Compressive Spectral Clustering - Error Analysis

Blake Hunter and **Thomas Strohmer**

Department of Mathematics
University of California, Davis
One Shields Ave.
Davis, California 95616

Abstract

Compressive spectral clustering combines the distance preserving measurements of compressed sensing with the power of spectral clustering. Our analysis provides rigorous bounds on how small errors in the affinity matrix can affect the spectral coordinates and clusterability. This work generalizes the current perturbation results of two-class spectral clustering to incorporate multiclass clustering using k eigenvectors.

One of the most common and powerful techniques for extracting meaningful information from a data set is spectral clustering. Spectral clustering uses local information to embed the data into a space which captures the global group structure. Standard learning techniques require an appropriate transformation to higher dimension where dimensionality reduction is done before clustering. Compressed sensing provides a mathematically rigorous way to obtain optimal dimensionality reduction for exact reconstruction. Hyperspectral images and MRIs are examples of high dimensional signals where the true underlying data may only have a few degrees of freedom or be sparse in some unknown basis. We show that the meaningful organization extracted from spectral clustering is preserved under the perturbation from making compressed sensing measurements.

Background

Traditional spectral clustering uses local Euclidean distances between data points to construct a graph with edge weights,

$$W(x_i, x_j) = \exp\left(-\frac{\|x_i, x_j\|_2^2}{2\sigma}\right),$$

that define the symmetric affinity matrix

$$A = D^{-\frac{1}{2}} W D^{-\frac{1}{2}} \qquad (1)$$

where $D(x_i, x_i) = \sum_k W(x_i, x_k)$ is a diagonal matrix of row sums of W. The data can be bipartitioned by applying k-means to the coordinates of the second eigenvector of the affinity matrix A (Ng, Jordan, and Weiss 2001). Multiclass clustering is achieved by repeatedly bipartitioning using the second eigenvector or by applying k-means to the top k eigenvectors of A (Jianbo, Yu, and Shi 2003).

Compressed sensing is used for exact recovery of sparse signals using fewer measurements than the ambient dimension. Compressed sensing provides a bound on the error derived from making these few measurements of a signal. Our goal is to take advantage of these compressed sensing techniques to perform spectral clustering using much fewer measurements than the ambient dimension (Donoho 2006; Candès and Tao 2005).

Earlier perturbation results have shown that spectral clustering using the second eigenvector is robust to small perturbation of the data (Huang et al. 2008). These results are based on the following perturbation theorem.

Theorem 1 *Given a perturbation of A $\tilde{A} = A + E$, let λ_i and v_i be the ith eigenvalue and eigenvector of A and $\tilde{v}_i$ be the ith eigenvalue of $\tilde{A}$ respectively, then*

$$\|\tilde{v_2} - v_2\| \leq \frac{1}{\lambda_2 - \lambda_3}\|E\| + O\left(\|E\|^2\right). \qquad (2)$$

This applies provided the gap between the second and third eigenvalue is not close to zero which is not the case of data sets with more than two underlying clusters. The number of eigenvalues close to one is equal to the number of separate clusters. Our analysis generalizes the results of spectral clustering on perturbed data to incorporate multi-class clustering using the top k eigenvectors.

Compressive Spectral Clustering

Assume that there is an underlying r-sparse representation y_i of the data x_i, where $y_i = Bx_i$ is a known

or unknown unitary transformation of x_i. Let Φ be a random $K \times N$ matrix, with Gaussian $\mathcal{N}(0,1)$ entries. Define the local distance $\tilde{d}(x_i, x_j) = \|\Phi x_i - \Phi x_j\|_2$ using m compressed sensing measurements. Construct a graph with edge weights,

$$\tilde{W}(x_i, x_j) = \exp\left(-\frac{\|\Phi x_i - \Phi x_j\|_2^2}{2\sigma}\right).$$

Define the symmetric matrix

$$\tilde{A} = \tilde{D}^{-\frac{1}{2}}\tilde{W}\tilde{D}^{-\frac{1}{2}} \qquad (3)$$

where $\tilde{D}_{i,i} = \sum_k \tilde{W}(x_i, x_k)$. Use the first k eigenvectors of $\tilde{A}$ as a k low-dimensional representation of the data and coordinates for clustering and classification.

