}}, \quad (56)$$

$$\text{diag}(\|\bar{\mathbf{m}}_1\|_2, \|\bar{\mathbf{m}}_2\|_2, \dots, \|\bar{\mathbf{m}}_{d_1}\|_2) = (2(\mathbf{I}_{d_1 \times d_1} + \cos(\Theta)))^{\frac{1}{2}}, \quad (57)$$

where  $\Theta = \text{diag}(\theta_1, \theta_2, \dots, \theta_{d_1})$  and  $\cos(\Theta) = \Sigma$ .

According to the definitions of  $\mathcal{D}$  and  $\mathcal{M}$ , we conclude that

$$\mathbf{D} = [\bar{\mathbf{d}}_1, \dots, \bar{\mathbf{d}}_{d_1}] \text{diag}(\|\bar{\mathbf{d}}_1\|_2, \dots, \|\bar{\mathbf{d}}_{d_1}\|_2)^{-1} = (\Phi\mathbf{U} - \Psi\mathbf{V})\{2(\mathbf{I} - \Sigma)\}^{-\frac{1}{2}}, \quad (58)$$

$$\mathbf{M} = [\bar{\mathbf{m}}_1, \dots, \bar{\mathbf{m}}_{d_1}] \text{diag}(\|\bar{\mathbf{m}}_1\|_2, \dots, \|\bar{\mathbf{m}}_{d_1}\|_2)^{-1} = (\Phi\mathbf{U} + \Psi\mathbf{V})\{2(\mathbf{I} + \Sigma)\}^{-\frac{1}{2}}. \quad (59)$$

$\square$

**Definition A.1** (Subspace projection). The subspace projection  $\omega(\mathcal{S})$  of  $d_1$ -dimensional subspace  $\mathcal{S}$  onto  $d_2(\geq d_1)$ -dimensional subspace  $\mathcal{W}$  is defined as follows:

$$\omega(\mathcal{S}) = \underset{\mathcal{S}' \in \mathcal{W}}{\operatorname{argmin}} \rho(\mathcal{S}, \mathcal{S}'), \quad (60)$$

where  $\rho(\mathcal{S}, \mathcal{S}')$  indicates the geodesic distance between  $\mathcal{S}$  and  $\mathcal{S}'$ .

**Lemma A.5.** The subspace projection  $\omega(\mathcal{S})$  of  $d_1$ -dimensional subspace  $\mathcal{S}$  onto  $d_2(\geq d_1)$ -dimensional subspace  $\mathcal{W}$  can be calculated by using the singular value decomposition (SVD) as flows:

$$\text{SVD} : \mathbf{W}^\top \mathbf{S} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top, \quad (61)$$

$$\omega(\mathcal{S}) : \mathcal{S} \rightarrow \mathcal{S}' \Rightarrow \mathbf{S} \rightarrow \mathbf{S}' = \mathbf{W} \mathbf{U}, \quad (62)$$

where  $\mathbf{W} \in \mathbb{R}^{n \times d_2}$  and  $\mathbf{S} \in \mathbb{R}^{n \times d_1}$  are the orthonormal basis matrices of  $\mathcal{W}$  and  $\mathcal{S}$ , respectively.

*Proof.* Eq.(60) can be written as a maximization problem of the following objective function  $f$  with respect to the canonical correlation between  $\mathcal{W}$  and  $\mathcal{S}$ :

$$f(\mathbf{A}, \mathbf{B}) = \underset{\mathbf{a}_i, \mathbf{b}_i}{\operatorname{maximize}} \sum_{i=1}^{d_1} (\mathbf{W} \mathbf{a}_i)^\top \mathbf{S} \mathbf{b}_i, \quad \text{subject to } \|\mathbf{a}_i\| = \|\mathbf{b}_i\| = 1, \quad (63)$$

where  $\mathbf{a}_i \in \mathbb{R}^{d_2}$  and  $\mathbf{b}_i \in \mathbb{R}^{d_1}$  are the weight coefficient vectors on each base.  $\mathbf{W} \mathbf{a}_i$  is a vector in  $\mathcal{W}$  and  $\mathbf{S} \mathbf{b}_i$  is a vector in  $\mathcal{S}$ .

We rewrite the above as follows:

$$f(\mathbf{A}, \mathbf{B}) = \underset{\mathbf{A}, \mathbf{B}}{\operatorname{maximize}} (\mathbf{W} \mathbf{A})^\top (\mathbf{S} \mathbf{B}), \quad \text{subject to } \|\mathbf{a}_i\| = \|\mathbf{b}_i\| = 1, \quad (64)$$

where  $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_{d_1}]$  and  $\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_{d_1}]$ .

