Grassmannian Manifold Optimization Assisted Sparse Spectral Clustering

Qiong Wang[a] Junbin Gao[b] Hong Li[a]

[a]School of Mathematics and Statistics Huazhong University of Science and Technology,Wuhan 430074, China

[b] Discipline of Business Analytics, The University of Sydney Business School The University of Sydney, NSW 2006, Australia

{wangqiong701,hongli}@hust.edu.cn ; junbin.gao@sydney.edu.au

IEEE 2017 Conference on Computer Vision and Pattern Recognition

CVPR July 21-26 HONOLULU 2017

Abstract

This paper provides a direct solution as solving a new Grassmann optimization problem. By this way calculating latent embedding becomes part of optimization on manifolds and the recently developed manifold optimization methods can be applied. The primary contributions of this paper are:

- We take a straightforward way to optimize the Sparse Spectral Clustering（SSC）objective introduced in[1]{Lu,Yan,Lin2016} by adopting Grassmannian manifold optimization strategy.
- We explore the application of the new algorithm for dimensionality reduction .

Summarization of SSC

Let $X = [x_1, x_2, \cdots x_N] \in R^{D\times N}$ be a set of N data points to be clustered, where D is the dimension of data. The Spectral Clustering (*SC*) method solve the following constrained optimization for computing $U \in \boldsymbol{R}^{N\times d}$

$$\min_{U\in R^{N\times d}} \langle UU^T, L\rangle \ , s.t.\ \ U^TU = I$$

where $d<N$, and L is the normalized graph Laplacian matrix, and then conduct the *k*-means on the normarized rows of U to cluster them into K groups.

In the ideal scenarios, UU^T can be permuted to block diagonal structure, The Sparse Spectral Clustering (SSC) in recent work [1] exploited the idea of inducing or enforcing sparsity in the Spectral Clustering,

$$\min_{U\in R^{N\times d}} \langle UU^T, L\rangle + \beta\|UU^T\|_1 \ , s.t.\ \ U^TU = I$$

The SSC aims at solving the following relaxed convex problem which optimize the new variable $P = UU^T$ instead.

$$\min_{P\in S^{N\times N}} \langle P, L\rangle + \beta\|P\|_1 \ , s.t.\ \ 0 \le P \le I,\ tr(P) = d$$

Introduction to GSC

For problem $\displaystyle\min_{U\in R^{N\times d}} \langle UU^T, L\rangle + \beta\|UU^T\|_1 \ , s.t.\ \ U^TU = I$

The orthogonal constraint $U^TU = I$ defines the Stiefel manifold $ST(d,N)$;

Let $O(d)=\{Q\in R^{d\times d} \mid Q^TQ = I\}$,the quotient space of Stiefel manifold under this equivalent relation: $ST(d,N)/O(d) := \{YQ: Y\in ST(d,N),\ Q\in O(d)\}$ is the representation of Grassmann manifold $G(d,N)$;

Let $f= \langle UU^T, L\rangle + \beta\|UU^T\|_1$, for any $Q\in O(d)$, we have $f(UQ) = f(U)$;

A better strategy is to re-form the problem on the Grassmann manifold as follows:

$$\min_{U\in G(d,N)} \langle UU^T, L\rangle + \beta\|UU^T\|_1$$

Computing the Gradient

For the first term in the objective function $f = \langle UU^T, L\rangle + \beta\|UU^T\|_1$

we note that: $\langle UU^T, L\rangle = tr(UU^TL) = tr(U^TLU)$.

Hence $\nabla\langle UU^T, L\rangle = LU + L^TU = 2LU$

Consider the second term of the objective function. First, according to the chain rule, we have:

$$vec\left(\frac{\partial\|UU^T\|_1}{\partial U}\right)^T = vec\left(sgn(UU^T)\right)^T \frac{\partial UU^T}{\partial U}$$

where $$\frac{\partial UU^T}{\partial U} = \left(I_{N^2} + T_{N\times N, N\times N}\right)\left(U\otimes I_N\right)$$

Define the column vector D as $$D = \frac{\partial UU^T}{\partial U} vec\left(sgn(UU^T)\right)^T$$

Thus the Euclidean derivative of the objective function $f(U)$ is:

$$\nabla f(U) = 2LU + \beta \mathrm{ivec}(D).$$

The Sparse Spectral Clustering Algorithm

At the representative U of a Grassmann point [U], the Riemann gradient can be simply calculated as: $\mathrm{grad}_{[U]} f = (I - UU^T)\nabla f(U)$.

Algorithm 1 Grassmann Manifold Optimization Assisted Spectral Clustering (GSC) Algorithm

Input: The data matrix X = [x_1,x_2,..., x_N], the number of latent dimension d and the trade-off parameter β.

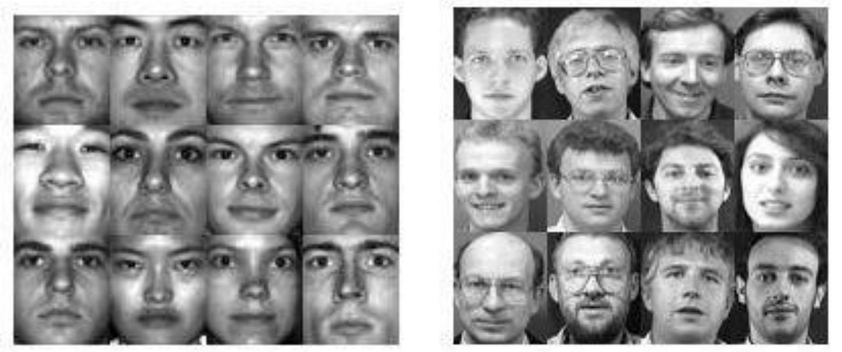
Output: The sparse latent representation U.

