Supplementary Document for Regularized ℓ^1-Graph for Data Clustering

Yingzhen Yang[1]
yyang58@ifp.uiuc.edu
Zhangyang Wang[1]
zwang119@ifp.uiuc.edu
Jianchao Yang[2]
jiayang@adobe.com
Jiawei Han[1]
hanj@cs.uiuc.edu
Thomas S. Huang[1]
huang@ifp.uiuc.edu

[1] University of Illinois at Urbana-Champaign
Urbana, IL 61801, USA
[2] Adobe Research
San Jose, CA 95110, USA

1 Proof of Proposition 1

The optimization problem in the preliminary form is:

$$\begin{aligned} \min_{\boldsymbol{\alpha},\mathbf{W}} \quad & \sum_{i=1}^{n} \|\mathbf{x}_i - \boldsymbol{X}\boldsymbol{\alpha}^i\|_2^2 + \lambda\|\boldsymbol{\alpha}^i\|_1 + \gamma \mathrm{Tr}(\boldsymbol{\alpha}\mathbf{L}_{\mathbf{W}}\boldsymbol{\alpha}^T) \\ & s.t.\ \mathbf{W} = (\mathbf{A}\circ|\boldsymbol{\alpha}| + \mathbf{A}^T\circ|\boldsymbol{\alpha}^T|)/2 \quad \boldsymbol{\alpha}\in S \end{aligned} \tag{1}$$

where $S := \{\boldsymbol{\alpha} \in \mathbb{R}^{n\times n} | \boldsymbol{\alpha}_{ii} = 0, 1 \le i \le n\}$, $\lambda > 0$ is the weight controlling the sparsity of the coefficients, and $\gamma > 0$ is the weight of the regularization term.

We reformulate 1) into the following optimization problem with simplified equality constraint:

$$\begin{aligned} \min_{\boldsymbol{\alpha},\mathbf{W}} \quad & \sum_{i=1}^{n} \|\mathbf{x}_i - \boldsymbol{X}\boldsymbol{\alpha}^i\|_2^2 + \lambda\|\boldsymbol{\alpha}^i\|_1 + \mathrm{Tr}(\boldsymbol{\alpha}\mathbf{L}_{\mathbf{A}\circ|\mathbf{W}|}\boldsymbol{\alpha}^T) \\ & s.t.\ \mathbf{W} = \boldsymbol{\alpha} \quad \boldsymbol{\alpha}\in S \end{aligned} \tag{2}$$

Proposition 1 establishes the equivalence between the problem (2) and problem (1).

Proposition 1. *The solution $\boldsymbol{\alpha}^*$ to the problem (2) is also the solution to the problem (1), and vice versa.*

Proof. When $\mathbf{W} = \boldsymbol{\alpha}$,

$$\begin{aligned}
\mathrm{Tr}(\boldsymbol{\alpha}\mathbf{L}_{\mathbf{A}\circ|\mathbf{W}|}\boldsymbol{\alpha}^T) &= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{A}_{ij}|\mathbf{W}_{ij}|\|\boldsymbol{\alpha}^i - \boldsymbol{\alpha}^j\|_2^2 \\
&= \frac{1}{4}\sum_{i,j=1}^{n}\mathbf{A}_{ij}|\boldsymbol{\alpha}_{ij}|\|\boldsymbol{\alpha}^i - \boldsymbol{\alpha}^j\|_2^2 + \frac{1}{4}\sum_{i,j=1}^{n}\mathbf{A}_{ji}|\boldsymbol{\alpha}_{ji}|\|\boldsymbol{\alpha}^j - \boldsymbol{\alpha}^i\|_2^2 \\
&= \frac{1}{2}\sum_{i,j=1}^{n}\frac{\mathbf{A}_{ij}|\boldsymbol{\alpha}_{ij}| + \mathbf{A}_{ji}|\boldsymbol{\alpha}_{ji}|}{2}\|\boldsymbol{\alpha}^i - \boldsymbol{\alpha}^j\|_2^2
\end{aligned}$$

Therefore, the objective function of problem (2) equals to that of problem (1) with their $\mathbf{W}$ determined by the corresponding equality constraints, so these two optimization problems are equivalent to each other. □

2 Solving the Optimization Problem (13) by ADMM

$$\begin{aligned}
&\min_{\boldsymbol{\alpha}^i \in \mathbf{R}^n} F(\boldsymbol{\alpha}^i) = \frac{1}{2}{\boldsymbol{\alpha}^i}^T\mathbf{P}_i\boldsymbol{\alpha}^i + \mathbf{b}_i^T\boldsymbol{\alpha}^i + \lambda\|\boldsymbol{\alpha}^i\|_1 \\
&s.t.\ \boldsymbol{\alpha}_{ii} = 0
\end{aligned} \tag{3}$$

