Supplementary Material of "Learning Canonical $\mathcal{F}$-Correlation Projection for Compact Multiview Representation"

Yun-Hao Yuan[1], Jin Li[1], Yun Li[1], Jipeng Qiang[1], Yi Zhu[1], Xiaobo Shen[2], Jianping Gou[3]
[1]School of Information Engineering, Yangzhou University, Yangzhou, China
[2]School of Computer Science, Nanjing University of Science and Technology, Nanjing, China
[3]School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang, China
{yhyuan, liyun, jpqiang, zhuyi}@yzu.edu.cn
cvjinli@outlook.com, njust.shenxiaobo@gmail.com, goujianping@ujs.edu.cn

Appendix A. Proof of Theorem 1

Before demonstrating the Theorem 1, let us first provide the following theorem:

Theorem 4 [1]. Suppose $\psi(z)$ is continuous on $z \geq 0$ and positive on $z > 0$ for $z \in \mathbb{R}$, $\mathrm{d}\psi/\mathrm{d}z$ is completely monotonic but not constant w.r.t. $z > 0$. Then for a set of any distinct vectors $\{\mathbf{f}^i \in \mathbb{R}^n\}_{i=1}^q$,

$$(-1)^{q-1}\det(\mathbf{H}) > 0 \tag{A1}$$

holds for any n and q, where $\det(\cdot)$ denotes the determinant of a matrix, $\mathbf{H} \in \mathbb{R}^{q\times q}$ with $(i,\ j)$-th entry as $\psi\left(||\mathbf{f}^i - \mathbf{f}^j||^2\right)$ and $||\cdot||$ denotes the 2-norm of a vector.

Theorem 4 reveals that the matrix $\mathbf{H}$, generated by some function $\psi(\cdot)$, must be nonsingular due to its nonzero determinant. Using Theorem 4, we can prove Theorem 1 as follows:

Proof of Theorem 1. In CCP, the $(j,\ t)$-th entry of $\mathbf{K}_{ii}^{\mathcal{F}}$ is calculated by

$$\ker(\mathbf{f}_i^j, \mathbf{f}_i^t) = \exp\left(\frac{-||\mathbf{f}_i^j - \mathbf{f}_i^t||^2}{2\sigma^2}\right), \tag{A2}$$

where $i = 1, 2$ and $j, t = 1, 2, \cdots, d_i$. Let $\psi(z)$ be

$$\psi(z) = \exp\left(\frac{-z}{2\sigma^2}\right),\ z \in \mathbb{R}. \tag{A3}$$

It follows from (A2) and (A3) that

$$\mathbf{K}_{ii}^{\mathcal{F}}(j,t) = \ker\left(\mathbf{f}_i^j, \mathbf{f}_i^t\right) = \psi\left(||\mathbf{f}_i^j - \mathbf{f}_i^t||^2\right). \tag{A4}$$

In addition, it is easy to show that $\psi(z)$ is continuous on $z \geq 0$ and positive on $z > 0$. Also, the derivative of $\psi(z)$ is

$$\frac{\mathrm{d}\psi}{\mathrm{d}z} = -\frac{1}{2\sigma^2}\exp\left(\frac{-z}{2\sigma^2}\right), \tag{A5}$$

which is strictly increasing on $z > 0$. Hence, $\det(\mathbf{K}_{ii}^{\mathcal{F}}) \neq 0$ holds according to Theorem 4. It follows immediately that $\mathbf{K}_{11}^{\mathcal{F}}$ and $\mathbf{K}_{22}^{\mathcal{F}}$ are nonsingular. ■

