Supplementary file to "Scaled Simplex Representation for Subspace Clustering"

Jun Xu[1,2], Mengyang Yu[2], Ling Shao[2], *Senior Member, IEEE*, Wangmeng Zuo[3], *Senior Member, IEEE*, Deyu Meng[4], Lei Zhang[5], *Fellow, IEEE*, David Zhang[5,6], *Fellow, IEEE*

[1]College of Computer Science, Nankai Univeristy, Tianjin, China

[2]Inception Institute of Artificial Intelligence, Abu Dhabi, UAE

[3]School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China

[4]School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, China

[5]Department of Computing, The Hong Kong Polytechnic University, Hong Kong SAR, China

[6]School of Science and Engineering, The Chinese University of Hong Kong (Shenzhen), Shenzhen, China

I. Solution of the NLSR Model

The NLSR model (Eqn. (20) in the main paper) does not have an analytical solution. We employ a variable splitting method [1], [2] to solve it. By introducing an auxiliary variable $\boldsymbol{Z}$, we can reformulate the NLSR model into a linear equality-constraint problem with two variables $\boldsymbol{C}$ and $\boldsymbol{Z}$:

$$\min_{\boldsymbol{C},\boldsymbol{Z}} \|\boldsymbol{X}-\boldsymbol{X}\boldsymbol{C}\|_F^2+\lambda\|\boldsymbol{C}\|_F^2 \quad \text{s.t.} \quad \boldsymbol{Z}=\boldsymbol{C}, \boldsymbol{Z}\geq 0. \tag{1}$$

Since the objective function is separable w.r.t. the variables $\boldsymbol{C}$ and $\boldsymbol{Z}$, problem (1) can be solved under the alternating direction method of multipliers (ADMM) [3] framework. The Lagrangian function of the problem (1) is

$$\begin{aligned}\mathcal{L}(\boldsymbol{C},\boldsymbol{Z},\boldsymbol{\Delta},\lambda,\rho)=&\|\boldsymbol{X}-\boldsymbol{X}\boldsymbol{C}\|_F^2+\lambda\|\boldsymbol{C}\|_F^2\\&+\langle\boldsymbol{\Delta},\boldsymbol{Z}-\boldsymbol{C}\rangle+\frac{\rho}{2}\|\boldsymbol{Z}-\boldsymbol{C}\|_F^2,\end{aligned} \tag{2}$$

where $\boldsymbol{\Delta}$ is the augmented Lagrangian multiplier and $\rho>0$ is the penalty parameter. We initialize the vector variables $\boldsymbol{C}_0$, $\boldsymbol{Z}_0$, and $\boldsymbol{\Delta}_0$ to be conformable zero matrices and set $\rho>0$ with a suitable value. Denote by ($\boldsymbol{C}_k$, $\boldsymbol{Z}_k$) and $\boldsymbol{\delta}_k$ the optimization variables and the Lagrange multiplier at iteration k ($k=0,1,2,...,K$), respectively. The variables can be updated by taking derivatives of the Lagrangian function (2) w.r.t. the variables $\boldsymbol{C}$ and $\boldsymbol{Z}$ and setting them to be zero.

(1) **Updating $\boldsymbol{C}$ while fixing $\boldsymbol{Z}$ and $\boldsymbol{\Delta}$:**

$$\min_{\boldsymbol{C}} \|\boldsymbol{X}-\boldsymbol{X}\boldsymbol{C}\|_F^2+\lambda\|\boldsymbol{C}\|_F^2+\frac{\rho}{2}\|\boldsymbol{C}-(\boldsymbol{Z}_k+\rho^{-1}\boldsymbol{\Delta}_k)\|_F^2. \tag{3}$$

This is a standard least squares regression problem with closed form solution:

$$\boldsymbol{C}_{k+1}=(\boldsymbol{X}^\top\boldsymbol{X}+\frac{2\lambda+\rho}{2}\boldsymbol{I})^{-1}(\boldsymbol{X}^\top\boldsymbol{X}+\frac{\rho}{2}\boldsymbol{Z}_k+\frac{1}{2}\boldsymbol{\Delta}_k) \tag{4}$$

(2) **Updating $\boldsymbol{Z}$ while fixing $\boldsymbol{C}$ and $\boldsymbol{\Delta}$:**

$$\min_{\boldsymbol{Z}} \|\boldsymbol{Z}-(\boldsymbol{C}_{k+1}-\rho^{-1}\boldsymbol{\Delta}_k)\|_F^2 \quad \text{s.t.} \quad \boldsymbol{Z}\geq 0. \tag{5}$$

Corresponding author: Jun Xu (email: nankaimathxujun@gmail.com).