To solve (3) by ADMM, (3) is rewritten as below by introducing an auxiliary variable $\mathbf{z}$

$$\begin{aligned}
&\min_{\boldsymbol{\alpha}^i \in \mathbf{R}^n} F(\boldsymbol{\alpha}^i) = \frac{1}{2}{\boldsymbol{\alpha}^i}^T\mathbf{P}_i\boldsymbol{\alpha}^i + {\mathbf{b}_i}^T\boldsymbol{\alpha}^i + \lambda\|\mathbf{z}\|_1 \\
&s.t.\ \boldsymbol{\alpha}_i = \mathbf{z}, \mathbf{z}_i = 0
\end{aligned} \tag{4}$$

Note that $\boldsymbol{\alpha}^i$ is now not involved in the ℓ^1-norm. The advantage of ADMM for the lasso problem is that it transforms the original problem into a sequence of subproblems where closed-form solutions exist, and the lasso problem can be solved efficiently in an iterative manner.

The augmented Lagrangian for the constrained convex optimization problem (4) is

$$\mathcal{L}^i(\boldsymbol{\alpha}^i, \mathbf{z}, \mathbf{y}) = \frac{1}{2}{\boldsymbol{\alpha}^i}^T\mathbf{P}_i\boldsymbol{\alpha}^i + {\mathbf{b}_i}^T\boldsymbol{\alpha}^i + \lambda\|z\|_1 + \mathbf{y}^T(\boldsymbol{\alpha}^i - \mathbf{z}) + \frac{\mu}{2}\|\boldsymbol{\alpha}^i - \mathbf{z}\|_2^2 \tag{5}$$

where $\mathbf{y}$ is the Lagrangian multiplier, μ is the penalty parameter which is a pre-set positive constant. The ADMM iterations for the optimization of (4) are listed below, with k being the iteration index:

- Update $\boldsymbol{\alpha}^i$ with fixed $\mathbf{z}$ and $\mathbf{y}$:

$$(\boldsymbol{\alpha}^i)^{(k)} = \min_{\boldsymbol{\alpha}^i}\mathcal{L}^i(\boldsymbol{\alpha}^i, \mathbf{z}^{(k-1)}, \mathbf{y}^{(k-1)}) \tag{6}$$

  and (6) has closed-form solution

$$(\boldsymbol{\alpha}^i)^{(k)} = (\mathbf{P}_i + \mu\mathbf{I}_n)^{-1}(\mu\mathbf{z}^{(k-1)} - \mathbf{y}^{(k-1)} - \mathbf{b}_i) \tag{7}$$

- Update $\mathbf{z}$ with fixed $\boldsymbol{\alpha}^i$ and $\mathbf{y}$:

$$\mathbf{z}^{(k)} = \min_{\mathbf{z}} \mathcal{L}^i((\boldsymbol{\alpha}^i)^{(k)}, \mathbf{z}, \mathbf{y}^{(k-1)}) \tag{8}$$

By soft thresholding,

$$\mathbf{z}_t^{(k)} = \begin{cases} \max\{0, |\mu(\boldsymbol{\alpha}_t^i)^{(k)} + \mathbf{y}_t^{(k-1)}| - \lambda\} \cdot \frac{\mathrm{sign}(\mu(\boldsymbol{\alpha}_t^i)^{(k)} + \mathbf{y}_t^{k-1})}{\mu} & t \neq i \\ 0 & t = i \end{cases} \tag{9}$$

where the subscript t indicates the t-th element of the vector.

- Update $\mathbf{y}$:

$$\mathbf{y}^{(k)} = \mathbf{y}^{(k-1)} + \mu((\boldsymbol{\alpha}^i)^{(k)} - \mathbf{z}^{(k)}) \tag{10}$$

The ADMM iterates (7), (9) and (10) until both the primal residual $\|(\boldsymbol{\alpha}^i)^{(k)} - \mathbf{z}^{(k)}\|_2$ and the dual residual $\mu\|\mathbf{z}^{(k+1)} - \mathbf{z}^{(k)}\|_2$ are smaller than a threshold or the iteration number k achieves the pre-set maximum number. It has been proved that ADMM iterations converge to the optimal solution to the convex optimization problem (4) [1].