Appendix B. Proof of Theorem 2

Proof. Let the i-th singular value of $\tilde{\mathbf{W}}_1\tilde{\mathbf{W}}_2^T \in \mathbb{R}^{d_1\times d_2}$ be η_i, $i = 1, 2, \cdots, \min(d_1, d_2)$. Since $\tilde{\mathbf{W}}_1^T\tilde{\mathbf{W}}_1 = \mathbf{I}_d$ and $\tilde{\mathbf{W}}_2^T\tilde{\mathbf{W}}_2 = \mathbf{I}_d$, it is easy to show that $\tilde{\mathbf{W}}_1\tilde{\mathbf{W}}_2^T$ has d unit singular values and the rest are zero, i.e., $\eta_1 = \eta_2 = \cdots = \eta_d = 1$. Using the Von-Neumann's trace inequality [2], we have

$$\begin{aligned}\mathrm{Tr}\left(\tilde{\mathbf{W}}_1^T\tilde{\mathbf{K}}_{12}^{\mathcal{F}}\tilde{\mathbf{W}}_2\right) &= \mathrm{Tr}\left((\tilde{\mathbf{W}}_1\tilde{\mathbf{W}}_2^T)^T\tilde{\mathbf{K}}_{12}^{\mathcal{F}}\right)\\ &\leq \sum_{i=1}^r \eta_i\sigma_i = \sum_{i=1}^d \sigma_i,\end{aligned} \tag{B1}$$

where the first equality applies the matrix property $\mathrm{Tr}(\mathbf{AB}) = \mathrm{Tr}(\mathbf{BA})$.

This means that the maximum value of optimization problem in (12) is the sum of the top d singular values of $\tilde{\mathbf{K}}_{12}^{\mathcal{F}}$. Using (13), we have

$$\begin{aligned}&\mathrm{Tr}\left(\tilde{\mathbf{W}}_1^T\tilde{\mathbf{K}}_{12}^{\mathcal{F}}\tilde{\mathbf{W}}_2\right)\\ &= \mathrm{Tr}\left((\mathbf{U}(:,1:d))^T\,\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T\,(\mathbf{V}(:,1:d))\right)\\ &= \mathrm{Tr}\left(\begin{bmatrix}\mathbf{I}_d & \mathbf{0}\end{bmatrix}\begin{bmatrix}\boldsymbol{\Sigma}_d & \\ & \boldsymbol{\Sigma}_{r-d}\end{bmatrix}\begin{bmatrix}\mathbf{I}_d\\ \mathbf{0}\end{bmatrix}\right)\\ &= \mathrm{Tr}(\boldsymbol{\Sigma}_d) = \sum_{i=1}^d \sigma_i,\end{aligned} \tag{B2}$$

where $\boldsymbol{\Sigma}_d = \mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_d)$ and $\boldsymbol{\Sigma}_{r-d} = \mathrm{diag}(\sigma_{d+1}, \cdots, \sigma_r)$. Hence, $\tilde{\mathbf{W}}_1 = \mathbf{U}(:, 1:d)$ and $\tilde{\mathbf{W}}_2 = \mathbf{V}(:, 1:d)$ are a solution of optimization problem in (12). ■

Appendix C. Derivation of Updating Rules in (18) and (19)

For optimization problem in (17), using the Lagrange multiplier technique we can obtain the following

$$\mathcal{L} = \sum_{i=1}^{m}\sum_{j=1}^{m} \tilde{\mathbf{w}}_i^T \mathbf{P}_i^T \tilde{\mathbf{K}}_{ij}^{\mathcal{F}} \mathbf{P}_j \tilde{\mathbf{w}}_j - \sum_{i=1}^{m} \lambda_i (\tilde{\mathbf{w}}_i^T \tilde{\mathbf{w}}_i - 1), \quad \text{(C1)}$$

where $\{\lambda_i \in \mathbb{R}\}_{i=1}^m$ are the Lagrange multipliers. Taking the derivative of $\mathcal{L}$ w.r.t. $\tilde{\mathbf{w}}_i$ and setting it to $\mathbf{0}$, we obtain

$$\frac{\partial \mathcal{L}}{\partial \tilde{\mathbf{w}}_i} = 2\sum_{j=1}^{m} \mathbf{P}_i^T \tilde{\mathbf{K}}_{ij}^{\mathcal{F}} \mathbf{P}_j \tilde{\mathbf{w}}_j - 2\lambda_i \tilde{\mathbf{w}}_i = \mathbf{0} \quad \text{(C2)}$$

with $i = 1, 2, \cdots, m$. It follows from (C2) that

$$\sum_{j=1}^{m} \mathbf{P}_i^T \tilde{\mathbf{K}}_{ij}^{\mathcal{F}} \mathbf{P}_j \tilde{\mathbf{w}}_j = \lambda_i \tilde{\mathbf{w}}_i, \ i = 1, 2, \cdots, m. \quad \text{(C3)}$$

