Supplementary Material: Efficient Stochastic Optimization for Low-Rank Distance Metric Learning

Jie Zhang and **Lijun Zhang**
National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China
zhangj@lamda.nju.edu.cn, zhanglj@lamda.nju.edu.cn

Proof of Theorem 2

Note that $\Pi_{\mathbb{S}_+ \cap \{W | \|W\|_2 \leq \tau\}}[X]$ is the optimal solution to the following problem

$$\begin{aligned} \min_{A \in \mathbb{R}^{d\times d}} \quad & \frac{1}{2}\|A - X\|_F^2 \\ \text{s.t.} \quad & A \in \mathbb{S}_+, \ \tau I - A \in \mathbb{S}_+ \ . \end{aligned} \tag{6}$$

The Lagrangian function associated with (6) is

$$L(A, Y_1, Y_2) = \frac{||A - X||_F^2}{2} - \mathrm{tr}(AY_1) + \mathrm{tr}((A - \tau I)Y_2)$$

where $Y_1 \in \mathbb{R}^{d\times d}$ and $Y_2 \in \mathbb{R}^{d\times d}$ are dual variables for constraints $A \in \mathbb{S}_+$ and $\tau I - A \in \mathbb{S}_+$. Let A^*,Y_1^*,Y_2^* be the optimal primal and dual solutions. The KKT conditions are

$$\begin{aligned} A^* &\in \mathbb{S}_+, \\ \tau I - A^* &\in \mathbb{S}_+, \\ A^* &= X + Y_1^* - Y_2^*, \\ \mathrm{tr}(A^* Y_1^*) &= 0, \\ \mathrm{tr}((A^* - \tau I)Y_2^*) &= 0, \\ Y_1^* \in \mathbb{S}_+, \ Y_2^* &\in \mathbb{S}_+. \end{aligned}$$

We complete the proof by noticing that

$$\begin{aligned} A^* &= \sum_{i:\lambda_i > \tau} \tau \mathbf{u}_i \mathbf{u}_i^\top + \sum_{i: 0 < \lambda_i \leq \tau} \lambda_i \mathbf{u}_i \mathbf{u}_i^\top, \\ Y_1^* &= -\sum_{i:\lambda_i < 0} \lambda_i \mathbf{u}_i \mathbf{u}_i^\top, \\ Y_2^* &= \sum_{i:\lambda_i > \tau} (\lambda_i - \tau) \mathbf{u}_i \mathbf{u}_i^\top, \end{aligned}$$

satisfy these KKT conditions.

Proof of Theorem 3

Note that $\Pi_{\mathbb{S}_+ \cap \{W | \|W\|_F \leq \tau\}}[X]$ is the optimal solution to the following problem

$$\begin{aligned} \min_{A \in \mathbb{R}^{d\times d}} \quad & \frac{1}{2}\|A - X\|_F^2 \\ \text{s.t.} \quad & A \in \mathbb{S}_+, \ \|A\|_F \leq \tau \ . \end{aligned} \tag{7}$$

The Lagrangian function associated with (7) is

$$L(A, Y_1, Y_2) = \frac{||A - X||_F^2}{2} - \mathrm{tr}(AY_1) + \frac{\nu}{2}(\|A\|_F^2 - \tau^2)$$

where $Y_1 \in \mathbb{R}^{d\times d}$ and $\nu \in \mathbb{R}$ are dual variables for constraints $A \in \mathbb{S}_+$ and $\|A\|_F \leq \tau$. Let A^*,Y_1^*,ν^* be the optimal primal and dual solutions. The KKT conditions are

$$\begin{aligned} A^* &\in \mathbb{S}_+, \\ \|A^*\|_F &\leq \tau \\ A^* &= \frac{1}{1+\nu^*}(X + Y_1^*), \\ \mathrm{tr}(A^* Y_1^*) &= 0, \\ \nu^*(\|A^*\|_F^2 - \tau^2) &= 0, \\ Y_1^* \in \mathbb{S}_+,\ \nu^* &\geq 0. \end{aligned}$$

We complete the proof by noticing that

$$\begin{aligned} Y_1^* &= -\sum_{i:\lambda_i<0} \lambda_i \mathbf{u}_i \mathbf{u}_i^\top, \\ \widehat{A} &= X + Y_1^* = \sum_{i:\lambda_i>0} \lambda_i \mathbf{u}_i \mathbf{u}_i^\top, \\ \nu^* &= \begin{cases} 0 & \|\widehat{A}\|_F \leq \tau \\ \dfrac{\|\widehat{A}\|_F}{\tau} - 1 & \|\widehat{A}\|_F > \tau \end{cases}, \\ A^* &= \begin{cases} \widehat{A} & \|\widehat{A}\|_F \leq \tau \\ \dfrac{\tau}{\|\widehat{A}\|_F}\widehat{A} & \|\widehat{A}\|_F > \tau \end{cases}, \end{aligned}$$

satisfy these KKT conditions.