

# [Supplementary] Membership Representation for Detecting Block-diagonal Structure in Low-rank or Sparse Subspace Clustering

Minsik Lee<sup>†</sup>    Jieun Lee<sup>‡</sup>    Hyeogjin Lee<sup>§</sup>    Nojun Kwak<sup>§</sup>

Division of EE, Hanyang University, Korea<sup>†</sup>

Department of ECE, Ajou University, Korea<sup>‡</sup>

Graduate School of CST, Seoul National University, Korea<sup>§</sup>

mleepaper@hanyang.ac.kr

mokona85@ajou.ac.kr

{hjinlee, nojunk}@snu.ac.kr

## I. PROOF OF EQUATION (11)

First, we give the following lemma:

**Lemma 1.** *Let  $\mathbf{P}$  be an orthogonal projection, i.e.,  $\mathbf{P}^2 = \mathbf{P} = \mathbf{P}^T$ , and all the elements are nonnegative. Then,  $\mathbf{P}$  is similar to a block-diagonal matrix by permuting the rows and columns for the same indices.*

Essentially, the proof of this lemma is based on the fact that all the off-diagonal elements of  $\mathbf{I} - \mathbf{P}$  are non-positive and it can be decomposed as  $\mathbf{I} - \mathbf{P} = \mathbf{V}^\perp \mathbf{V}^{\perp T}$  for a thin orthogonal matrix  $\mathbf{V}^\perp$ . The only case that satisfies both conditions is when  $\mathbf{P}$  is similar to a block-diagonal matrix:

*Proof.* Since  $\mathbf{P}$  is an orthogonal projection, it can be factorized as  $\mathbf{P} = \mathbf{V}\mathbf{V}^T$  where  $\mathbf{V} \in \mathbb{R}^{m_r \times m_c}$  is a thin orthogonal matrix. Since  $\mathbf{P}$  is nonnegative, the off-diagonal elements of  $\mathbf{P}' = \mathbf{I} - \mathbf{P} = \mathbf{V}^\perp \mathbf{V}^{\perp T}$  will be non-positive, where  $\mathbf{V}^\perp \in \mathbb{R}^{m_r \times (m_r - m_c)}$  is an orthogonal matrix satisfying  $\mathbf{V}^T \mathbf{V}^\perp = \mathbf{0}$ . By applying  $QR$ -decomposition to  $\mathbf{V}^{\perp T} = \mathbf{Q}\mathbf{V}'^T$ , it becomes

$$\mathbf{P}' = \mathbf{V}^\perp \mathbf{V}^{\perp T} = \mathbf{V}' \mathbf{Q}^T \mathbf{Q} \mathbf{V}'^T = \mathbf{V}' \mathbf{V}'^T, \quad (\text{S1})$$

where  $\mathbf{V}'$  is a lower-triangular matrix. Without loss of generality, we assume that  $\mathbf{V}'$  has strictly positive diagonal elements. Along with this proof, we will show a conceptual example of  $\mathbf{V}'$  for  $m_r = 7$  and  $m_c = 4$  to help the readers' understanding, as follows:

$$\mathbf{V}' = \begin{bmatrix} + & 0 & 0 \\ ? & + & 0 \\ ? & ? & + \\ ? & ? & ? \\ ? & ? & ? \\ ? & ? & ? \\ ? & ? & ? \end{bmatrix}. \quad (\text{S2})$$

Let us denote  $\mathbf{V}'_{:i}$  and  $\mathbf{V}'_{j\cdot}$  as the  $i$ th column and  $j$ th row vectors of  $\mathbf{V}'$ , respectively. Then, all the elements in  $\mathbf{V}'_{:1}$  except  $\mathbf{V}'_{11} > 0$  are zeros. Since  $\mathbf{V}'_{:1} \mathbf{V}'_{j\cdot}^T = \mathbf{P}'_{1j} \leq 0$  for all  $j \neq 1$ , all the elements in  $\mathbf{V}'_{:1}$  except  $\mathbf{V}'_{11}$  must be non-positive. We can similarly show that all the elements of  $\mathbf{V}'$  except the diagonal elements are non-positive, by observing