According to (C3) and $\tilde{\mathbf{w}}_i^T \tilde{\mathbf{w}}_i = 1$, we can obtain the following updating rules:

$$\lambda_i \leftarrow \left\| \sum_{j=1}^{m} \mathbf{P}_i^T \tilde{\mathbf{K}}_{ij}^{\mathcal{F}} \mathbf{P}_j \tilde{\mathbf{w}}_j \right\|,$$
$$\tilde{\mathbf{w}}_i \leftarrow \frac{1}{\lambda_i} \sum_{j=1}^{m} \mathbf{P}_i^T \tilde{\mathbf{K}}_{ij}^{\mathcal{F}} \mathbf{P}_j \tilde{\mathbf{w}}_j.$$

Appendix D. Proof of Theorem 3

We can rewrite (C3) as the following concise form:

$$\tilde{\mathbf{K}}\tilde{\mathbf{w}} = \mathbf{\Lambda}\tilde{\mathbf{w}}, \quad \text{(D1)}$$

where $\tilde{\mathbf{K}} \in \mathbb{R}^{h\times h}$ is a block matrix with the (i, j)-th block element as $\mathbf{P}_i^T \tilde{\mathbf{K}}_{ij}^{\mathcal{F}} \mathbf{P}_j$, $\tilde{\mathbf{w}} = \left[\tilde{\mathbf{w}}_1^T, \tilde{\mathbf{w}}_2^T, \cdots, \tilde{\mathbf{w}}_m^T\right]^T \in \mathbb{R}^h$, $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1 \mathbf{I}_{d_1}, \lambda_2 \mathbf{I}_{d_2}, \cdots, \lambda_m \mathbf{I}_{d_m}) \in \mathbb{R}^{h\times h}$, and $h = \sum_{i=1}^m d_i$.

Next, let us first provide the following two important lemmas, which play the key roles to complete the proof of Theorem 3.

Lemma 1. The matrix $\tilde{\mathbf{K}}$ in (D1) is symmetric positive semi-definite.

Proof. Let $\phi(\mathbf{X}_i) = \left[\phi(\mathbf{f}_i^1), \phi(\mathbf{f}_i^2), \cdots, \phi(\mathbf{f}_i^{d_i})\right]^T \in \mathbb{R}^{d_i \times N}$, $i = 1, 2, \cdots, m$. Let us denote

$$\mathbf{P} = \operatorname{diag}(\mathbf{P}_1, \mathbf{P}_2, \cdots, \mathbf{P}_m) \in \mathbb{R}^{h\times h},$$
$$\mathbf{D} = \operatorname{diag}\left((\mathbf{K}_{11}^{\mathcal{F}})^{-\frac{1}{2}}, (\mathbf{K}_{22}^{\mathcal{F}})^{-\frac{1}{2}}, \cdots, (\mathbf{K}_{mm}^{\mathcal{F}})^{-\frac{1}{2}}\right) \in \mathbb{R}^{h\times h},$$
$$\phi(\mathbf{X}) = \left[\phi(\mathbf{X}_1)^T, \phi(\mathbf{X}_2)^T, \cdots, \phi(\mathbf{X}_m)^T\right]^T \in \mathbb{R}^{h\times N}$$

with $\mathbf{X} = \left[\mathbf{X}_1^T, \mathbf{X}_2^T, \cdots, \mathbf{X}_m^T\right]^T \in \mathbb{R}^{h\times n}$. Note that $\mathbf{D}^T = \mathbf{D}$ due to the symmetry of each $(\mathbf{K}_{ii}^{\mathcal{F}})^{-\frac{1}{2}}$. Then, we have that

$$\tilde{\mathbf{K}} = \mathbf{P}^T \mathbf{D} \phi(\mathbf{X}) \phi(\mathbf{X})^T \mathbf{D} \mathbf{P}. \quad \text{(D2)}$$

