Supplemental Material for A-Optimal Projection for Image Representation

Xiaofei He, *Senior Member, IEEE*, Chiyuan Zhang
Lijun Zhang, *Member, IEEE*, and Xuelong Li, *Fellow, IEEE*

1 Proof of Theorem 3.1

By using Woodbury matrix identity, Eq. (14) can be rewritten as follows:

$$\begin{aligned}&\mathrm{Tr}\Big(\big(A^T\widetilde{X}\widetilde{X}^TA+\lambda_2 I\big)^{-1}\Big)\\ =&\ \mathrm{Tr}\Big(\frac{1}{\lambda_2}I-\frac{1}{\lambda_2}A^T\widetilde{X}\big(I+\frac{1}{\lambda_2}\widetilde{X}^TAA^T\widetilde{X}\big)^{-1}\widetilde{X}^TA\Big)\\ =&\ \frac{k}{\lambda_2}-\frac{1}{\lambda_2}\mathrm{Tr}\Big(A^T\widetilde{X}\big(\lambda_2 I+\widetilde{X}^TAA^T\widetilde{X}\big)^{-1}\widetilde{X}^TA\Big)\end{aligned}$$

Noticing that $\mathrm{Tr}(AB)=\mathrm{Tr}(BA)$, we have:

$$\begin{aligned}&\mathrm{Tr}\Big(\big(A^T\widetilde{X}\widetilde{X}^TA+\lambda_2 I\big)^{-1}\Big)\\ =&\ \frac{k}{\lambda_2}-\frac{1}{\lambda_2}\mathrm{Tr}\Big(\big(\lambda_2 I+\widetilde{X}^TAA^T\widetilde{X}\big)^{-1}\widetilde{X}^TAA^T\widetilde{X}\Big)\\ =&\ \frac{k}{\lambda_2}-\frac{1}{\lambda_2}\mathrm{Tr}\Big(\big(\lambda_2 I+\widetilde{X}^TAA^T\widetilde{X}\big)^{-1}\\ &\qquad\big(\widetilde{X}^TAA^T\widetilde{X}+\lambda_2 I-\lambda_2 I\big)\Big)\\ =&\ \frac{k}{\lambda_2}-\frac{1}{\lambda_2}\mathrm{Tr}\Big(I-\lambda_2\big(\lambda_2 I+\widetilde{X}^TAA^T\widetilde{X}\big)^{-1}\Big)\\ =&\ \frac{k-n}{\lambda_2}+\mathrm{Tr}\Big(\big(\lambda_2 I+\widetilde{X}^TAA^T\widetilde{X}\big)^{-1}\Big)\end{aligned}$$

This completes the proof.

2 Proof of Theorem 4.1

Let $H=g(M)=\lambda I+\widetilde{X}^TM\widetilde{X}$. It is easy to see that g is an affine function of M. Therefore, g is convex. Let $f(H)=\mathrm{Tr}(H^{-1})$ which is also convex [1]. Thus, $f\circ g(M)=\mathrm{Tr}\Big(\big(\lambda I+\widetilde{X}^TM\widetilde{X}\big)^{-1}\Big)$ is convex. This completes the proof.

- *X. He and C. Zhang are with the State Key Lab of CAD&CG, College of Computer Science, Zhejiang University, Hangzhou 310058, China. E-mail: xiaofeihe@cad.zju.edu.cn, pluskid@gmail.com*
- *L. Zhang is with National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China. E-mail: zhanglj@lamda.nju.edu.cn*
- *X. Li is with the Center for OPTical IMagery Analysis and Learning (OPTIMAL), State key Laboratory of Transicent Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China. E-mail: xuelong_li@opt.ac.cn.*

3 Proof of Theorem 4.2

Let M_a^* be the optimal solution of problem (17), and (P^*, M_b^*) be the optimal solutions of problem (18). Then $M_a^*=M_b^*$ is a sufficient condition for Theorem 4.2. We define

$$f(M)=(\lambda I+\widetilde{X}^TM\widetilde{X})^{-1}.$$

Assume $M_a^*\neq M_b^*$. Since M_a^* minimizes the problem (17), we have

$$\mathrm{Tr}f(M_a^*)<\mathrm{Tr}f(M_b^*).$$

Note that (P^*, M_b^*) satisfies the conditions in problem (18), so we have

$$\begin{aligned}&P^*\succeq_{\mathbb{S}_n^+} f(M_b^*)\\ \Leftrightarrow\ &P^*-f(M_b^*)\in\mathbb{S}_n^+\\ \Leftrightarrow\ &\mathrm{Tr}P^*>\mathrm{Tr}f(M_b^*)\end{aligned}$$

It is clear that $(f(M_b^*), M_a^*)$ satisfies the conditions in problem (18). Thus, for problem (18), $(f(M_b^*), M_a^*)$ is more optimal than P^*, M_b^*, which contradicts our assumption. Therefore, we have $M_a^*=M_b^*$.