that  $\mathbf{V}'_{:i}$  must have non-positive elements except for  $\mathbf{V}'_{ii}$ . This can be shown as in the following example:

$$\mathbf{V}' \mathbf{V}'^T = \begin{bmatrix} + & 0 & 0 \\ \ominus & + & 0 \\ \ominus & \ominus & + \\ \ominus & \ominus & \ominus \end{bmatrix} \begin{bmatrix} + & \ominus & \ominus & \ominus & \ominus & \ominus & \ominus \\ 0 & + & \ominus & \ominus & \ominus & \ominus & \ominus \\ 0 & 0 & + & \ominus & \ominus & \ominus & \ominus \end{bmatrix}, \quad (\text{S3})$$

where  $\ominus$  represents non-positive elements.

Thus,  $\mathbf{V}'_{j\cdot}$  only contains non-positive elements for  $j > m'_c$ , where  $m'_c = m_r - m_c$ . However, the inner product of nonnegative vectors is positive, which violates the assumption, unless they are orthogonal in which case the inner product will be zero. Hence, there can be no more than  $m'_c$  row vectors for  $j > m'_c$  that have at least one strictly negative element, i.e., there will be at least  $m_r - 2m'_c (= 2m_c - m_r)$  zero rows in  $\mathbf{V}'$ . Hence, we assume that  $\mathbf{V}'_{j\cdot} = \mathbf{0}$  for  $j > m'_c + m''_c$  without loss of generality. Here,  $m''_c (\leq m'_c)$  is the number of nonzero rows  $\mathbf{V}'_{j\cdot}$  for  $j > m'_c$ . Now, (S3) becomes

$$\mathbf{V}' \mathbf{V}'^T = \begin{bmatrix} + & 0 & 0 \\ \ominus & + & 0 \\ \ominus & \ominus & + \\ \ominus & \ominus & \ominus \\ \ominus & \ominus & \ominus \\ \ominus & \ominus & \ominus \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} + & \ominus & \ominus & \ominus & \ominus & \ominus & 0 \\ 0 & + & \ominus & \ominus & \ominus & \ominus & 0 \\ 0 & 0 & + & \ominus & \ominus & \ominus & 0 \end{bmatrix}. \quad (\text{S4})$$

Moreover, if  $\mathbf{V}'_{ji} < 0$  for some  $i$  and  $m'_c < j \leq m'_c + m''_c$ ,

$$\mathbf{V}'_{j'i} = 0, \quad \text{for } j' \neq j, m'_c < j' \leq m'_c + m''_c, \quad (\text{S5})$$

i.e., for each  $i$ , there can be only one element  $(j, i)$  for  $j \in (m'_c, m'_c + m''_c]$  which is strictly negative. Thus, the columns  $\{i\}$  can be partitioned into  $m''_c$  disjoint sets, i.e.,

$$\mathbf{i}_j = \{i | \mathbf{V}'_{ji} < 0\} \quad \text{for } m'_c < j \leq m'_c + m''_c, \quad (\text{S6})$$

and  $\mathbf{i}_j \cap \mathbf{i}_{j'} = \emptyset$  for  $j \neq j'$ . The following example shows

this property:

$$\mathbf{V}'\mathbf{V}'^T = \begin{bmatrix} + & 0 & 0 \\ \ominus & + & 0 \\ \ominus & \ominus & + \\ 0 & - & - \\ - & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} + & \ominus & \ominus & 0 & - & 0 & 0 \\ 0 & + & \ominus & - & 0 & 0 & 0 \\ 0 & 0 & + & - & 0 & 0 & 0 \end{bmatrix}. \quad (\text{S7})$$

