

Leveraging Union of Subspace Structure to Improve Constrained Clustering

John Lipor and Laura Balzano

University of Michigan Department of Electrical and Computer Engineering
 {lipor,girasole}@umich.edu

I. INTRODUCTION

The union of subspaces (UoS) model, in which data vectors lie near one of several subspaces, has been used actively in the computer vision community [1], [2]. Subspace clustering algorithms such as Sparse Subspace Clustering (SSC) [3] achieve state-of-the-art unsupervised clustering performance, but in many cases supervision is possible, e.g., a person could answer whether two images are of the same person or whether two digits are the same.

Pairwise-constrained clustering (PCC) algorithms use supervision in the form of *must-link* and *cannot-link* constraints on pairs of samples in the dataset. *Active* query selection methods, in which sample pairs are intelligently selected based on a number of heuristics, have been shown to make best use of supervised input. Active methods such as Uncertainty Reducing Active Spectral Clustering (URASC) [4] significantly reduce clustering error with a modest number of pairwise constraints, but do not take advantage of any structure in the data.

We introduce a method of active query selection that leverages the known UoS structure to improve pairwise-constrained clustering. We describe two ways to measure the margin of a point in the subspace clustering context using the subspace residual as well as margin based on the affinity matrix. For margin based on subspace residual, we prove that points lying near the intersection of subspaces are points with low margin under a common noise scenario. We present a novel PCC algorithm, called SUPERPAC (SUBspace CLUSTERing with Pairwise Active Constraints), that utilizes the subspace estimates provided by algorithms such as SSC [3] to reduce clustering error as quickly as possible.

II. UOS-BASED PAIRWISE-CONSTRAINED CLUSTERING

Let $\mathcal{X} = \{x_i \in \mathbb{R}^D\}_{i=1}^N$ be a set of data points lying on a union of K subspaces $\{\mathcal{S}_k\}_{k=1}^K$, each having dimension d . Denote the true clustering of a point $x \in \mathcal{X}$ by $C(x)$. Let the output of a clustering algorithm be an affinity/similarity matrix A and a set of label estimates $\{\hat{C}(x_i)\}_{i=1}^N$. Our algorithm for PCC consists of an initialization and three main steps. To initialize, we build a set of certain sets \mathcal{Z} using an EXPLORE-like algorithm similar to that of [5]. A test point is then obtained using either the *min-margin* criterion for subspaces [6] or a notion of margin based on the affinity matrix, both of which we define below. The test point is queried against representatives from the certain sets until a *must-link* is found; otherwise test point becomes its own certain set. Finally, the certain sets are used to impute values in the affinity matrix, and spectral clustering is performed. These steps (excluding the initialization) are then repeated until the maximum number of pairwise comparisons has been obtained. A diagram of our algorithm is given in Fig. 1. Pseudocode is given in Fig. 2, where UOS-EXPLORE is an initialization algorithm described in [7].

III. SAMPLE SELECTION VIA MARGIN

The key step in our algorithm is choosing a test point that is likely to be misclassified. Min-margin points have been studied extensively

in active learning; intuitively, these are points that lie near the decision boundary of the current classifier. We now describe two notions of margin—one based on subspace distances and one based on the input affinity matrix.

For a subspace \mathcal{S}_k with orthonormal basis U_k , let the distance of a point to that subspace be $\text{dist}(x, \mathcal{S}_k) = \min_{y \in \mathcal{S}_k} \|x - y\|_2 = \|x - U_k U_k^T x\|_2$. Let $k^* = \arg \min_{k \in [K]} \text{dist}(x, \mathcal{S}_k)$ be the index of the closest subspace. Then the subspace margin of a point $x \in \mathcal{X}$ is defined as [6]

$$\mu_1(x) = \max_{j \neq k^*, j \in [K]} \frac{\text{dist}(x, \mathcal{S}_{k^*})}{\text{dist}(x, \mathcal{S}_j)}. \quad (1)$$

We denote our algorithm with this definition of margin by SUPERPAC-R (residual). The point of minimum margin is then defined as $\arg \max_{x \in \mathcal{X}} \mu_1(x)$. Note that $0 \leq \mu_1(x) \leq 1$, where the larger $\mu_1(x)$ the closer x is to the decision boundary. The following theorem shows that points lying near the intersection of subspaces are included among those of minimum margin with high probability.