3 Time Complexity for Solving the Optimization Problem (7) by ADMM

Based on the previous section and the algorithm description in the paper, the subproblem (12) accounts for the most of the time complexity. Suppose the maximum iteration number of ADMM is N_1, and the maximum iteration number of the coordinate descent for solving subproblem (12) is N_2. In our implementation, we store the matrix inversion result in equation (7), and the time complexity for computing the inversion of a $n \times n$ matrix is $n^{2.376}$ by the CoppersmithÍCWinograd algorithm, the overall time complexity for solving the optimization Problem (7) by ADMM is $\mathcal{O}(N_1 n^{2.376} + N_1 N_2 n^2)$.

4 Supplementary Clustering Results

Figure 1 shows several examples from the ORL face database.

Figure 1: Example images of the ORL face database

We also examine the changes of the clustering performance on Yale face database with respect to γ and K, and illustrate the result in Figure 2 and Figure 3 respectively.

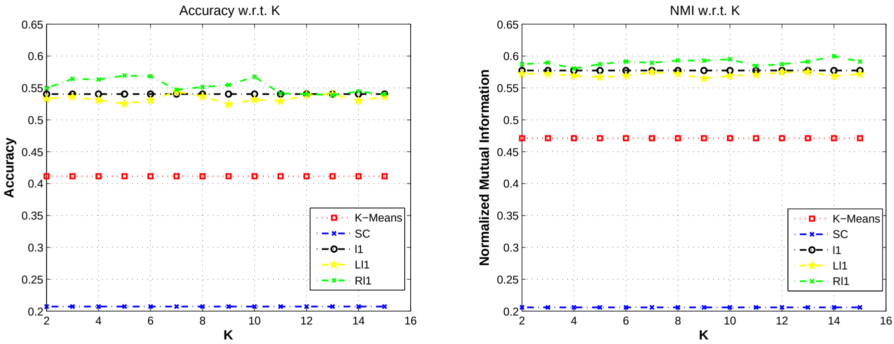

Figure 2: Clustering performance with different values of K, i.e. the number of nearest neighbors, on Yale face database when $\gamma = 0.5$. Left: Accuracy; Right: NMI

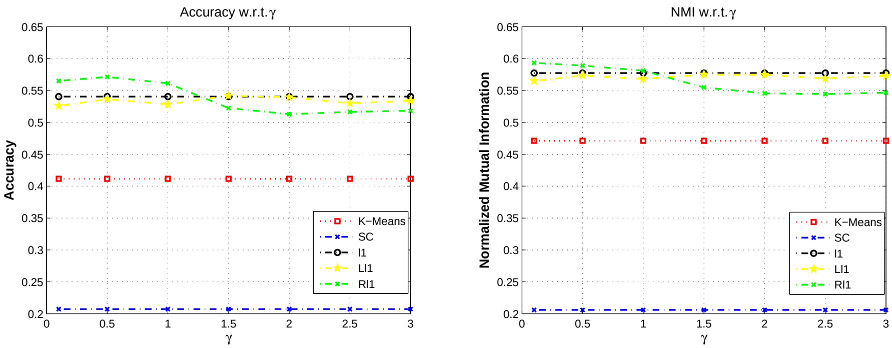

Figure 3: Clustering performance with different values of γ, i.e. the weight of the regularization term, on Yale face database when $K = 5$. Left: Accuracy; Right: NMI

5 Empirical Convergence

We show the convergence curve of ADMM for ORL face database and Yale face database in Figure 4.

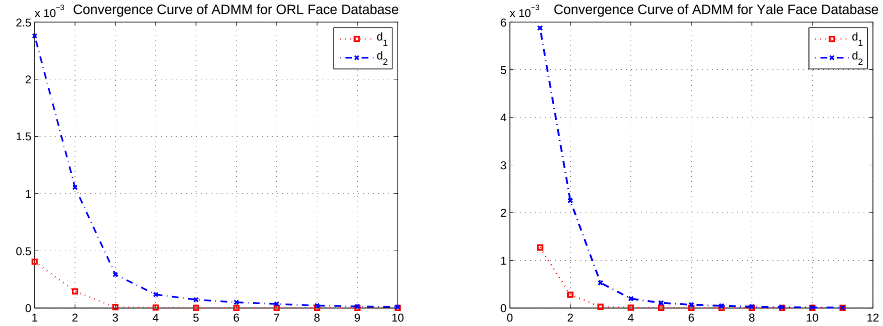

Figure 4: Convergence curves of ADMM for ORL (Left) and Yale (Right) face database, $d_1 = \|\mathbf{W} - \boldsymbol{\alpha}\|_F$, $d_2 = \|\boldsymbol{\alpha}^{(k)} - \boldsymbol{\alpha}^{(k-1)}\|_F$.

References

[1] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.*, 3(1):1–122, January 2011. ISSN 1935-8237.