Membership Representation for Detecting Block-diagonal Structure in Low-rank or Sparse Subspace Clustering

Minsik Lee[1], Jieun Lee[2], Hyeogjin Lee[3], Nojun Kwak[3]
[1]Division of EE, Hanyang University, Korea. [2]Department of ECE, Ajou University, Korea. [3]Graduate School of CST, Seoul National University, Korea.

Recently, there have been many proposals [2, 5] with state-of-the-art results in subspace clustering [3] that take advantages of the low-rank or sparse optimization techniques [4]. These methods are based on self-expressive models, which have well-defined theoretical aspects. They produce matrices with (approximately) block-diagonal structure, which is then applied to spectral clustering. However, there is no definitive way to construct affinity matrices from these block-diagonal matrices and it is ambiguous how the performance will be affected by the construction method.

In this paper, we propose an alternative approach to detect block-diagonal structures from these matrices. The proposed method shares the philosophy of the above subspace clustering methods, in that it is a self-expressive system based on a Hadamard product of a membership matrix, which frequently appears in correlation clustering [1]. A membership matrix $\mathbf{M}$ is a symmetric matrix whose elements are either one or zero, which can be transformed into a block-diagonal matrix by permuting the same indices of rows and columns. It can be shown that a matrix is a membership matrix if and only if it is a matrix of ones and zeros with diagonal elements being ones and is PSD.

If $\mathbf{W}$ is a clean block-diagonal matrix without any error and $\mathbf{M}$ is a membership matrix whose all-one block locations are identical to those of diagonal blocks of $\mathbf{W}$, then obviously

$$\mathbf{W} = \mathbf{W} \odot \mathbf{M}, \tag{1}$$

where $\odot$ is the Hadamard product. Since there is a trivial solution $\mathbf{M} = \mathbf{1}\mathbf{1}^T$, and there can be errors in $\mathbf{W}$, we consider the following alternative problem:

$$\min_{\mathbf{M}} \quad \|\mathbf{W} - \mathbf{W} \odot \mathbf{M}\|_1 + \lambda_M \|\mathbf{M}\|_2, \quad \text{s.t.} \quad \mathbf{M} \in \mathbb{M}, \tag{2}$$

where $\mathbb{M}$ is the set of membership matrices. Here, we use the entry-wise l_1-norm for $\mathbf{W} - \mathbf{W} \odot \mathbf{M}$ to handle outliers. However, this problem is NP-hard because it has both discrete and PSD constraints, so we relax them as

$$\text{diag}(\mathbf{M}) = \mathbf{1}, \quad \mathbf{M} \succeq \mathbf{0}, \quad \mathbf{M} \geq 0, \tag{3}$$

where $\succeq$ and $\geq$ are semidefinite and element-wise inequality, respectively. Based on this relaxation, (2) becomes a convex problem which can be efficiently solved by an augmented Lagrangian method [4]. The solution $\breve{\mathbf{M}}$ is a boosted version of $\mathbf{W}$ that emphasizes the block-diagonal structure.

Even though $\breve{\mathbf{M}}$ might be better than $\mathbf{W}$ to apply spectral clustering, we still do not know the number of clusters. To resolve this issue, we transform the representation to a normalized membership matrix [6], which is a variant of the membership matrix. A normalized membership matrix $\mathbf{F}$ is similar to a membership matrix, except that the cluster blocks are filled with $\frac{1}{n_k}$ instead of ones, where n_k is the number of samples in the kth cluster. It can be shown that a matrix is a normalized membership matrix if and only if it is doubly stochastic, i.e., $\mathbf{F}\mathbf{1} = \mathbf{F}^T\mathbf{1} = \mathbf{1}$ and is an orthogonal projection, i.e., $\mathbf{F}^2 = \mathbf{F} = \mathbf{F}^T$. One of the important properties of a normalized membership matrix is that it has eigenvalues either one or zero, thus the sum of eigenvalues is identical to its rank, i.e., the number of clusters. Therefore, it can be more reliable in estimating the number of clusters.

