Supplementary Documents for "Semi-Crowdsource Clustering: Generalizing Crowd Labeling by Robust Distance Metric Learning"

Jinfeng Yi[†], Rong Jin[†], Anil K. Jain[†], Shaili Jain[♮], Tianbao Yang[‡]
[†]Michigan State University, East Lansing, MI 48824, USA
[♮]Yale University, New Haven, CT 06520, USA
[‡]Machine Learning Lab, GE Global Research, San Ramon, CA 94583, USA
{yijinfen, rongjin, jain}@cse.msu.edu, shaili.jain@yale.edu, tyang@ge.com

1 Theoretical Analysis for Perfect Recovery using Equation (2)

The following discussion about the perfect recovery result using Eq. (2) comes from [3]. We repeat it in the supplementary document for the completeness of this study.

To discuss the perfect recovery result for using Eq. (2), we first need to make a few assumptions about A^* besides its low rank. Let A^* be a low-rank matrix of rank r, with a singular value decompsition $A^* = U\Sigma V^\top$, where $U = (\mathbf{u}_1, \ldots, \mathbf{u}_r) \in \mathbb{R}^{N\times r}$ and $V = (\mathbf{v}_1, \ldots, \mathbf{v}_r) \in \mathbb{R}^{N\times r}$ are the left and right eigenvectors of A^*, satisfying the following incoherence assumptions.

- **A1** The row and column spaces of A^* have coherence bounded above by some positive number μ_0, i.e.,

$$\max_{i\in[N]} \|P_U(\mathbf{e}_i)\|_2^2 \le \frac{\mu_0 r}{N}, \quad \max_{i\in[N]} \|P_V(\mathbf{e}_i)\|_2^2 \le \frac{\mu_0 r}{N}$$

  where $\mathbf{e}_i$ is the standard basis vector.

- **A2** The matrix $E = UV^\top$ has a maximum entry bounded by $\frac{\mu_1\sqrt{r}}{N}$ in absolute value for some positive μ_1, i.e. $|E_{i,j}| \le \frac{\mu_1\sqrt{r}}{N}, \forall (i,j) \in [N]\times[N]$,

where P_U and P_V denote the orthogonal projections on the column space and row space of A^*, respectively, i.e.

$$P_U = UU^\top, \quad P_V = VV^\top$$

To state our theorem, we need to introduce a few notations. Let $\xi(A')$ and $\mu(A')$ denote the low-rank and sparsity incoherence of matrix A' defined by [1], i.e.

$$\xi(A') = \max_{E\in T(A'), \|E\|\le 1} \|E\|_\infty \tag{1}$$

$$\mu(A') = \max_{E\in \Omega(A'), \|E\|_\infty\le 1} \|E\| \tag{2}$$

where $T(A')$ denotes the space spanned by the elements of the form $\mathbf{u}_k\mathbf{y}^\top$ and $\mathbf{x}\mathbf{v}_k^\top$, for $1 \le k \le r$, $\Omega(A')$ denotes the space of matrices that have the same support to A', $\|\cdot\|$ denotes the spectral norm and $\|\cdot\|_\infty$ denotes the largest entry in magnitude.

Lemma 1. *Let $A^* \in \mathbb{R}^{N\times N}$ be a similarity matrix of rank r obeying the incoherence properties (**A1**) and (**A2**), with $\mu = \max(\mu_0, \mu_1)$. Suppose we observe m_1 entries of A^* recorded in $\tilde{A}$*

with locations sampled uniformly at random, denoted by $\mathcal{S}$. Under the assumption that m_0 entries randomly sampled from m_1 observed entries are corrupted, denoted by Ω, i.e. $A^*_{ij} \neq \tilde{A}_{ij}, (i,j) \in \Omega$. Given $\mathcal{P}_{\mathcal{S}}(\tilde{A}) = \mathcal{P}_{\mathcal{S}}(A^* + E^*)$, where E^* corresponds to the corrupted entries in Ω. With

$$\mu(E^*)\xi(A^*) \leq \frac{1}{4r+5}, \quad m_1 - m_0 \geq C_1 \mu^4 n (\log n)^2,$$

and C_1 is a constant, we have, with a probability at least $1 - N^{-3}$, the solution $(A', E) = (A^*, E^*)$ is the unique optimizer to (2) provided that

$$\frac{\xi(A^*) - (2r-1)\xi^2(A^*)\mu(E^*)}{1 - 2(r+1)\xi(A^*)\mu(E^*)} < \lambda < \frac{1 - (4r+5)\xi(A^*)\mu(E^*)}{(r+2)\mu(E^*)}$$

2 Proof of Theorem 1

To prove Theorem 1, we need the following theorem for matrix concentration.

