Robust Convex Clustering Analysis

Supplemental Materials

Qi Wang, Pinghua Gong, Shiyu Chang, Thomas S. Huang, Jiayu Zhou

APPENDIX

Here we show the proof of Theorem 1.

Assume a data matrix $X \in \mathbb{R}^{d\times n}$ with n samples and d features satisfy:

$$X = P^* + Q^* + \delta, \tag{12}$$

where $P^* \in \mathbb{R}^{d\times n}$ and $Q^* \in \mathbb{R}^{d\times n}$ are the underlying true decomposition of X. P^* is the component containing cluster information of X. Q^* is the component containing outliers. $\delta \in \mathbb{R}^{d\times n}$ is the stochastic noise matrix. For the i^{th} row of δ, $\delta_{i,j} \sim \mathcal{N}(0, \sigma^2)$, where $i \in \mathbb{N}_d$ and $j \in \mathbb{N}_n$, i.e., for each feature there exists a normal distributed noise for all data points.

The optimization problem in Eq. (2) is:

$$(\hat{P}, \hat{Q}) = \underset{P,Q}{\operatorname{argmin}} \frac{1}{2}\|X - (P + Q)\|_F^2 + \alpha\|P\|_{FU} + \beta\|Q\|_{2,1}, \tag{13}$$

where $\hat{P} \in \mathbb{R}^{d\times n}$ and $\hat{Q} \in \mathbb{R}^{d\times n}$ are the optimal solution pair obtained by solving Eq. (13) and $\|P\|_{FU}$ is defined as:

$$\|P\|_{FU} = \sum_{i<j} w_{i,j}\|P_i - P_j\|_2.$$

We first present a theorem for the optimal solution pair which are important for our following theoretical analysis.

Theorem 2. *Consider the optimization problem in Eq. (2) for $n, d > 1$. Take the regularization parameters α and β as:*

$$\alpha > \lambda, \beta > \frac{\lambda}{\sqrt{n}}, \lambda = \sigma\sqrt{n(d+t)}, \tag{14}$$

where t is a positive scalar and σ is the standard deviation for each row of δ. Then with a probability of at lease $1 - \exp(-\frac{1}{2}(t - d\log(1 + \frac{t}{d})))$, for a global minimizer $\hat{P}, \hat{Q}$ in Eq. (13) we have:

$$\frac{1}{2}\|(P^* + Q^*) - (\hat{P} + \hat{Q})\|_F^2 \le \alpha(gn + 1)\|(\hat{P} - P^*)^T\|_{2,1} + 2\beta\|\hat{Q} - Q^*\|_{2,1}, \tag{15}$$

where P^* and Q^* are the ground truth that generates the data matrix X and g is the maximum value of weight $w_{i,j}$.

Proof. Since $\hat{P}$ and $\hat{Q}$ are global minimizers, from Eq. (13) we have:

$$\begin{aligned}&\frac{1}{2}\|X - (\hat{P} + \hat{Q})\|_F^2 + \alpha\|\hat{P}\|_{FU} + \beta\|\hat{Q}\|_{2,1}\\ &\le \frac{1}{2}\|X - (P^* + Q^*)\|_F^2 + \alpha\|P^*\|_{FU} + \beta\|Q^*\|_{2,1},\end{aligned}$$

By substituting the assumption we made about X in Eq. (12) we obtain:

$$\begin{aligned}&\frac{1}{2}\|(P^* + Q^*) + \delta - (\hat{P} + \hat{Q})\|_F^2 + \alpha\|\hat{P}\|_{FU} + \beta\|\hat{Q}\|_{2,1}\\ &\le \frac{1}{2}\|\delta\|_F^2 + \alpha\|P^*\|_{FU} + \beta\|Q^*\|_{2,1}.\end{aligned} \tag{16}$$

Next we expand the first term in left hand side as:

$$\begin{aligned}&\frac{1}{2}\|(P^* + Q^*) + \delta - (\hat{P} + \hat{Q})\|_F^2\\ =&\frac{1}{2}\|(P^* + Q^*) - (\hat{P} + \hat{Q})\|_F^2 + \frac{1}{2}\|\delta\|_F^2\\ &+ \sum_{i,j}((P^* + Q^*) - (\hat{P} + \hat{Q}))\delta_{i,j}.\end{aligned} \tag{17}$$

