# Second-order Spectral Transform Block for 3D Shape Classification and Retrieval (supplemental material) 

Ruixuan Yu, Jian Sun, and Huibin Li

## I. Proof of Property 2

Proof. We present the proof of this property by two steps, i.e., we first prove the spectral transform operation keeps symmetry, then we prove that the transform keeps positive definite property.

Denoting the input matrix of spectral transform as $H$, with its singular value decomposition as $H=U \Lambda V^{\top}$, denoting the spectral transform as a function $f_{S T}(\cdot)$ which operates on the singular values of a given symmetric matrix, i.e., $f_{S T}(H)=$ $U f_{S T}(\Lambda) V^{\top}$.
Step 1. In this step, we prove this: if $H$ is symmetric, i.e., $H=H^{\top}$, then $f_{S T}(H)=f_{S T}(H)^{\top}$.

Notice that $H=U \Lambda V^{\top}$ is the singular value decomposition, then

$$
\begin{equation*}
U U^{\top}=I, V V^{\top}=I, U^{-1}=U^{\top}, V^{-1}=V^{\top} \tag{1}
\end{equation*}
$$

By $H=H^{\top}$, we achieve

$$
\begin{gather*}
U \Lambda V^{\top}=\left\{U \Lambda V^{\top}\right\}^{\top},  \tag{2}\\
\Lambda=U^{\top} V \Lambda U^{\top} V=\left\{U^{\top} V\right\} \Lambda\left\{U^{\top} V\right\}, \tag{3}
\end{gather*}
$$

and $\left\{U^{\top} V\right\} \Lambda\left\{U^{\top} V\right\}$ is a singular value decomposition of $\Lambda$. By the definition of $f_{S T}(\cdot)$, we have

$$
\begin{equation*}
f_{S T}(\Lambda)=\left\{U^{\top} V\right\} f_{S T}(\Lambda)\left\{U^{\top} V\right\} \tag{4}
\end{equation*}
$$

by which

$$
\begin{equation*}
U f_{S T}(\Lambda) V^{\top}=V f_{S T}(\Lambda) U^{\top} \tag{5}
\end{equation*}
$$

Notice that by the definition of $f_{S T}(\cdot)$, for the left term in Equation (5), we have

$$
\begin{equation*}
U f_{S T}(\Lambda) V^{\top}=f_{S T}\left(U \Lambda V^{\top}\right)=f_{S T}(H) \tag{6}
\end{equation*}
$$

for the right term in Equation (5),

$$
\begin{align*}
V f_{S T}(\Lambda) U^{\top} & =\left\{U f_{S T}(\Lambda) V^{\top}\right\}^{\top} \\
& =\left\{f_{S T}\left(U \Lambda V^{\top}\right)\right\}^{\top}  \tag{7}\\
& =\left\{f_{S T}(H)\right\}^{\top} .
\end{align*}
$$

So, we finally derive

$$
\begin{equation*}
f_{S T}(H)=f_{S T}(H)^{\top} \tag{8}
\end{equation*}
$$

That is, $f_{S T}(H)$ is a symmetric matrix, so the spectral transform keeps symmetry.
Step 2. In this step, we prove this: if $H$ is symmetric positive definite, then $f_{S T}(H)$ is also symmetric positive definite.

Notice that $H$ is symmetric positive definite, so $H=$ $U \Lambda U^{\top}$ and $f_{S T}(H)=U f_{S T}(\Lambda) U^{\top}$ are both singular value decomposition, then we have

$$
\begin{align*}
& U U^{\top}=I, U^{-1}=U^{\top}  \tag{9}\\
& f_{S T}(H) U=U f_{S T}(\Lambda) \tag{10}
\end{align*}
$$

and Equation (10) is in fact the eigenvalue decomposition of $f_{S T}(H)$. So $f_{S T}(\Lambda)$ is the eigenvalue matrix of $f_{S T}(H)$, it is a diagonal matrix.

Further more, $f_{S T}(\Lambda)$ is an element-wise operation on the diagonal elements of $\Lambda$ by first $l_{2}$-normalization and mixture of power function, if a diagonal elements of $\Lambda$ is non-negative, then its corresponding element in $f_{S T}(\Lambda)$ is also non-negative. So we achieve this: if the diagonal elements of $\Lambda$ are all nonnegative, then the diagonal elements of $f_{S T}(\Lambda)$ are all nonnegative.

Finally, recall that a matrix is positive definite if and only if its eigenvalues are all positive. So if $H$ is positive definite, then the diagonal elements of $\Lambda$ are all positive, and by above proof, the diagonal elements of $f_{S T}(\Lambda)$ are also positive, i.e., the eigenvalues of $f_{S T}(H)$ are all positive, which demonstrates that $f_{S T}(H)$ is positive definite. Combining with step 1 , it is easily to have this: if $H$ is symmetric positive definite, then $f_{S T}(H)$ is also symmetric positive definite.