Clearly $\tilde{\mathbf{K}}$ is symmetric. For an arbitrary nonzero vector $\boldsymbol{\xi} \in \mathbb{R}^h$, it follows from (D2) that

$$\begin{aligned}\boldsymbol{\xi}^T \tilde{\mathbf{K}} \boldsymbol{\xi} &= \boldsymbol{\xi}^T \mathbf{P}^T \mathbf{D} \phi(\mathbf{X}) \phi(\mathbf{X})^T \mathbf{D} \mathbf{P} \boldsymbol{\xi} \\ &= \left(\phi(\mathbf{X})^T \mathbf{D} \mathbf{P} \boldsymbol{\xi}\right)^T \left(\phi(\mathbf{X})^T \mathbf{D} \mathbf{P} \boldsymbol{\xi}\right) \\ &\geq 0.\end{aligned} \quad \text{(D3)}$$

Hence, $\tilde{\mathbf{K}}$ is symmetric positive semi-definite. ■

Lemma 2. Let the largest eigenvalue of $\tilde{\mathbf{K}}$ be δ_1 and $\tilde{\mathbf{w}}_i^T \tilde{\mathbf{w}}_i = 1, i = 1, 2, \cdots, m$. Then, $\tilde{\mathbf{w}}^T \tilde{\mathbf{K}} \tilde{\mathbf{w}} \leq m\delta_1$, where $\tilde{\mathbf{w}}$ is defined in (D1).

Proof. Let the eigenvalue decomposition of $\tilde{\mathbf{K}}$ be

$$\tilde{\mathbf{K}} = \mathbf{G}\mathbf{\Delta}\mathbf{G}^T = \sum_{i=1}^{h} \delta_i \mathbf{g}_i \mathbf{g}_i^T, \quad \text{(D4)}$$

where $\mathbf{G} = [\mathbf{g}_1, \mathbf{g}_2, \cdots, \mathbf{g}_h] \in \mathbb{R}^{h\times h}$ is an orthogonal matrix, and $\mathbf{\Delta} = \operatorname{diag}(\delta_1, \delta_2, \cdots, \delta_h) \in \mathbb{R}^{h\times h}$ is a diagonal matrix consisting of h nonnegative eigenvalues in descending order.

Using (D4), we have

$$\begin{aligned}\tilde{\mathbf{w}}^T \tilde{\mathbf{K}} \tilde{\mathbf{w}} &= \sum_{i=1}^{h} \delta_i \tilde{\mathbf{w}}^T \mathbf{g}_i \mathbf{g}_i^T \tilde{\mathbf{w}} = \sum_{i=1}^{h} \delta_i (\tilde{\mathbf{w}}^T \mathbf{g}_i)^2 \\ &\leq \delta_1 \sum_{i=1}^{h} (\tilde{\mathbf{w}}^T \mathbf{g}_i)^2 = \delta_1 \tilde{\mathbf{w}}^T \left(\sum_{i=1}^{h} \mathbf{g}_i \mathbf{g}_i^T\right) \tilde{\mathbf{w}} \\ &= \delta_1 \tilde{\mathbf{w}}^T \mathbf{G}\mathbf{G}^T \tilde{\mathbf{w}} = \delta_1 \sum_{i=1}^{m} \tilde{\mathbf{w}}_i^T \tilde{\mathbf{w}}_i = m\delta_1.\end{aligned} \quad \text{(D5)}$$

Thus, Lemma 2 is true. ■

Using Lemmas 1 and 2, now let us prove the Theorem 3:

Proof of Theorem 3. Using (D1), we can rewrite the updating rules in (18) and (19) as

$$\tilde{\mathbf{K}}\tilde{\mathbf{w}}^{(t)} = \mathbf{\Lambda}^{(t)} \tilde{\mathbf{w}}^{(t+1)}, \quad \text{(D6)}$$

where $\tilde{\mathbf{w}}^{(t)} = \left[\tilde{\mathbf{w}}_1^{(t)T}, \tilde{\mathbf{w}}_2^{(t)T}, \cdots, \tilde{\mathbf{w}}_m^{(t)T}\right]^T$ and $\mathbf{\Lambda}^{(t)} = \operatorname{diag}\left(\lambda_1^{(t)} \mathbf{I}_{d_1}, \lambda_2^{(t)} \mathbf{I}_{d_2}, \cdots, \lambda_m^{(t)} \mathbf{I}_{d_m}\right)$, and t is an iterative variable.

For optimization problem in (17), let us define

$$f(\tilde{\mathbf{w}}) = \sum_{i=1}^{m}\sum_{j=1}^{m} \tilde{\mathbf{w}}_i^T \mathbf{P}_i^T \tilde{\mathbf{K}}_{ij}^{\mathcal{F}} \mathbf{P}_j \tilde{\mathbf{w}}_j = \tilde{\mathbf{w}}^T \tilde{\mathbf{K}} \tilde{\mathbf{w}}. \quad \text{(D7)}$$

It follows from (D7) that

$$f(\tilde{\mathbf{w}}^{(t)}) = \tilde{\mathbf{w}}^{(t)T}\tilde{\mathbf{K}}\tilde{\mathbf{w}}^{(t)}, \tag{D8}$$

$$f(\tilde{\mathbf{w}}^{(t+1)}) = \tilde{\mathbf{w}}^{(t+1)T}\tilde{\mathbf{K}}\tilde{\mathbf{w}}^{(t+1)}. \tag{D9}$$

Using (D6), we are able to obtain the following

$$\begin{aligned} & f(\tilde{\mathbf{w}}^{(t+1)}) - f(\tilde{\mathbf{w}}^{(t)}) \\ & \quad = \tilde{\mathbf{w}}^{(t+1)T}\tilde{\mathbf{K}}\tilde{\mathbf{w}}^{(t+1)} - \tilde{\mathbf{w}}^{(t)T}\tilde{\mathbf{K}}\tilde{\mathbf{w}}^{(t)} \\ & \quad = \tilde{\mathbf{w}}^{(t+1)T}\tilde{\mathbf{K}}\tilde{\mathbf{w}}^{(t+1)} - \tilde{\mathbf{w}}^{(t+1)T}\boldsymbol{\Lambda}^{(t)}\tilde{\mathbf{w}}^{(t)} \\ & \qquad + \tilde{\mathbf{w}}^{(t+1)T}\left(\boldsymbol{\Lambda}^{(t)}\tilde{\mathbf{w}}^{(t+1)} - \tilde{\mathbf{K}}\tilde{\mathbf{w}}^{(t)}\right) \\ & \quad = \tilde{\mathbf{w}}^{(t+1)T}\tilde{\mathbf{K}}\left(\tilde{\mathbf{w}}^{(t+1)} - \tilde{\mathbf{w}}^{(t)}\right) \\ & \qquad + \tilde{\mathbf{w}}^{(t+1)T}\boldsymbol{\Lambda}^{(t)}\left(\tilde{\mathbf{w}}^{(t+1)} - \tilde{\mathbf{w}}^{(t)}\right). \end{aligned} \tag{D10}$$