1: Form the affinity matrix W, and compute the initial latent representation $U^{(0)}$;

2: Compute the normalized Laplacian matrix L;

3: With the initial $U^{(0)}$, call the Riemannian trust-region (RTR)algorithm in ManOpt toolbox to optimize the objective, until a pre-defined termination criterion is satisfied.

Experiments on Clustering



(a) (b)

Figure 1. Examples of the face datasets: (a) Extended Yale B and (b) ORL faces.

Method	Ncut	SSC	GSC
$K = 5$	61.56(9.34)	95.64(5.90)	**96.58(3.94)**
$K = 8$	56.77(8.84)	88.95(4.76)	**91.65(4.64)**
$K = 10$	48.39(7.61)	82.86(4.86)	**85.24(4.34)**
$K = 12$	46.94(4.82)	80.17(4.40)	**82.64(4.20)**
$K = 15$	45.51(4.06)	76.91(1.57)	**77.82(1.63)**
$K = 18$	45.33(3.80)	75.06(1.59)	**76.84(1.49)**

Table 1. Clustering results in terms of accuracy (%) and standard deviation on Yale B dataset.

Method	Ncut	SSC	GSC
$K = 5$	59.85(6.98)	97.25(6.62)	**97.80(5.41)**
$K = 8$	55.75(6.31)	91.25(5.89)	**93.50(5.41)**
$K = 10$	55.25(5.91)	80.95(10.67)	**82.77(6.32)**
$K = 12$	51.35(6.98)	79.55(11.32)	**82.50(10.82)**
$K = 15$	50.47(5.07)	78.85(7.06)	**79.67(4.79)**
$K = 18$	50.10(4.76)	77.95(7.41)	**78.96(5.21)**

Table 2. Clustering results in terms of accuracy (%) and standard deviation on ORL dataset.

Experiments on Dimensionality Reduction

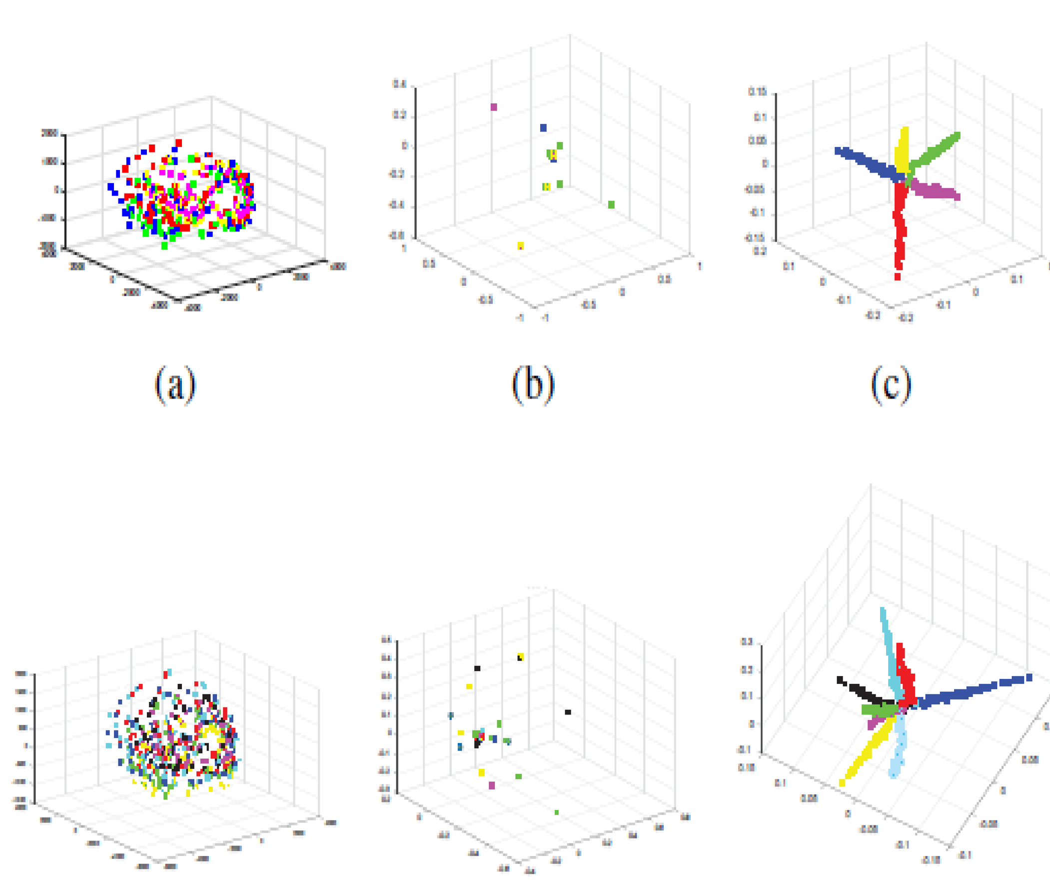


Figure 2. Visualization of the data dimension reduction of PCA of original data set (a) and (d); matrix U of SSC (b) and (e), and U of GDR (c) and (f): 5 classes case on the first row and 8 classes case on the second row for the YaleB faces data set.

Conclusion

1. This paper proposes the GSC model which adopts Grassmann manifold optimization strategy to optimize the sparse spectral clustering objective introduced in [1] in a straight forward way.
2. The major difference between our method and [1] is that ours guarantees $UU^T = I$ (bez on Grassmann) while a big relaxation to this constraint in [1].
3. We also propose the GDR model which visualizes the latent representation of original data as the results from dimensionality reduction.

Reference

[1] C. Lu, S. Yan, and Z. Lin. Convex sparse spectral clustering: Single-view to multi-view. IEEE Transactions on Image Processing, 25(6):2833–2843, 2016