Theorem 1. Consider two d -dimensional subspaces \mathcal{S}_1 and \mathcal{S}_2 . Let $y = x + n$, where $x \in \mathcal{S}_1$ and $n \sim \mathcal{N}(0, \sigma^2 I_D)$. Define $\mu(y) = \frac{\text{dist}(y, \mathcal{S}_1)}{\text{dist}(y, \mathcal{S}_2)}$. Then

$$\frac{(1 - \varepsilon) \sqrt{\sigma^2(D - d)}}{(1 + \varepsilon) \sqrt{\sigma^2(D - d) + \text{dist}(x, \mathcal{S}_2)^2}} \leq \mu(y)$$

and

$$\mu(y) \leq \frac{(1 + \varepsilon) \sqrt{\sigma^2(D - d)}}{(1 - \varepsilon) \sqrt{\sigma^2(D - d) + \text{dist}(x, \mathcal{S}_2)^2}},$$

with probability at least $1 - 4e^{-c\varepsilon^2(D-d)}$, where c is an absolute constant.

Thm. 1 can be interpreted as follows: suppose we have two noisy points y_1, y_2 that were drawn from a subspace, i.e., $y_i = x_i + n_i$ where $x_i \in \mathcal{S}_1$, $i = 1, 2$. Then the point closer to the intersection with another subspace \mathcal{S}_2 will have smaller margin as long as the gap between $\text{dist}(x_1, \mathcal{S}_2)^2$ and $\text{dist}(x_2, \mathcal{S}_2)^2$ grows linearly with the noise variance σ^2 .

We also consider a version of margin calculated from the entries of the affinity matrix itself, denoted by SUPERPAC-A (affinity). Given an affinity matrix A , we estimate the probability that a point x_i is in subspace k as $\mathbb{P}(k|x_i) = \sum_{j: \hat{C}(x_j)=k} A_{ij} / \sum_{j=1}^N A_{ij}$. Let $l^* = \arg \max_{l \in [K]} \mathbb{P}(l|x)$. Define the affinity margin of a point x as

$$\mu_2(x) = \max_{j \neq l^*, j \in [K]} \frac{\mathbb{P}(j|x)}{\mathbb{P}(l^*|x)}. \quad (2)$$

The point of minimum margin again follows as $\arg \max_{x \in \mathcal{X}} \mu_2(x)$.

We show the error as a function of number of pairwise constraints in Figs. 2-4. Our method achieves superior performance compared to the state-of-the-art [4] and random query selection. Further, it is agnostic to the input algorithm chosen, meaning that our approach remains applicable even as unsupervised algorithms are improved.

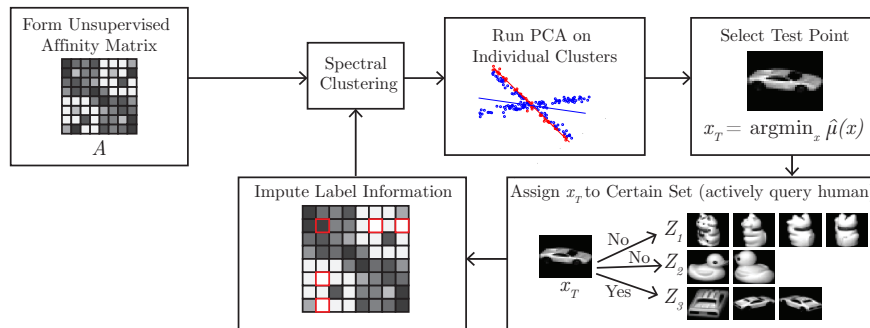


Fig. 1: Diagram of SUPERPAC algorithm for pairwise constrained clustering.

Algorithm 1 SUPERPAC

- 1: **Input:** $\mathcal{X} = \{x_1, x_2, \dots, x_N\}$: data, K : number of clusters, d : subspace dimension, A : affinity matrix, maxQueries: maximum number of pairwise comparisons
- 2: **Estimate Labels:** $\hat{C} \leftarrow \text{SPECTRALCLUSTERING}(A, K)$
- 3: **Initialize Certain Sets:** Initialize $\mathcal{Z} = \{Z_1, \dots, Z_{n_c}\}$ and numQueries via UOS-EXPLORE
- 4: **while** numQueries < maxQueries **do**
- 5: **Obtain Test Point:** select $x_T \leftarrow \arg \max_{x \in \mathcal{X}} \mu_1(x)$ or $\arg \max_{x \in \mathcal{X}} \mu_2(x)$
- 6: **Assign x_T to Certain Set:**
 Sort $\{Z_1, \dots, Z_{n_c}\}$ in order of most likely must-link (via subspace residual for x_T), query x_T against representatives from Z_k until must-link constraint is found or $k = n_c$. If no must-link constraint is found, set $\mathcal{Z} \leftarrow \{Z_1, \dots, Z_{n_c}, \{x_T\}\}$ and increment n_c .
- 8: **Impute Constraints:** Set $A_{ij} = A_{ji} = 1$ for (x_i, x_j) in the same certain set and $A_{ij} = A_{ji} = 0$ for (x_i, x_j) in different certain sets (do not impute for points absent from certain sets)
- 9: **Estimate Labels:** $\hat{C} \leftarrow \text{SPECTRALCLUSTERING}(A, K)$
- 10: **end while**