Now, we will show that the indices of rows can also be partitioned based on  $\mathbf{i}_j$ . To do this, we have to show that  $\{\mathbf{V}'_{ki}\}$  is partitioned based on  $i$  for  $1 \leq k \leq m'_c$ , because it is already partitioned for  $k > m'_c$ . Let  $j$  be the row index satisfying  $m'_c \in \mathbf{i}_j$ , i.e.,  $\mathbf{V}'_{jm'_c} < 0$ , and let  $i' \in \mathbf{i}_{j'}$  ( $\mathbf{V}'_{j'i'} < 0$ ) for some  $j' \neq j$ . Then,

$$0 = \mathbf{V}'_{:m'_c}^T \mathbf{V}'_{:i'} = \sum_{k=1}^{m_r} \mathbf{V}'_{km'_c} \mathbf{V}'_{ki'} = \mathbf{V}'_{m'_cm'_c} \mathbf{V}'_{m'_ci'}, \quad (\text{S8})$$

because  $\mathbf{V}'_{km'_c} = 0$  for  $k < m'_c$  and  $\mathbf{V}'_{km'_c} \mathbf{V}'_{ki'} = 0$  for  $k > m'_c$ . This means that  $\mathbf{V}'_{m'_ci'} = 0$  for any  $i' \in \mathbf{i}_{j'}, j' \neq j$ . Likewise, we can easily show that  $\mathbf{V}'_{m'_ci''} < 0$  for any  $i'' \in \mathbf{i}_j$ . This can be shown as in the following:

$$\mathbf{V}'\mathbf{V}'^T = \begin{bmatrix} + & 0 & 0 \\ \ominus & + & 0 \\ 0 & - & + \\ 0 & - & - \\ - & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} + & \ominus & 0 & 0 & - & 0 & 0 \\ 0 & + & - & - & 0 & 0 & 0 \\ 0 & 0 & + & - & 0 & 0 & 0 \end{bmatrix}. \quad (\text{S9})$$

Similarly, we can show that  $\mathbf{V}'_{ii'} = 0$  if

$$\begin{aligned} i \in \mathbf{i}_j, i' \in \mathbf{i}_{j'}, \quad \text{for } j' \neq j, \\ \mathbf{V}'_{ii'} = 0, \quad \text{for } i < i'' \leq m'_c, \end{aligned} \quad (\text{S10})$$

and otherwise  $\mathbf{V}'_{ii'} < 0$ . This recursively proves that  $\mathbf{V}'_{ii'} = 0$  for  $i \in \mathbf{i}_j, i' \in \mathbf{i}_{j'}, j' \neq j$ , which also proves that the indices of rows can also be partitioned based on  $\mathbf{i}_j$ . Therefore, we can define another  $m'_c$  disjoint sets, i.e.,

$$\mathbf{k}_j = \mathbf{i}_j \cup \{j\} \quad \text{for } m'_c < j \leq m'_c + m'_c, \quad (\text{S11})$$

and  $\mathbf{k}_j \cap \mathbf{k}_{j'} = \emptyset$  for  $j \neq j'$ . Then,  $\mathbf{V}'_{ki}$  can be nonzero only if  $k \in \mathbf{k}_j, i \in \mathbf{i}_j$ . Finally, our example looks like

$$\mathbf{V}'\mathbf{V}'^T = \begin{bmatrix} + & 0 & 0 \\ 0 & + & 0 \\ 0 & - & + \\ 0 & - & - \\ - & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} + & 0 & 0 & 0 & - & 0 & 0 \\ 0 & + & - & - & 0 & 0 & 0 \\ 0 & 0 & + & - & 0 & 0 & 0 \end{bmatrix}. \quad (\text{S12})$$

Since we have shown that the nonzero elements of  $\mathbf{V}'$  can be partitioned into disjoint sets of rows and columns,  $\mathbf{V}'$  can be permuted<sup>1</sup> into a block-diagonal matrix  $\mathbf{V}''$ , as in the

following example:

$$\mathbf{V}'' = \begin{bmatrix} + & 0 & 0 \\ - & 0 & 0 \\ 0 & + & 0 \\ 0 & - & + \\ 0 & - & - \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \quad (\text{S13})$$

Note that the size of each block matrix is  $N(\mathbf{k}_j) \times N(\mathbf{i}_j) = (N(\mathbf{i}_j) + 1) \times N(\mathbf{i}_j)$ . Hence, there must be a unit vector  $\mathbf{v}_j$ , which is orthogonal to  $\mathbf{V}''$  and whose elements are nonzero only for the row indices of the  $j$ th block of  $\mathbf{V}''$ , for each  $j$ . Note that  $\mathbf{v}_j$ s are mutually orthogonal. For the above example, there are four  $\mathbf{v}_j$ s which look like

$$\mathbf{v}_1 = \begin{bmatrix} + \\ + \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 0 \\ + \\ + \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}. \quad (\text{S14})$$

