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 \( \Pi_{S_+ \cap \{W \mid \|W\|_2 \leq \tau \}} [X] \) is the optimal solution to the following problem
\[
\min_{A \in \mathbb{R}^{d \times d}} \frac{1}{2} \|A - X\|_F^2 \\
\text{s.t.} \quad A \in S_+, \tau I - A \in S_+ .
\]
(6)

The Lagrangian function associated with (6) is
\[
L(A, Y_1, Y_2) = \frac{\|A - X\|_F^2}{2} - \text{tr}(AY_1) + \text{tr}((A - \tau I)Y_2)
\]
where \( Y_1 \in \mathbb{R}^{d \times d} \) and \( Y_2 \in \mathbb{R}^{d \times d} \) are dual variables for constraints \( A \in S_+ \) and \( \tau I - A \in S_+ \). Let \( A^*, Y_1^*, Y_2^* \) be the optimal primal and dual solutions. The KKT conditions are
\[
A^* \in S_+, \tau I - A^* \in S_+, \\
A^* = X + Y_1^* - Y_2^*, \\
\text{tr}(A^*Y_1^*) = 0, \\
\text{tr}((A^* - \tau I)Y_2^*) = 0, \\
Y_1^* \in S_+, Y_2^* \in S_+ .
\]

We complete the proof by noticing that
\[
A^* = \sum_{i : \lambda_i > \tau} \tau u_i u_i^T + \sum_{i : 0 < \lambda_i \leq \tau} \lambda_i u_i u_i^T, \\
Y_1^* = -\sum_{i : \lambda_i < 0} \lambda_i u_i u_i^T, \\
Y_2^* = \sum_{i : \lambda_i > \tau} (\lambda_i - \tau) u_i u_i^T,
\]
satisfy these KKT conditions.

Proof of Theorem 3
Note that \( \Pi_{S_+ \cap \{W \mid \|W\|_F \leq \tau \}} [X] \) is the optimal solution to the following problem
\[
\min_{A \in \mathbb{R}^{d \times d}} \frac{1}{2} \|A - X\|_F^2 \\
\text{s.t.} \quad A \in S_+, \|A\|_F \leq \tau .
\]
(7)

The Lagrangian function associated with (7) is
\[
L(A, Y_1, Y_2) = \frac{\|A - X\|_F^2}{2} - \text{tr}(AY_1) + \nu \left( \frac{\|A\|_F^2}{2} - \tau^2 \right)
\]
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 $A^*, Y_1^*, \nu^*$ be the optimal primal and dual solutions. The KKT conditions are

$$A^* \in \mathbb{S}_+,$$

$$\|A^*\|_F \leq \tau$$

$$A^* = \frac{1}{1 + \nu^*}(X + Y_1^*),$$

$$\text{tr}(A^*Y_1^*) = 0,$$

$$\nu^*(\|A^*\|_F^2 - \tau^2) = 0,$$

$$Y_1^* \in \mathbb{S}_+, \ \nu^* \geq 0.$$

We complete the proof by noticing that

$$Y_1^* = - \sum_{i : \lambda_i < 0} \lambda_i u_i u_i^\top,$$

$$\hat{A} = X + Y_1^* = \sum_{i : \lambda_i > 0} \lambda_i u_i u_i^\top,$$

$$\nu^* = \begin{cases} 0 & \|\hat{A}\|_F \leq \tau \\ \frac{\|\hat{A}\|_F}{\tau} - 1 & \|\hat{A}\|_F > \tau \end{cases},$$

$$A^* = \begin{cases} \hat{A} & \|\hat{A}\|_F \leq \tau \\ \frac{\tau}{\|\hat{A}\|_F} \hat{A} & \|\hat{A}\|_F > \tau \end{cases},$$

satisfy these KKT conditions.