The perturbation of the local distances calculated in the compressed domain, $\tilde{d}(x_i, x_j)$, can be made arbitrarily close to the true local distance $d(x_i, x_j)$ by taking enough measurements. This is guaranteed by the restricted isometry property of compressed sensing. This small error can cause large error between the true and perturbed eigenvectors, $\|v_i - \tilde{v}_i\|$. With some conditions on the separation of eigenvalues, we show that spectral coordinates of first k eigenvectors of $\tilde{A}$ from taking compressed sensing measurements, can be made arbitrarily close to the traditional spectral coordinate when using the full ambient dimension. Thus the meaningful organization extracted from spectral clustering is preserved under the perturbation from making compressed sensing measurements.

Theorem 2 *Let A be the adjacency matrix from standard spectral clustering defined in* (1). *If $\tilde{A}$ is the adjacency matrix formed by taking compressed sensing measurements defined in* (3), *where $\tilde{W}_{i,j} = e^{-\frac{\|\Phi x_i - \Phi x_j\|_2^2}{2\sigma}}$, the x_is are r-sparse and Φ satisfies the RIP with*

$$\delta = \frac{\epsilon}{4\max_{i,j}\left\{\frac{\|x_i - x_j\|_2^2}{2\sigma}\right\}}.$$

Then for $0 < \epsilon < 1$,

$$|\tilde{A}_{i,j} - A_{i,j}| \leq \epsilon.$$

We show that the first k eigenvectors of $\tilde{A}$ provide the same low dimensional embedding as the k eigenvectors of traditional spectral clustering.

Let

$$V = [v_1 \ v_2 \ \ldots \ v_k],$$

where v_l is the column eigenvector corresponding to the lth largest eigenvalue of A and similarly define

$$\tilde{V} = [\tilde{v_1} \ \tilde{v_2} \ \ldots \ \tilde{v_k}],$$

for first k column eigenvector of $\tilde{A}$. Each data point can be clustered by applying k-means to the rows of V or $\tilde{V}$.

Theorem 3 *Let A and $\tilde{A}$ be the affinity matrices from standard spectral clustering and from compressive spectral clustering as defined above. If there is a $\alpha > 0$ such that $\lambda_k - \lambda_{k+1} \geq \alpha$ and $\lambda_k \geq \alpha$, then the canonical angle Θ between the column space of the first k eigenvectors of $\tilde{A}$ and the column space of the first k eigenvectors of A will satisfy,*

$$\|\sin\Theta\|_F \leq \frac{N\epsilon}{\alpha}.$$

This bounds the low dimensional embedding of projecting onto the first k eigenvectors of $\tilde{A}$,

$$\|P_{\mathcal{V}_k} - P_{\tilde{\mathcal{V}}_k}\|_F = \sqrt{2}\|\sin\Theta\|_F \leq \frac{\sqrt{2}}{\alpha}\|A - \tilde{A}\|_F.$$

Theorem 4 *Given V formed by the top k column eigenvectors of A and $\tilde{V}$, the matrix formed by the top k eigenvectors of $\tilde{A}$. If Q is the orthogonal matrix that minimizes $\|\tilde{V} - VQ\|_F$ defined above then*

$$\|\tilde{V} - VQ\|_2 \leq (1 + \sqrt{2})\frac{N}{\alpha}\epsilon.$$

Thus $\tilde{V}$ can be made arbitrarily close to a rotation of V.

Corollary 5 *If $v(i)$ is the ith row of V formed by the top k eigenvectors of A and $\tilde{v}(i)$ is the ith row of $\tilde{V}$ formed by the top k eigenvectors of $\tilde{A}$. Then*

$$\|\tilde{v}(i) - v(i)Q\|_2 \leq (1 + \sqrt{2})\frac{N}{\alpha}\epsilon$$

where Q is the orthogonal matrix that minimizes $\|\tilde{V} - VQ\|_F$.

The rows of V and $\tilde{V}$ provide interchangeable coordinates for clustering and classification. Thus spectral clustering is achievable in the compressed domain.

References

Candès, E. J., and Tao, T. 2005. Decoding by linear programming. *IEEE Trans. Inform. Theory* 51:4203–4215.

Donoho, D. L. 2006. Compressed sensing. *IEEE Trans. Inform. Theory* 52(4):1289–1306.

Huang, L.; Yan, D.; Jordan, M. I.; and Taft, N. 2008. Spectral clustering with perturbed data. In *Advances in Neural Information Processing Systems (NIPS)*.

Jianbo, S. Y.; Yu, S. X.; and Shi, J. 2003. Multiclass spectral clustering. In *In International Conference on Computer Vision*, 313–319.

Ng, A. Y.; Jordan, M. I.; and Weiss, Y. 2001. On spectral clustering: Analysis and an algorithm. *Advances in Neural Information Processing Systems 14* 849–856.