Therefore, if we let  $\bar{\mathbf{V}}$  be a matrix formed with  $\mathbf{v}_j$ , then  $\bar{\mathbf{V}}\bar{\mathbf{V}}^T$  is block-diagonal, where the blocks are rank one with eigenvalue one. Moreover, if we form  $\tilde{\mathbf{V}}$  by permuting the rows of  $\bar{\mathbf{V}}$  in the inverse order of the permutation from  $\mathbf{V}'$  to  $\mathbf{V}''$ , then  $\tilde{\mathbf{V}}$  and  $\mathbf{V}$  must have the same span because of the assumption. This concludes the proof.  $\square$

This lemma leads to the following theorem:

**Theorem 1.**  $\mathbf{F}'$  is a normalized membership matrix iff it is doubly stochastic and is an orthogonal projection.

*Proof.*  $\mathbf{F}'$  is an orthogonal projection because it is symmetric and its eigenvalues are either zero or one. This proves the only-if part.

Based on Lemma 1, if  $\mathbf{F}'$  is an orthogonal projection,  $\mathbf{F}'$  is similar to a block-diagonal matrix  $\mathbf{F}^*$  by permutation<sup>2</sup>. Since  $\mathbf{F}^*$  is also doubly stochastic, the row sums of  $\mathbf{F}^*$  are 1, i.e.,  $\mathbf{F}^* \mathbf{1} = \mathbf{1}$ . In order for this to be valid, every block  $\mathbf{F}_k^*$  must satisfy  $\mathbf{F}_k^* \mathbf{1} = \mathbf{1}$ . Because  $\mathbf{F}_k^*$  is rank one with eigenvalue one, all of the elements of  $\mathbf{F}_k^*$  must be equal to  $\frac{1}{n_k}$  where  $n_k$  is the dimension of  $\mathbf{F}_k^*$ . This proves the if part.  $\square$

## II. MOTION CLUSTERING FOR CMU MOTION CAPTURE SEQUENCES

In the field of non-rigid structure from motion (NRSfM) [S1], the non-rigid shape changes are often assumed to have a low-rank basis, under an assumption that the changes in the data sequence is simple enough. If this assumption does not hold, then the data sequence can be divided into several scenes, where each of them can be expressed by a low-rank basis, as in [S1]. Hence, a complex non-rigid motion sequence can also be a good example to show the effectiveness of a

<sup>1</sup>Care should be taken that, in this case, the permutation *can* be applied to different indices for rows and columns.

<sup>2</sup>The elements of a doubly stochastic matrix are nonnegative.

subspace clustering technique. Accordingly, we have applied the proposed algorithm to some of the CMU motion capture sequences<sup>3</sup>. Since these sequences have large numbers of frames with redundant shapes, we reduced them by sampling every  $m$ th frame so that the number of frames in each sequence becomes less than 1500. Since only the non-rigid changes are assumed to be low-rank in NRSfM, we excluded any rigid motions in the data sets by using the generalized Procrustes analysis (GPA) [S2], and normalized each data set so that the mean-squared value of the elements becomes one. The parameters of LRR and SSC have been set to  $\lambda = 0.05$  and  $\alpha = 2000$ , respectively, and those of MR were set as  $\lambda_M = 0.02$  and  $\beta = 0.5$  for LRR, and  $\lambda_M = 10^{-8}$  and  $\beta = 0.6$  for SSC, respectively. The video results of the motion clustering are provided as another supplementary material. An objective evaluation is not possible because there is no ground truth for these data sets, but we can confirm subjectively that the results are good, based on the videos.

#### REFERENCES

- [S1] Y. Zhu, D. Huang, F. De la Torre, and S. Lucey, "Complex non-rigid motion 3d reconstruction by union of subspaces," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, June 2014.
- [S2] J. C. Gower, "Generalized procrustes analysis," *Psychometrika*, vol. 40, no. 1, pp. 33–51, March 1975.

<sup>3</sup><http://mocap.cs.cmu.edu/subjects.php>