In addition, it is easy to obtain that $\tilde{\mathbf{w}}^{(t)T}\Lambda^{(t)}\tilde{\mathbf{w}}^{(t)} = \sum_{i=1}^{m}\lambda_i^{(t)} = \tilde{\mathbf{w}}^{(t+1)T}\Lambda^{(t)}\tilde{\mathbf{w}}^{(t+1)}$. Together with (D6), we have

$$\begin{aligned} 0 &= \left(\tilde{\mathbf{w}}^{(t+1)T}\Lambda^{(t)}\tilde{\mathbf{w}}^{(t+1)} - \tilde{\mathbf{w}}^{(t)T}\Lambda^{(t)}\tilde{\mathbf{w}}^{(t)}\right) \\ & \quad + \left(\tilde{\mathbf{w}}^{(t)T}\tilde{\mathbf{K}}\tilde{\mathbf{w}}^{(t)} - \tilde{\mathbf{w}}^{(t)T}\tilde{\mathbf{K}}\tilde{\mathbf{w}}^{(t)}\right) \\ &= \left(\tilde{\mathbf{w}}^{(t)T}\tilde{\mathbf{K}}\tilde{\mathbf{w}}^{(t+1)} - \tilde{\mathbf{w}}^{(t)T}\Lambda^{(t)}\tilde{\mathbf{w}}^{(t)}\right) \\ & \quad + \left(\tilde{\mathbf{w}}^{(t)T}\Lambda^{(t)}\tilde{\mathbf{w}}^{(t+1)} - \tilde{\mathbf{w}}^{(t)T}\tilde{\mathbf{K}}\tilde{\mathbf{w}}^{(t)}\right) \\ &= \tilde{\mathbf{w}}^{(t)T}\tilde{\mathbf{K}}\left(\tilde{\mathbf{w}}^{(t+1)} - \tilde{\mathbf{w}}^{(t)}\right) \\ & \quad + \tilde{\mathbf{w}}^{(t)T}\Lambda^{(t)}\left(\tilde{\mathbf{w}}^{(t+1)} - \tilde{\mathbf{w}}^{(t)}\right). \end{aligned} \tag{D11}$$

Subtracting (D11) from (D10) leads to

$$\begin{aligned} & f(\tilde{\mathbf{w}}^{(t+1)}) - f(\tilde{\mathbf{w}}^{(t)}) \\ & \quad = \left(\tilde{\mathbf{w}}^{(t+1)T} - \tilde{\mathbf{w}}^{(t)T}\right)\tilde{\mathbf{K}}\left(\tilde{\mathbf{w}}^{(t+1)} - \tilde{\mathbf{w}}^{(t)}\right) \\ & \qquad + \left(\tilde{\mathbf{w}}^{(t+1)T} - \tilde{\mathbf{w}}^{(t)T}\right)\Lambda^{(t)}\left(\tilde{\mathbf{w}}^{(t+1)} - \tilde{\mathbf{w}}^{(t)}\right) \\ & \quad = \left(\tilde{\mathbf{w}}^{(t+1)} - \tilde{\mathbf{w}}^{(t)}\right)^T\left(\tilde{\mathbf{K}} + \Lambda^{(t)}\right)\left(\tilde{\mathbf{w}}^{(t+1)} - \tilde{\mathbf{w}}^{(t)}\right). \end{aligned} \tag{D12}$$

In (D12), $\Lambda^{(t)}$ is a positive semi-definite diagonal matrix because its each diagonal entry is not less than 0 (see the updating rule in (18)). Together with Lemma 1, we have that $\tilde{\mathbf{K}} + \Lambda^{(t)}$ is bound to be symmetric positive semi-definite. Thus, we obtain

$$f(\tilde{\mathbf{w}}^{(t+1)}) - f(\tilde{\mathbf{w}}^{(t)}) \geq 0 \Leftrightarrow f(\tilde{\mathbf{w}}^{(t+1)}) \geq f(\tilde{\mathbf{w}}^{(t)}), \tag{D13}$$

which shows that the objective function in (17) is nondecreasing. In terms of Lemma 2, the objective function has a upper bound. Putting these two conclusions together results in the convergence of the objective function. This completes the proof of Theorem 3. ■

References

[1] C. A. Micchelli. Interpolation of scattered data: Distance matrices and conditionally positive definite functions. *Constructive Approximation*, 2:11–12, 1986. 1

[2] J. Von Neumann. Some matrix-inequalities and metrization of matrix-space. *Tomsk University Review*, 1:286–300, 1937. 1