Fig. 2: SUPERPAC Algorithm

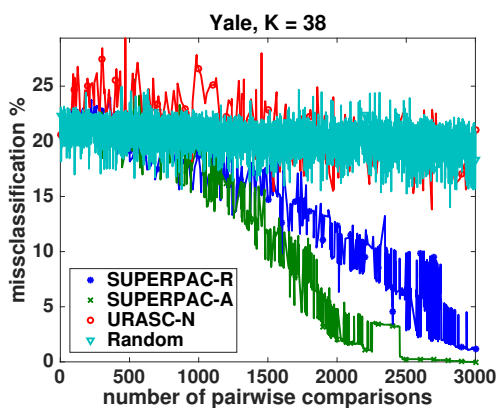


Fig. 3: Misclassification rate versus number of pairwise comparisons for extended Yale face database B with $K = 38$ subjects. Input affinity matrix is taken from SSC-OMP.

REFERENCES

- [1] R. Basri and D. Jacobs, "Lambertian reflectance and linear subspaces," *IEEE TPAMI*, vol. 25, no. 2, pp. 218–233, February 2003.
- [2] N. Oliver, B. Rosario, and A. Pentland, "A bayesian computer vision

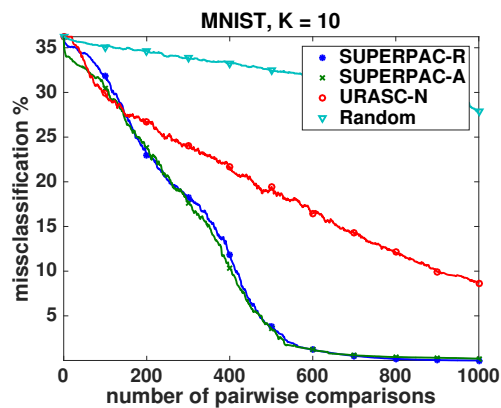


Fig. 4: Misclassification rate versus number of pairwise comparisons for MNIST handwritten digits with $K = 10$ and $N_k = 100$ points per subspace. Input affinity matrix is taken from SSC.

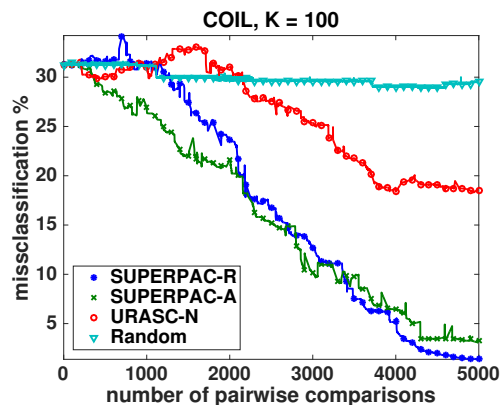


Fig. 5: Misclassification rate versus number of pairwise comparisons for COIL-100 ($K = 100$) database. Input affinity matrix is taken from EnSC.

system for modeling human interactions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 831–843, 2000.

- [3] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 35, pp. 2765–2781, Nov. 2013.
- [4] C. Xiong, D. M. Johnson, and J. J. Corso, "Active clustering with model-based uncertainty reduction," *IEEE Trans. Pattern Anal. Mach. Intelligence*, 2016, accepted for publication.
- [5] S. Basu, A. Banerjee, and R. J. Mooney, "Active semi-supervision for pairwise constrained clustering," in *Proc. SIAM Int. Conf. on Data Mining*, 2004.
- [6] J. Lipor and L. Balzano, "Margin-based active subspace clustering," in *Proc. Int. Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, 2015.
- [7] —, "Leveraging union of subspace structure to improve constrained clustering," *arXiv preprint arXiv:1608.02146*, 2016.