The solution of $\boldsymbol{Z}$ is

$$\boldsymbol{Z}_{k+1}=\max(0,\boldsymbol{C}_{k+1}-\rho^{-1}\boldsymbol{\Delta}_k), \tag{6}$$

where the "$\max(\cdot)$" operator outputs element-wisely the maximal value of the inputs.

(3) **Updating the Lagrangian multiplier $\boldsymbol{\Delta}$:**

$$\boldsymbol{\Delta}_{k+1}=\boldsymbol{\Delta}_k+\rho(\boldsymbol{Z}_{k+1}-\boldsymbol{C}_{k+1}). \tag{7}$$

The above alternative updating steps are repeated until the convergence condition is satisfied or the number of iterations exceeds a preset threshold K. The convergence condition of the ADMM algorithm is: $\|\boldsymbol{Z}_{k+1}-\boldsymbol{C}_{k+1}\|_F\leq$ Tol, $\|\boldsymbol{C}_{k+1}-\boldsymbol{C}_k\|_F\leq$ Tol, and $\|\boldsymbol{Z}_{k+1}-\boldsymbol{Z}_k\|_F\leq$ Tol are simultaneously satisfied, where Tol >0 is a small tolerance value. Since the objective function and constraints are all strictly convex, the NLSR model solved by the ADMM algorithm [3] is guaranteed to converge to a global optimal solution.

II. Solution of the SLSR Model

We solve the SLSR model (Eqn. (21) in the main paper) by employing variable splitting methods [1], [2]. Specifically, we introduce an auxiliary variable $\boldsymbol{Z}$ into the SLSR model, which can then be equivalently reformulated as a linear equality-constrained problem:

$$\begin{aligned}&\min_{\boldsymbol{C},\boldsymbol{Z}}\|\boldsymbol{X}-\boldsymbol{X}\boldsymbol{C}\|_F^2+\lambda\|\boldsymbol{Z}\|_F^2\\&\text{s.t.}\quad \boldsymbol{1}^\top\boldsymbol{Z}=s\boldsymbol{1}^\top,\boldsymbol{Z}=\boldsymbol{C},\end{aligned} \tag{8}$$

whose solution for $\boldsymbol{C}$ coincides with the solution of Eqn. (20) in the main paper. Since its objective function is separable w.r.t. the variables $\boldsymbol{C}$ and $\boldsymbol{Z}$, problem (8) can also be solved via the ADMM method [3]. The corresponding augmented Lagrangian function is the same as in Eqn. (11) in the main paper. Denote by ($\boldsymbol{C}_k,\boldsymbol{Z}_k$) and $\boldsymbol{\Delta}_k$ the optimization variables and Lagrange multiplier at iteration k ($k=0,1,2,...$), respectively. We initialize the variables $\boldsymbol{C}_0$, $\boldsymbol{Z}_0$, and $\boldsymbol{\Delta}_0$ to be conformable zero matrices. By taking derivatives of the Lagrangian function $\mathcal{L}$ (Eqn. (11) in the main paper) w.r.t. $\boldsymbol{C}$ and $\boldsymbol{Z}$, and setting them to be zeros, we can alternatively update the variables as follows:

Algorithm 4: Projection of the vector $\boldsymbol{v}_{k+1}$ onto a scaled affine space

Input: Data point $\boldsymbol{v}_{k+1} \in \mathbb{R}^N$, scalar s.
1. Sort $\boldsymbol{v}_{k+1}$ into $\boldsymbol{w}$: $w_1 \geq w_2 \geq ... \geq w_N$;
2. Find $\alpha = \max\{1 \leq j \leq N : w_j + \frac{1}{j}(s - \sum_{i=1}^{j} w_i) > 0\}$;
3. Define $\beta = \frac{1}{\alpha}(s - \sum_{i=1}^{\alpha} w_i)$;

Output: $\boldsymbol{z}_{k+1}$: $z_{k+1}^i = v_{k+1}^i + \beta$, $i = 1, ..., N$.

