

# Regularized $\ell^1$ -Graph for Data Clustering

Yingzhen Yang<sup>1</sup>  
yyang58@ifp.uiuc.edu

Zhangyang Wang<sup>1</sup>  
zwang119@ifp.uiuc.edu

Jianchao Yang<sup>2</sup>  
jjayang@adobe.com

Jiawei Han<sup>1</sup>  
hanj@cs.uiuc.edu

Thomas S. Huang<sup>1</sup>  
huang@ifp.uiuc.edu

<sup>1</sup> University of Illinois at Urbana-Champaign  
Urbana, IL 61801, USA

<sup>2</sup> Adobe Research  
San Jose, CA 95110, USA

$\ell^1$ -Graph has been proven to be effective in data clustering, which partitions the data space by using the sparse representation of the data as the similarity measure. However, the sparse representation is performed for each datum independently without taking into account the geometric structure of the data. Motivated by  $\ell^1$ -Graph and manifold learning, we propose Regularized  $\ell^1$ -Graph ( $R\ell^1$ -Graph) for data clustering. Compared to  $\ell^1$ -Graph, the sparse representations of  $R\ell^1$ -Graph are regularized by the geometric information of the data. In accordance with the manifold assumption, the sparse representations vary smoothly along the geodesics of the data manifold through the graph Laplacian constructed by the sparse codes. Experimental results on various data sets demonstrate the superiority of our algorithm compared to  $\ell^1$ -Graph and other competing clustering methods.

$\ell^1$ -graph [2, 3], which builds the graph by reconstructing each datum with all the other data, has been shown to be robust to noise and capable of producing superior results for high dimensional data, compared to K-means and spectral clustering. Compared to  $k$ -nearest-neighbor graph and  $\mathcal{E}$ -ball graph,  $\ell^1$ -graph adaptively determines the neighborhood of each datum by solving sparse representation problem locally. Given the data  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$ ,  $\ell^1$ -graph seeks for the robust sparse representation for the entire data by solving the  $\ell_1$ -norm optimization problem for each data point:

$$\min_{\alpha^i} \|\alpha^i\|_1 \quad s.t. \quad \mathbf{x}_i = \mathbf{X}\alpha^i \quad i = 1, \dots, n \quad (1)$$

where  $\alpha^i \in \mathbb{R}^{n \times 1}$ , and we denote by  $\alpha$  the coefficient matrix  $\alpha = [\alpha^1, \dots, \alpha^n] \in \mathbb{R}^{n \times n}$  with the element  $\alpha_{ij} = \alpha_{ij}^j$ . Let  $G = (\mathbf{X}, \mathbf{W})$  be the  $\ell^1$ -graph where  $\mathbf{X}$  is the set of vertices,  $\mathbf{W}$  is the graph weight matrix and  $\mathbf{W}_{ij}$  indicates the similarity between  $\mathbf{x}_i$  and  $\mathbf{x}_j$ .  $\ell^1$ -graph sets the  $n \times n$  matrix  $\mathbf{W}$  as

$$\mathbf{W} = (|\alpha| + |\alpha^T|)/2 \quad (2)$$

where  $|\alpha|$  is the matrix whose elements are the absolute values of  $\alpha$ , and then feed  $\mathbf{W}$  as the pairwise similarity matrix into the spectral clustering algorithm to get the clustering result.

While  $\ell^1$ -graph demonstrates better performance than many traditional similarity-based clustering methods, it performs sparse representation for each datum independently without considering the geometric information and manifold structure of the entire data. In order to obtain the sparse representations that account for the geometric information and manifold structure of the data, we employ the manifold assumption [1] and propose a novel Regularized  $\ell^1$ -Graph ( $R\ell^1$ -Graph). The manifold assumption in this case requires that if two points  $\mathbf{x}_i$  and  $\mathbf{x}_j$  are close in the intrinsic geometry of the submanifold, their corresponding sparse codes  $\alpha^i$  and  $\alpha^j$  are also expected to be similar to each other. The following objective function for  $R\ell^1$ -Graph is given below:

$$\min_{\alpha, \mathbf{W}} \sum_{i=1}^n \|\mathbf{x}_i - \mathbf{X}\alpha^i\|_2^2 + \lambda \|\alpha\|_1 + \gamma \text{Tr}(\alpha \mathbf{L}_\mathbf{W} \alpha^T) \quad (3)$$

$$s.t. \quad \mathbf{W} = (\mathbf{A} \circ |\alpha| + \mathbf{A}^T \circ |\alpha^T|)/2 \quad \alpha \in S$$

where  $S = \{\alpha \in \mathbb{R}^{n \times n} | \alpha_{ii} = 0, 1 \leq i \leq n\}$ ,  $\lambda > 0$  is the weight controlling the sparsity of the coefficients, and  $\gamma > 0$  is the weight of the regularization term,  $\mathbf{L}_\mathbf{W}$  is the graph Laplacian matrix constructed by the pairwise similarity matrix  $\mathbf{W}$ ,  $\mathbf{A}$  is a KNN adjacency matrix.

We simplified the optimization problem (3), and employ Alternating Direction Method of Multipliers (ADMM) to solve the nonconvex optimization problem. ADMM decomposes the original problem into a sequence of tractable subproblems which can be solved efficiently.

We demonstrate the performance of  $R\ell^1$ -Graph with comparison to other competing methods, i.e. K-means (KM), Spectral Clustering (SC),  $\ell^1$ -Graph and Laplacian regularized  $\ell^1$ -Graph. There are two parameters that influence the regularization term in  $R\ell^1$ -Graph, namely the weight of the regularization  $\gamma$  and the number of nearest neighbors  $K$  of the KNN adjacency matrix. The regularization term imposes stronger smoothness constraint on the sparse codes with larger  $\gamma$  and  $K$ , and vice versa. We investigate how the clustering accuracy on ORL face database changes when varying these two parameters, and illustrate the result in Figure 1. We observe that the performance of  $R\ell^1$ -Graph is much better than other algorithms over a large range of both  $\gamma$  and  $K$ , revealing the robustness of our algorithm. Please refer to the paper for detailed description of our algorithm and more experimental results.

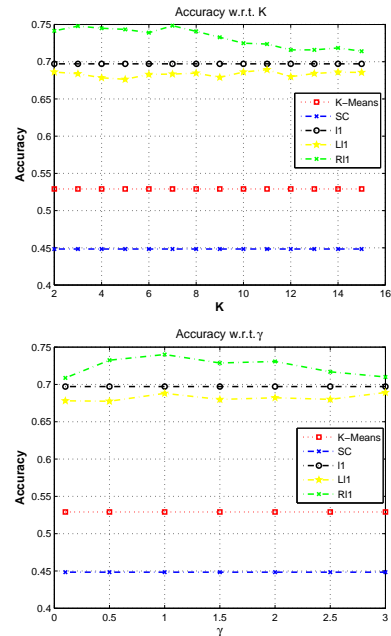


Figure 1: Clustering accuracy with different values of  $K$  and  $\gamma$  on ORL face database. Upper:  $K$ ; Down:  $\gamma$

- [1] Mikhail Belkin, Partha Niyogi, and Vikas Sindhwani. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *Journal of Machine Learning Research*, 7:2399–2434, 2006.
- [2] Bin Cheng, Jianchao Yang, Shuicheng Yan, Yun Fu, and Thomas S. Huang. Learning with  $\ell^1$ -graph for image analysis. *IEEE Transactions on Image Processing*, 19(4):858–866, 2010.
- [3] Shuicheng Yan and Huan Wang. Semi-supervised learning by sparse representation. In *SDM*, pages 792–801, 2009.