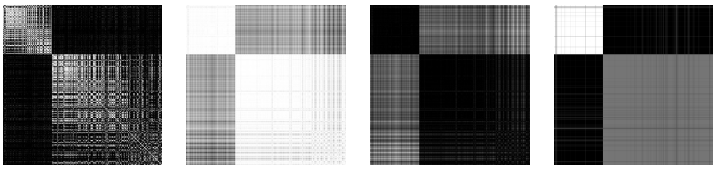
In order to transform $\breve{\mathbf{M}}$ to a normalized membership matrix, we utilize the equation (1) conversely: In an ideal case where $\breve{\mathbf{M}} \in \mathbb{M}$, it is obvious that

$$\mathbf{F} = \mathbf{F} \odot \breve{\mathbf{M}}, \tag{4}$$

where $\mathbf{F}$ is the normalized membership counterpart of $\breve{\mathbf{M}}$. To avoid a trivial solution $\mathbf{F} = \mathbf{I}$, we need a regularization term $\|\mathbf{F}\|_* = \text{tr}(\mathbf{F})$. Since $\|(\mathbf{1}\mathbf{1}^T - \breve{\mathbf{M}}) \odot \mathbf{F}\|_1$ has a different scale with $\text{tr}(\mathbf{F})$, we instead minimize $\text{tr}(\mathbf{F})$ with a constraint $\|(\mathbf{1}\mathbf{1}^T - \breve{\mathbf{M}}) \odot \mathbf{F}\|_1 \leq c$, where the constant c is decided based on $\mathbf{H}$. Since handling a normalized membership matrix is also difficult, we may relax the constraint to make it tractable as

$$\min_{\mathbf{F}} \quad \text{tr}(\mathbf{F}), \quad \text{s.t.} \quad \mathbf{F} \geq 0, \ \mathbf{F}\mathbf{1} = \mathbf{1}, \ \mathbf{F} \succeq \mathbf{0}, \ \|\mathbf{H} \odot \mathbf{F}\|_1 \leq c. \tag{5}$$

The solution $\breve{\mathbf{F}}$ of the above problem can be considered as a doubly stochastic normalization [6] of $\breve{\mathbf{M}}$, which is more appropriate for spectral clustering. In performing spectral clustering on $\breve{\mathbf{F}}$, we do not need the normalization step of the normalized cut, because our affinity matrix is already doubly stochastic. Since all the eigenvalues of $\breve{\mathbf{F}}$ are in between zero and one regardless of the input data, it can be more reliable to infer the number of clusters from $\breve{\mathbf{F}}$ than from $\mathbf{W}'$. In our experiments, we just counted the eigenvalues above $\frac{1}{2}$ to estimate K. We observe that in many cases, $\breve{\mathbf{F}}$ is indeed very close to a normalized membership matrix, as shown in Fig. 1(d).



(a) Input (latent matrix $\mathbf{W}$ = output of SSC/LRR) (b) Output of the 1st stage: Approximation of membership matrix ($\breve{\mathbf{M}}$) (c) Dissimilarity matrix ($\mathbf{H} = \mathbf{1}\mathbf{1}^T - \breve{\mathbf{M}}$) (d) Output of the 2nd stage: Approximation of normalized membership matrix ($\breve{\mathbf{F}}$)

Figure 1: A typical example of the intermediate results of the proposed algorithm (Hopkins155, $K = 2$, $n = 276$).

The details of the optimization procedure is described in the paper. Figure 1 shows the intermediate results of the proposed algorithm, which can give a general sense of the method. MR showed better or at least similar results than the plain spectral clustering approach in our experiments. In summary, MR can act as a robust post-processing step for subspace clustering. It depends much less on the characteristics of the data than the combination of a heuristic estimator and the normalized cut, and gives a more reliable form to estimate the number of clusters.

[1] Yudong Chen, Ali Jalali, Sujay Sanghavi, and Huan Xu. Clustering partially observed graphs via convex optimization. *J. Machine Learning Research*, 15:2213–2238, June 2014. URL http://jmlr.org/papers/v15/chen14a.html.

[2] Ehsan Elhamifar and René Vidal. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 35(11):2765–2781, January 2013.

[3] Fabien Lauer and Christoph Schnorr. Spectral clustering of linear subspaces for motion segmentation. In *Proc. IEEE Int'l Conf. Computer Vision*, September 2009.

[4] Z. Lin, M. Chen, L. Wu, and Y. Ma. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. Technical Report UILU-ENG-09-2215, University of Illinois at Urbana-Champaign, November 2009.

[5] Guangcan Liu, Zhouchen Lin, Shuicheng Yan, Ju Sun, Yong Yu, and Yi Ma. Robust recovery of subspace structures by low-rank representation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 35(1): 171–184, January 2013.

[6] Ron Zass and Amnon Shashua. A unifying approach to hard and probabilistic clustering. In *Proc. IEEE Int'l Conf. Computer Vision*, October 2005.