Lemma 2. *(Lemma 2 from [2]) Let $\mathcal{H}$ be a Hilbert space and ξ be a random variable on (Z, ρ) with values in $\mathcal{H}$. Assume $\|\xi\| \leq M < \infty$ almost surely. Denote $\sigma^2(\xi) = \mathrm{E}(\|\xi\|^2)$. Let $\{z_i\}_{i=1}^m$ be independent random drawers of ρ. For any $0 < \delta < 1$, with confidence $1 - \delta$,*

$$\left\| \frac{1}{m} \sum_{i=1}^m (\xi_i - \mathrm{E}[\xi_i]) \right\| \leq \frac{4M \ln(2/\delta)}{\sqrt{m}}$$

Using the assumption that $|\mathbf{x}|_2 \leq 1$ and Lemma 2, we have, with a probability $1 - n^{-3}$,

$$\left| \frac{1}{m} \widehat{X}\widehat{X}^\top - C_X \right|_2 \leq \frac{12 \ln n}{\sqrt{n}}$$

and therefore

$$\left| \left(\frac{1}{m} \widehat{X}\widehat{X}^\top + \lambda I \right)^{-1} - (C_X + \lambda I)^{-1} \right|_2 \leq \frac{12 \ln n}{\lambda \sqrt{n}}$$

Second, according to Lemma 1, with a probability $1 - n^{-3}$, we have $\widehat{A} = YY^\top$ and therefore $\widehat{X}\widehat{A}\widehat{X}^\top = \widehat{X}YY^\top\widehat{X}^\top$. Again, using the matrix concentration theory, we have, with a probability $1 - n^{-3}$,

$$\left| \frac{1}{m} \widehat{X}Y - B \right|_2 \leq \frac{12 \ln n}{\sqrt{n}}$$

Finally, we rewrite $|M_s - \widehat{M}_s|_2$ as

$$\begin{aligned}
& \|M_s - \widehat{M}_s\|_2 \\
\leq & \left| M_s - \left(\frac{1}{m} \widehat{X}\widehat{X}^\top + \lambda I \right)^{-1} BB^\top C_X \right|_2 + \\
& \left| \left(\frac{1}{m} \widehat{X}\widehat{X}^\top + \lambda I \right)^{-1} BB^\top C_X - \left(\frac{1}{m} \widehat{X}\widehat{X}^\top + \lambda I \right) BB^\top \left(\frac{1}{m} \widehat{X}\widehat{X}^\top + \lambda I \right)^{-1} \right|_2 + \\
& \left| \left(\frac{1}{m} \widehat{X}\widehat{X}^\top + \lambda I \right)^{-1} BB^\top \left(\frac{1}{m} \widehat{X}\widehat{X}^\top + \lambda I \right)^{-1} - \left(\frac{1}{m} \widehat{X}\widehat{X}^\top + \lambda I \right)^{-1} \frac{\widehat{X}Y}{m} B^\top \left(\frac{1}{m} \widehat{X}\widehat{X}^\top + \lambda I \right)^{-1} \right| + \\
& \left| \left(\frac{1}{m} \widehat{X}\widehat{X}^\top + \lambda I \right)^{-1} \frac{\widehat{X}Y}{m} B^\top \left(\frac{1}{m} \widehat{X}\widehat{X}^\top + \lambda I \right)^{-1} - \widehat{M}_s \right|
\end{aligned}$$

It is easy to see that with a probability $1 - 3n^{-3}$, each term on the right hand side of the above inequality is bounded by $\frac{12 \ln n}{\lambda^2 \sqrt{n}}$, leading to the result of the theorem.

References

[1] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky. Rank-sparsity incoherence for matrix decomposition. volume 21, pages 572–596, 2011.

[2] S. Smale and D.-X. Zhou. Geometry on probability spaces. *Constr Approx*, 30:311–323, 2009.

[3] J. Yi, R. Jin, A. K. Jain, and S. Jain. Crowdclustering with sparse pairwise labels: A matrix completion approach. In *AAAI Workshop on Human Computation*, 2012.