By substituting Eq. (17) into Eq. (16) and rearranging all terms we obtain:

$$\begin{aligned}&\frac{1}{2}\|(P^* + Q^*) - (\hat{P} + \hat{Q})\|_F^2\\ &\le \alpha\|P^*\|_{FU} + \beta\|Q^*\|_{2,1} - \alpha\|\hat{P}\|_{FU} - \beta\|\hat{Q}\|_{2,1}\\ &\quad + \sum_{i,j}(\hat{P}_{i,j} - P^*_{ij})\delta_{i,j} + \sum_{i,j}(\hat{Q}_{i,j} - Q^*_{ij})\delta_{i,j}\\ &\le \alpha\|P^*\|_{FU} + \beta\|Q^*\|_{2,1} - \alpha\|\hat{P}\|_{FU} - \beta\|\hat{Q}\|_{2,1}\\ &\quad + \sum_j \|P^*_j - \hat{P}_j\|_2\|\delta_j\|_2 + \sum_i \|Q^*_i - \hat{Q}_i\|_2\|\delta_i\|_2,\end{aligned} \tag{18}$$

where the last line we used Cauchy-Schwartz inequality. P^*_j, $\hat{P}_j$,δ_j are the j^{th} column of P^*, $\hat{P}$ and δ respectively and Q^*_i, $\hat{Q}_i$,δ_i are the i^{th} row of Q^*, $\hat{Q}$ and δ respectively.

Next we compute the upper bounds for $\|\delta_i\|_2$ where i is the row index with the following Lemma about chi-squared random variable.

Lemma 1. *Let $\chi^2(d)$ be a chi-squared random variable with d degrees of freedom. Then the following holds [8]:*

$$Pr(\chi^2(d) \ge d + t) \le \exp(-\frac{1}{2}(t - d\ \log(1 + \frac{t}{d}))), t > 0. \tag{19}$$

Hence with the probability of at least $1 - \exp(-\frac{1}{2}(t - d\log(1 + \frac{t}{d})))$, we have $\|\delta_i\|_2 \le \sigma\sqrt{(d+t)} = \lambda/(\sqrt{n})$.

Similarly, we can compute the upper bound for $\|\delta_j\|_2$ where j is the column index:

$$\|\delta_j\|_2 \le \sqrt{\|\delta\|_F^2} \le \sigma\sqrt{n(d+t)} = \lambda. \tag{20}$$

Substitute the upper bound of $\|\delta_i\|_2$ and $\|\delta_j\|_2$ into Eq. (18) we obtain:

$$\begin{aligned}&\frac{1}{2}\|(P^*+Q^*) - (\hat{P} + \hat{Q})\|_F^2\\ &\le \alpha\|P^*\|_{FU} + \beta\|Q^*\|_{2,1} - \alpha\|\hat{P}\|_{FU} - \beta\|\hat{Q}\|_{2,1}\\ &\quad + \lambda\sum_j \|P^*_j - \hat{P}_j\|_2 + \frac{\lambda}{\sqrt{n}}\sum_i \|Q^*_i - \hat{Q}_i\|_2.\end{aligned}$$

Now take the regularization parameter as Eq. (14), we obtain:

$$\begin{aligned}&\frac{1}{2}\|(P^* + Q^*) - (\hat{P} + \hat{Q})\|_F^2\\ &\le \alpha\|P^*\|_{FU} + \beta\|Q^*\|_{2,1} - \alpha\|\hat{P}\|_{FU} - \beta\|\hat{Q}\|_{2,1}\\ &\quad + \alpha\|(\hat{P} - P^*)^T\|_{2,1} + \beta\|\hat{Q} - Q^*\|_{2,1}.\end{aligned} \tag{21}$$

We divide the right hand side of Eq. (21) into two parts, the terms that only contain P and the terms that only contain

Q, and bound both term accordingly. First we bound the part only contain P:

$$
\begin{aligned}
&\alpha\|P^*\|_{FU} - \alpha\|\hat{P}\|_{FU} + \alpha\|(P^* - \hat{P})^T\|_{2,1} \\
=&\alpha \sum_{i<j} w_{i,j}(\|P^*_i - P^*_j\|_2 - \|\hat{P}_i - \hat{P}_j\|_2) + \alpha\|(P^* - \hat{P})^T\|_{2,1} \\
\leq&\alpha \sum_{i<j} w_{i,j}\|P^*_i - P^*_j - \hat{P}_i + \hat{P}_j\|_2 + \alpha\|(P^* - \hat{P})^T\|_{2,1} \\
\leq&\alpha \sum_{i<j} w_{i,j}(\|P^*_i - \hat{P}_i\|_2 + \|\hat{P}_j - P^*_j\|_2) + \alpha\|(P^* - \hat{P})^T\|_{2,1} \\
\leq&\alpha \sum_{i<j} g(\|P^*_i - \hat{P}_i\|_2 + \|\hat{P}_j - P^*_j\|_2) + \alpha\|(P^* - \hat{P})^T\|_{2,1} \\
\leq&\alpha gn\|(P^* - \hat{P})^T\|_{2,1} + \alpha\|(P^* - \hat{P})^T\|_{2,1} \\
\leq&\alpha(gn+1)\|(\hat{P} - P^*)^T\|_{2,1},
\end{aligned}
\tag{22}
$$

where indices i and j are both used to denote the column of P. Next the following terms that only contain Q can be bounded:

$$
\beta\|Q^*\|_{2,1} - \beta\|\hat{Q}\|_{2,1} + \beta\|\hat{Q} - Q^*\|_{2,1} \leq 2\beta\|\hat{Q} - Q^*\|_{2,1}. \tag{23}
$$

