Efficient Learning of Mahalanobis Metrics for Ranking

1. Derivation of Algorithm 2

We first reproduce the relevant lemmas from the main paper:

Lemma 1.1 *Given a point $W = YY^\top \in S^+_{d,m}$, the orthogonal projection of a matrix Z in the ambient space $\mathbb{R}^{d\times d}$ onto $\mathcal{T}_W S^+_{d,m}$, is given by $P_{T_W}(Z) = \xi$, where*

$$\xi = \xi^s + \xi^p; \quad \xi^s = P_y \frac{Z+Z^\top}{2} P_y, \tag{1}$$

$$\xi^p = P_y^\perp \frac{Z+Z^\top}{2} P_y + P_y \frac{Z+Z^\top}{2} P_y^\perp$$

and $P_y = YY^\dagger$, $P_y^\perp = I - P_y$

Lemma 1.2 *Let $W \in S^+_{d,m}$ and ξ, ξ^p, ξ^s be as defined in Lemma 1.1. Then, the function $\mathcal{R}_W(\xi) = VW^\dagger V$ where*

$$V = W + \frac{1}{2}\xi^s + \xi^p - \frac{1}{8}\xi^s W^\dagger \xi^s - \frac{1}{2}\xi^p W^\dagger \xi^s$$

is a second-order retraction from the tangent space $\mathcal{T}_W S^+_{d,m}$ to $S^+_{d,m}$.

1.1. Useful Identities

If we assume Y is full column rank, then we have the identity

$$W^\dagger = (Y^\dagger)^\top Y^\dagger$$

We also introduce the identities

$$P_y W^\dagger = W^\dagger P_y = W^\dagger$$
$$Y^\dagger P_y = Y^\dagger$$
$$P_y (Y^\dagger)^\top = (Y^\dagger)^\top$$

for simplifying expressions in the sequel.

1.2. Derivation

Given the current estimate $W_t = YY^\top$ and the gradient $G = UV^\top$, we need to calculate M such that $MM^\top = W_{t+1}$.

From Lemma 1.2,

$$\begin{aligned} W_{t+1} &= VW_t^\dagger V^\top \\ &= V(Y^\dagger)^\top Y^\dagger V^\top \\ &= MM^\top, M = V(Y^\dagger)^\top \end{aligned}$$

Thus, we just need to compute $V(Y^\dagger)^\top$ to obtain W_{t+1}:

$$\begin{aligned} M &= V(Y^\dagger)^\top \\ &= W(Y^\dagger)^\top + \frac{1}{2}\xi^s (Y^\dagger)^\top + \xi^p (Y^\dagger)^\top \\ &\quad - \frac{1}{8}\xi^s W^\dagger \xi^s (Y^\dagger)^\top - \frac{1}{2}\xi^p W^\dagger \xi^s (Y^\dagger)^\top \end{aligned} \tag{2}$$

Now, using the expressions for ξ^s, ξ^p, $P_y^\perp$ and P_y in Lemma 1.1, and substituting G for $\frac{Z+Z^\top}{2}$ (assuming G is symmetric) we can derive the following:

$$\begin{aligned} \xi^p &= (I - P_y) G P_y + P_y G (I - P_y) \\ &= P_y G + G P_y - 2 P_y G P_y \\ W^\dagger \xi^s (Y^\dagger)^\top &= W^\dagger P_y G P_y (Y^\dagger)^\top \\ &= W^\dagger G (Y^\dagger)^\top \end{aligned}$$

Now we can calculate each term in (2):

$$\begin{aligned} W(Y^\dagger)^\top &= YY^\top (Y^\dagger)^\top = Y \\ \xi^s (Y^\dagger)^\top &= P_y G P_y (Y^\dagger)^\top = P_y G (Y^\dagger)^\top \\ \xi^p (Y^\dagger)^\top &= P_y G (Y^\dagger)^\top + G P_y (Y^\dagger)^\top - 2 P_y G P_y (Y^\dagger)^\top \\ &= P_y G (Y^\dagger)^\top + G (Y^\dagger)^\top - 2 P_y G (Y^\dagger)^\top \\ &= G (Y^\dagger)^\top - P_y G (Y^\dagger)^\top \\ \xi^s W^\dagger \xi^s (Y^\dagger)^\top &= P_y G P_y W^\dagger G (Y^\dagger)^\top = P_y G W^\dagger G (Y^\dagger)^\top \\ \xi^p W^\dagger \xi^s (Y^\dagger)^\top &= (P_y G + G P_y - 2 P_y G P_y) W^\dagger G (Y^\dagger)^\top \\ &= P_y G W^\dagger G (Y^\dagger)^\top + G P_y W^\dagger G (Y^\dagger)^\top \\ &\quad - 2 P_y G P_y W^\dagger G (Y^\dagger)^\top \\ &= P_y G W^\dagger G (Y^\dagger)^\top + G W^\dagger G (Y^\dagger)^\top \\ &\quad - 2 P_y G W^\dagger G (Y^\dagger)^\top \\ &= G W^\dagger G (Y^\dagger)^\top - P_y G W^\dagger G (Y^\dagger)^\top \end{aligned}$$

Substituting into (2), we get:

$$\begin{aligned} M &= Y + G(Y^\dagger)^\top - \frac{1}{2} P_y G (Y^\dagger)^\top \\ &\quad - \frac{1}{2} G W^\dagger G (Y^\dagger)^\top + \frac{3}{8} P_y G W^\dagger G (Y^\dagger)^\top \end{aligned}$$

By substituting $G = UV^\top$ and choosing the order of multiplication appropriately, Algorithm 2 naturally follows.