4 Proof of Theorem 4.3

Let $\phi=\|I-A^T\widetilde{X}B\|^2+\lambda\|B\|^2$. Thus, we have

$$\begin{aligned}\phi&=\|I-A^T\widetilde{X}B\|^2+\lambda\|B\|^2\\ &=\mathrm{Tr}\Big((I-A^T\widetilde{X}B)(I-A^T\widetilde{X}B)^T\Big)+\lambda\mathrm{Tr}(BB^T)\\ &=k-2\mathrm{Tr}(A^T\widetilde{X}B)+\mathrm{Tr}(A^T\widetilde{X}BB^T\widetilde{X}^TA)+\lambda\mathrm{Tr}(BB^T)\end{aligned}\tag{1}$$

Noticing that

$$\frac{\partial\mathrm{Tr}(A^T\widetilde{X}B)}{\partial B^T}=A^T\widetilde{X},$$

$$\frac{\partial\mathrm{Tr}(A^T\widetilde{X}BB^T\widetilde{X}^TA)}{\partial B^T}=2B^T\widetilde{X}^TAA^T\widetilde{X},$$

and

$$\frac{\partial\mathrm{Tr}(BB^T)}{\partial B^T}=2B^T.$$

By requiring the gradient of ϕ with respect to B^T vanish, we have

$$
\begin{aligned}
&\frac{\partial\phi}{\partial B^T} = 0 \\
\Rightarrow\ & B^T\widetilde{X}^T AA^T\widetilde{X} + B^T - A^T\widetilde{X} = 0 \\
\Rightarrow\ & B = (\widetilde{X}^T AA^T\widetilde{X} + \lambda I)^{-1}\widetilde{X}^T A.
\end{aligned} \tag{2}
$$

Substituting Eq. (2) into Eq. (1) and noticing that $\text{Tr}(AB) = \text{Tr}(BA)$, we have

$$
\begin{aligned}
&\text{Tr}(A^T\widetilde{X}B) \\
=\ & \text{Tr}\Big(A^T\widetilde{X}\big(\widetilde{X}^T AA^T\widetilde{X} + \lambda I\big)^{-1}\widetilde{X}^T A\Big) \\
=\ & \text{Tr}\Big(\big(\widetilde{X}^T AA^T\widetilde{X} + \lambda I\big)^{-1}\widetilde{X}^T AA^T\widetilde{X}\Big) \\
=\ & \text{Tr}\Big(\big(\widetilde{X}^T AA^T\widetilde{X} + \lambda I\big)^{-1}\big(\widetilde{X}^T AA^T\widetilde{X} + \lambda I - \lambda I\big)\Big) \\
=\ & n - \lambda\text{Tr}\Big(\big(\widetilde{X}^T AA^T\widetilde{X} + \lambda I\big)^{-1}\Big),
\end{aligned} \tag{3}
$$

and,

$$
\begin{aligned}
&\text{Tr}(A^T\widetilde{X}BB^T\widetilde{X}^T A) + \lambda\text{Tr}(BB^T) \\
=\ & \text{Tr}(BB^T\widetilde{X}^T AA^T\widetilde{X}) + \lambda\text{Tr}(BB^T) \\
=\ & \text{Tr}\Big(BB^T\big(\widetilde{X}^T AA^T\widetilde{X} + \lambda I\big)\Big) \\
=\ & \text{Tr}\Big(\big(\widetilde{X}^T AA^T\widetilde{X} + \lambda I\big)^{-1}\widetilde{X}^T AA^T\widetilde{X}\big(\widetilde{X}^T AA^T\widetilde{X} + \lambda I\big)^{-1} \\
&\quad \big(\widetilde{X}^T AA^T\widetilde{X} + \lambda I\big)\Big) \\
=\ & \text{Tr}\Big(\big(\widetilde{X}^T AA^T\widetilde{X} + \lambda I\big)^{-1}\widetilde{X}^T AA^T\widetilde{X}\Big) \\
=\ & n - \lambda\text{Tr}\Big(\big(\widetilde{X}^T AA^T\widetilde{X} + \lambda I\big)^{-1}\Big).
\end{aligned} \tag{4}
$$

Finally, we have

$$
\phi = k - n + \lambda\text{Tr}\Big(\big(\widetilde{X}^T AA^T\widetilde{X} + \lambda I\big)^{-1}\Big). \tag{5}
$$

Thus, the optimal A is given by solving the following problem:

$$
\min_A \text{Tr}\Big(\big(\widetilde{X}^T AA^T\widetilde{X} + \lambda I\big)^{-1}\Big). \tag{6}
$$

This completes the proof.

REFERENCES

[1] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.