Combine Eq. (22), Eq. (23) and Eq. (21), we obtain:

$$
\begin{aligned}
\frac{1}{2}\|(P^* + Q^*) - (\hat{P} + \hat{Q})\|_F^2 \leq& \alpha(gn+1)\|(\hat{P} - P^*)^T\|_{2,1} \\
&+ 2\beta\|\hat{Q} - Q^*\|_{2,1}.
\end{aligned}
\tag{24}
$$

This completes the proof of Theorem 2. □

Before we move on, we introduce two notations: $\mathcal{J}(\cdot)$ and $\mathcal{J}_\perp(\cdot)$ denote the index sets for nonzero rows and zero rows and the definitions are:

$$
\mathcal{J}(Q) = \{i | Q_i \neq 0\}, \mathcal{J}_\perp(Q) = \{i | Q_i = 0\}. \tag{25}
$$

Next we make the following assumptions about the X, P^* and Q^* and give a theoretical bound between $P^* + Q^*$ and $\hat{P} + \hat{Q}$ based on these assumptions:

Assumption 1. *For a matrix pair $\Gamma_P \in \mathbb{R}^{d\times n}$ and $\Gamma_Q \in \mathbb{R}^{d\times n}$, let r and c ($1 \leq r \leq n$, $1 \leq c \leq d$) be the upper bounds of $|\mathcal{J}((P^*)^T)|$ and $|\mathcal{J}(Q^*)|$, respectively. Let γ_1 and γ_2 be positive scalars. We assume there exists positive scalars $\kappa_1(r)$ and $\kappa_2(c)$ such that:*

$$
\kappa_1(r) = \min_{\Gamma_P, \Gamma_Q \in R(r,c)} = \frac{\|\Gamma_P + \Gamma_Q\|_F}{\|(\Gamma_P^T)^{\mathcal{J}(P^T)}\|_{2,1}} > 0, \tag{26}
$$

$$
\kappa_2(c) = \min_{\Gamma_P, \Gamma_Q \in R(r,c)} = \frac{\|\Gamma_P + \Gamma_Q\|_F}{\|(\Gamma_Q)^{\mathcal{J}(Q)}\|_{2,1}} > 0, \tag{27}
$$

where $R(r,c)$ is defined as:

$$
\begin{aligned}
R(r,c) =&\{\Gamma_P \in \mathbb{R}^{d\times n}, \Gamma_Q \in \mathbb{R}^{d\times n} | \Gamma_P \neq 0, \Gamma_Q \neq 0, | \\
&|\mathcal{J}(P^T)| \leq r, |\mathcal{J}(Q)| \leq c, \\
&\|(\Gamma_P^T)^{\mathcal{J}_\perp(P^T)}\|_{2,1} \leq \gamma_1 \|(\Gamma_P^T)^{\mathcal{J}(P^T)}\|_{2,1}, \\
&\|(\Gamma_Q)^{\mathcal{J}_\perp(Q)}\|_{2,1} \leq \gamma_2 \|(\Gamma_Q)^{\mathcal{J}(Q)}\|_{2,1}\},
\end{aligned}
$$

where $\mathcal{J}(\cdot)$ and $\mathcal{J}_\perp(\cdot)$ are defined in Eq. (25) and $|\mathcal{J}|$ denotes the number of elements in the set $\mathcal{J}$.

Lemma 2. *Under Assumption 1 the following results hold with probability of at least $1 - \exp(-\frac{1}{2}(t - d\log(1 + \frac{t}{d})))$ $(t > 0)$:*

$$
\|(P^* - \hat{P})^T\|_{2,1} \leq \frac{(\gamma_1 + 1)\sqrt{r}}{\kappa_1(r)} \left(\frac{2\alpha\sqrt{r}}{\kappa_1(r)} + \frac{2\beta\sqrt{c}}{\kappa_2(c)} \right), \tag{28}
$$

$$
\|Q^* - \hat{Q}\|_{2,1} \leq \frac{(\gamma_2 + 1)\sqrt{c}}{\kappa_2(c)} \left(\frac{2\alpha\sqrt{r}}{\kappa_1(r)} + \frac{2\beta\sqrt{c}}{\kappa_2(c)} \right). \tag{29}
$$

The proof of a similar lemma can be found in [8].

Theorem 1 can be obtained by substituting Eq. (28) and Eq. (29) into Eq. (15).