Supplemental Material of Regularized Distance Metric Learning: Theory and Algorithm

Rong Jin[1] **Shijun Wang**[2] **Yang Zhou**[1]
[1]Dept. of Computer Science & Engineering, Michigan State University, East Lansing, MI 48824
[2]Radiology and Imaging Sciences, National Institutes of Health, Bethesda, MD 20892
`rongjin@cse.msu.edu` `wangshi@cc.nih.gov` `zhouyang@msu.edu`

Theorem 1. *The optimal solution λ_t to the problem in (14) is expressed as*

$$\lambda_t = \begin{cases} \lambda & y_t = -1 \\ \min\left(\lambda, [(x_t - x_t')^\top A_{t-1}^{-1}(x_t - x_t')]^{-1}\right) & y_t = +1 \end{cases}$$

Proof. By using the Schur complement, this condition $A_{t-1} - \lambda y_t (x_t - x_t')(x_t - x_t')^\top \succeq 0$ can be expressed as

$$\begin{pmatrix} A_{t-1} & x - x' \\ (x - x')^T & 1/(\lambda y_t) \end{pmatrix} \succeq 0.$$

If we apply the Schur complemnet again on the above matrix, then it is equivalent to

$$1/(\lambda y_t) \geq (x - x')^T A_{t-1}^{-1}(x - x').$$

When $y_t = +1$, we have

$$\lambda \leq \frac{1}{(x - x')^T A_{t-1}^{-1}(x - x')}.$$

In order to keep an appropriate and stable learning speed, we prefer a pre-selected λ by imposing the above limitation through $\min\left(\lambda, [(x_t - x_t')^\top A_{t-1}^{-1}(x_t - x_t')]^{-1}\right)$.

When $y_t = -1$, because the summation of two positive definite matrices is still a positive definite matrix, we simply choose $\lambda_t = \lambda$.

□