(1) **Updating $\boldsymbol{C}$ while fixing $\boldsymbol{Z}_k$ and $\boldsymbol{\Delta}_k$:**

$$\boldsymbol{C}_{k+1} = \arg\min_{\boldsymbol{C}} \|\boldsymbol{X} - \boldsymbol{X}\boldsymbol{C}\|_F^2 + \frac{\rho}{2}\|\boldsymbol{C} - (\boldsymbol{Z}_k + \frac{1}{\rho}\boldsymbol{\Delta}_k)\|_F^2. \tag{9}$$

This is a standard least square regression problem and has a closed-from solution given by

$$\boldsymbol{C}_{k+1} = (\boldsymbol{X}^\top \boldsymbol{X} + \frac{\rho}{2}\boldsymbol{I})^{-1}(\boldsymbol{X}^\top \boldsymbol{X} + \frac{\rho}{2}\boldsymbol{Z}_k + \frac{1}{2}\boldsymbol{\Delta}_k). \tag{10}$$

(2) **Updating $\boldsymbol{Z}$ while fixing $\boldsymbol{C}_k$ and $\boldsymbol{\Delta}_k$:**

$$\begin{aligned} \boldsymbol{Z}_{k+1} = \arg\min_{\boldsymbol{Z}} & \|\boldsymbol{Z} - \frac{\rho}{2\lambda + \rho}(\boldsymbol{C}_{k+1} - \rho^{-1}\boldsymbol{\Delta}_k)\|_F^2 \\ \text{s.t.} \quad & \boldsymbol{1}^\top \boldsymbol{Z} = s\boldsymbol{1}^\top. \end{aligned} \tag{11}$$

This is a quadratic programming problem and the objective function is strictly convex, with a close and convex constraint, so there is a unique solution. Here, we employ the projection based method [4], whose computational complexity is $\mathcal{O}(N \log N)$ to process a vector of length N. Denote by $\boldsymbol{v}_{k+1}$ an arbitrary column of $\frac{\rho}{2\lambda+\rho}(\boldsymbol{C}_{k+1} - \rho^{-1}\Delta_k)$, the solution of $\boldsymbol{z}_{k+1}$ (the corresponding column in $\boldsymbol{Z}_{k+1}$) can be solved by projecting $\boldsymbol{v}_{k+1}$ onto a scaled affine space [4]. The solution of problem (11) is summarized in Algorithm 4.

(3) **Updating $\boldsymbol{\Delta}$ while fixing $\boldsymbol{C}_k$ and $\boldsymbol{Z}_k$:**

$$\boldsymbol{\Delta}_{k+1} = \boldsymbol{\Delta}_k + \rho(\boldsymbol{Z}_{k+1} - \boldsymbol{C}_{k+1}). \tag{12}$$

We repeat the above alternative updates until a certain convergence condition is satisfied or the number of iterations reaches a preset threshold K. The convergence condition of the ADMM algorithm is met when $\|\boldsymbol{C}_{k+1} - \boldsymbol{Z}_{k+1}\|_F \leq$ Tol, $\|\boldsymbol{C}_{k+1} - \boldsymbol{C}_k\|_F \leq$ Tol, and $\|\boldsymbol{Z}_{k+1} - \boldsymbol{Z}_k\|_F \leq$ Tol are simultaneously satisfied, where Tol > 0 is a small tolerance value. Since the objective function and constraints are convex, the SLSR model solved by the ADMM algorithm, is guaranteed to converge to a global optimal solution.

References

[1] R. Courant. Variational methods for the solution of problems of equilibrium and vibrations. *Bulletin of the American Mathematical Society*, 49(1):1–23, 1943. 1

[2] J. Eckstein and D. P. Bertsekas. On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1):293–318, 1992. 1

[3] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, January 2011. 1

[4] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra. Efficient projections onto the l 1-ball for learning in high dimensions. In *ICML*, pages 272–279. ACM, 2008. 2