Using the Lagrangian undetermined multiplier method, we obtain the following objective function:

$$f(\mathbf{A}, \mathbf{B}) = \operatorname{tr} \left( (\mathbf{W} \mathbf{A})^\top \mathbf{S} \mathbf{B} \right) - \operatorname{tr} \left( ((\mathbf{W} \mathbf{A})^\top \mathbf{W} \mathbf{A} - \mathbf{I}) \mathbf{\Sigma}_1 \right) - \operatorname{tr} \left( ((\mathbf{S} \mathbf{B})^\top \mathbf{S} \mathbf{B} - \mathbf{I}) \mathbf{\Sigma}_2 \right), \quad (65)$$

$$= \operatorname{tr} (\mathbf{A}^\top \mathbf{W}^\top \mathbf{S} \mathbf{B}) - \operatorname{tr} ((\mathbf{A}^\top \mathbf{W}^\top \mathbf{W} \mathbf{A} - \mathbf{I}) \mathbf{\Sigma}_1) - \operatorname{tr} ((\mathbf{B}^\top \mathbf{S}^\top \mathbf{S} \mathbf{B} - \mathbf{I}) \mathbf{\Sigma}_2), \quad (66)$$

where  $\mathbf{\Sigma}_1 \in \mathbb{R}^{d_1 \times d_1}$  and  $\mathbf{\Sigma}_2 \in \mathbb{R}^{d_1 \times d_1}$  are the diagonal matrices representing undetermined multipliers.

To find  $\mathbf{A}$  and  $\mathbf{B}$  for maximizing the objective function  $f$ , we differentiate  $f(\mathbf{A}, \mathbf{B})$  with respect to  $\mathbf{A}$  and  $\mathbf{B}$  and set the derivatives to zeros as follows:

$$\frac{\partial f}{\partial \mathbf{A}} = \mathbf{W}^\top \mathbf{S} \mathbf{B} - \mathbf{W}^\top \mathbf{W} \mathbf{A} \mathbf{\Sigma}_1 = \mathbf{0}, \quad (67)$$

$$\frac{\partial f}{\partial \mathbf{B}} = \mathbf{S}^\top \mathbf{W} \mathbf{A} - \mathbf{S}^\top \mathbf{S} \mathbf{B} \mathbf{\Sigma}_2 = \mathbf{0}. \quad (68)$$

From Eqs.(67) and (68),

$$\mathbf{W}^\top \mathbf{S} \mathbf{B} = \mathbf{W}^\top \mathbf{W} \mathbf{A} \mathbf{\Sigma}_1 = \mathbf{A} \mathbf{\Sigma}_1, \quad (69)$$

$$\mathbf{S}^\top \mathbf{W} \mathbf{A} = \mathbf{S}^\top \mathbf{S} \mathbf{B} \mathbf{\Sigma}_2 = \mathbf{B} \mathbf{\Sigma}_2. \quad (70)$$

Further, from Eq.(70),

$$\mathbf{B} = \mathbf{S}^\top \mathbf{W} \mathbf{A} \mathbf{\Sigma}_2^{-1}. \quad (71)$$

By substituting the above equation into Eq.(69), we obtain

$$\mathbf{W}^\top \mathbf{S} \mathbf{S}^\top \mathbf{W} \mathbf{A} \mathbf{\Sigma}_2^{-1} = \mathbf{A} \mathbf{\Sigma}_1, \quad (72)$$

$$\mathbf{W} \mathbf{W}^\top \mathbf{S} \mathbf{S}^\top \mathbf{W} \mathbf{A} = \mathbf{W} \mathbf{A} \mathbf{\Sigma}_1 \mathbf{\Sigma}_2, \quad (73)$$

$$\mathbf{P}_1 \mathbf{P}_2 (\mathbf{W} \mathbf{A}) = (\mathbf{W} \mathbf{A}) \hat{\mathbf{\Sigma}}, \quad (74)$$

where  $\mathbf{P}_1$  and  $\mathbf{P}_2$  are the orthogonal projection matrices defined as  $\mathbf{W}\mathbf{W}^\top$  and  $\mathbf{S}\mathbf{S}^\top$ , respectively, and  $\widehat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2$ . The last equation implies that  $\mathbf{W}\mathbf{A}$  represents a set of eigenvectors of  $\mathbf{P}_1\mathbf{P}_2$ .

Next, we show that  $\mathbf{W}\mathbf{U}$  also represents a set of the eigenvectors of  $\mathbf{P}_1\mathbf{P}_2$ .

$$\mathbf{P}_1\mathbf{P}_2 = \mathbf{W}\mathbf{W}^\top\mathbf{S}\mathbf{S}^\top, \quad (75)$$

$$(76)$$

By substituting  $\mathbf{W}^\top\mathbf{S} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$  into the above equation, we obtain the following:

$$\mathbf{P}_1\mathbf{P}_2 = \mathbf{W}\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top\mathbf{S}^\top. \quad (77)$$

$$(78)$$

By postmultiplying both sides of the above equation by  $\mathbf{W}$ , we obtain the following:

$$\mathbf{P}_1\mathbf{P}_2\mathbf{W} = \mathbf{W}\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top\mathbf{S}^\top\mathbf{W}, \quad (79)$$

$$= \mathbf{W}\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top\mathbf{V}\boldsymbol{\Sigma}\mathbf{U}^\top, \quad (80)$$

$$= \mathbf{W}\mathbf{U}\boldsymbol{\Sigma}^2\mathbf{U}^\top, \quad (81)$$

$$\mathbf{P}_1\mathbf{P}_2(\mathbf{W}\mathbf{U}) = (\mathbf{W}\mathbf{U})\boldsymbol{\Sigma}^2. \quad (82)$$

Equation (82) denotes that  $\mathbf{W}\mathbf{U}$  also represents a set of eigenvectors of  $\mathbf{P}_1\mathbf{P}_2$ . Thus, as  $\mathbf{W}\mathbf{U} = \mathbf{W}\mathbf{A}$ , we conclude that  $\mathbf{W}\mathbf{U}$  represents the orthonormal basis of the closest subspace  $\mathcal{S}' \in \mathcal{W}$  to  $\mathcal{S}$ . Consequently, the following relationship can be held:

$$\omega(\mathcal{S}) = \underset{\mathcal{S}' \in \mathcal{W}}{\operatorname{argmin}} \rho(\mathcal{S}, \mathcal{S}'). \quad (83)$$

It is worth to confirm that  $\mathbf{S}\mathbf{V}$  represents a set of eigenvectors of  $\mathbf{P}_2\mathbf{P}_1$ .

From Eq.(69), we obtain

$$\mathbf{A} = \mathbf{W}^\top\mathbf{S}\mathbf{B}\boldsymbol{\Sigma}_1^{-1}. \quad (84)$$

By substituting the above equation into Eq.(70), we obtain

$$\mathbf{S}^\top\mathbf{W}\mathbf{W}^\top\mathbf{S}\mathbf{B}\boldsymbol{\Sigma}_1^{-1} = \mathbf{B}\boldsymbol{\Sigma}_2, \quad (85)$$

$$\mathbf{S}\mathbf{S}^\top\mathbf{W}\mathbf{W}^\top\mathbf{S}\mathbf{B} = \mathbf{S}\mathbf{B}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1, \quad (86)$$

$$\mathbf{P}_2\mathbf{P}_1(\mathbf{S}\mathbf{B}) = (\mathbf{S}\mathbf{B})\widehat{\boldsymbol{\Sigma}}. \quad (87)$$

$\mathbf{P}_2\mathbf{P}_1$  can be transformed by using  $\mathbf{W}^\top\mathbf{S} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$  as follows:

$$\mathbf{P}_2\mathbf{P}_1 = \mathbf{S}\mathbf{S}^\top\mathbf{W}\mathbf{W}^\top, \quad (88)$$

$$= \mathbf{S}\mathbf{V}\boldsymbol{\Sigma}\mathbf{U}^\top\mathbf{W}^\top, \quad (89)$$

$$\mathbf{P}_2\mathbf{P}_1\mathbf{S} = \mathbf{S}\mathbf{V}\boldsymbol{\Sigma}\mathbf{U}^\top\mathbf{W}^\top\mathbf{S}, \quad (90)$$

$$= \mathbf{S}\mathbf{V}\boldsymbol{\Sigma}\mathbf{U}^\top\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top, \quad (91)$$

$$= \mathbf{S}\mathbf{V}\boldsymbol{\Sigma}^2\mathbf{V}^\top, \quad (92)$$

$$\mathbf{P}_2\mathbf{P}_1(\mathbf{S}\mathbf{V}) = (\mathbf{S}\mathbf{V})\boldsymbol{\Sigma}^2. \quad (93)$$

The last equation implies that  $\mathbf{S}\mathbf{V}$  represents a set of eigenvectors of  $\mathbf{P}_2\mathbf{P}_1$ .  $\square$