So far, we have proved property 2 that the transformed descriptors after spectral transform still lie in SPDM-manifold / space of symmetric matrix.

## II. Computation of gradients in the spectral TRANSFORM OPERATION

We now present the computation of gradients (Eqs. (11), (12), (13)) in the spectral transform operation.

Given the partial derivative of loss $L$ with respect to $H^{\prime}$, i.e., $\frac{\partial L}{\partial H^{\prime}}$, we first compute $\frac{\partial L}{\partial \Lambda^{\prime}}$ with inspiration of matrix gradient computation in [1]. Given $H^{\prime}=U \Lambda^{\prime} U^{T}$, denoting the variation of $H^{\prime}$ as $d H^{\prime}$, and colon "." as matrix inner product operator that $X: Y=\operatorname{Trace}\left(X^{\top} Y\right)$ for any two matrices $X, Y$, we have

$$
\begin{align*}
\frac{\partial L}{\partial \Lambda^{\prime}}: d \Lambda^{\prime} & =\frac{\partial L}{\partial \Lambda^{\prime}}:\left(U^{\top} d H^{\prime} U\right)_{d i a g} \\
& =\left(\frac{\partial L}{\partial \Lambda^{\prime}}\right)_{d i a g}: U^{\top} d H^{\prime} U  \tag{11}\\
& =U\left(\frac{\partial L}{\partial \Lambda^{\prime}}\right)_{d i a g} U^{\top}: d H^{\prime}
\end{align*}
$$

where $(\cdot)_{\text {diag }}$ denotes an operator on matrix that sets all nondiagonal elements as 0 . Considering $\frac{\partial L}{\partial \Lambda^{\prime}}: d \Lambda^{\prime}=\frac{\partial L}{\partial H^{\prime}}: d H^{\prime}$, it is obvious that

$$
\begin{equation*}
\frac{\partial L}{\partial H^{\prime}}=U\left(\frac{\partial L}{\partial \Lambda^{\prime}}\right)_{d i a g} U^{\top} \tag{12}
\end{equation*}
$$

by which we derive

$$
\begin{equation*}
\frac{\partial L}{\partial \Lambda^{\prime}}=\left(U^{\top} \frac{\partial L}{\partial H^{\prime}} U\right)_{d i a g} \tag{13}
\end{equation*}
$$

Then, we compute the partial derivative of loss $L$ with respect to $\Gamma$, i.e., $\frac{\partial L}{\partial \Gamma}$, where $(\Gamma)_{i}=\gamma_{i}, i=0, \cdots, N_{m}$. Considering that $\Lambda_{j j}^{\prime}=f_{M P F}\left(\tilde{\Lambda}_{j j}\right)=\sum_{i=0}^{N_{m}} \gamma_{i} \tilde{\Lambda}_{j j}^{\alpha_{i}}, j=$ $1, \cdots, N_{\Lambda}$, then for $i=0,1, \ldots, N_{m}$ we have

$$
\begin{align*}
\frac{\partial \Lambda_{j j}^{\prime}}{\partial \gamma_{i}} & =\tilde{\Lambda}_{j j}^{\alpha_{i}}, j=1, \cdots, N_{\Lambda}  \tag{14}\\
\frac{\partial L}{\partial \gamma_{i}} & =\sum_{j=1}^{N_{\Lambda}} \frac{\partial L}{\partial \Lambda_{j j}^{\prime}} \frac{\partial \Lambda_{j j}^{\prime}}{\partial \gamma_{i}} \\
& =\sum_{j=1}^{N_{\Lambda}} \tilde{\Lambda}_{j j}^{\alpha_{i}} \frac{\partial L}{\partial \Lambda_{j j}^{\prime}}  \tag{15}\\
& =\sum_{j=1}^{N_{\Lambda}}\left(\tilde{\Lambda}^{\alpha_{i}} \frac{\partial L}{\partial \Lambda^{\prime}}\right)_{j j} \\
& =\left(\tilde{\Lambda}^{\alpha_{i}} \frac{\partial L}{\partial \Lambda^{\prime}}\right)_{G d i a g}^{\top} \mathbf{1}
\end{align*}
$$

where $(\cdot)_{\text {Gdiag }}$ is the operator of vectorizing the diagonal elements of a matrix. Then we have

$$
\begin{equation*}
\left(\frac{\partial L}{\partial \Gamma}\right)_{i}=\left(\tilde{\Lambda}^{\alpha_{i}} \frac{\partial L}{\partial \Lambda^{\prime}}\right)_{G d i a g}^{\top} \mathbf{1}, i=0,1, \ldots, N_{m} \tag{16}
\end{equation*}
$$

Finally, we present the computation of $\frac{\partial L}{\partial \Omega}$. Considering that $\gamma_{i}=\frac{e^{\omega_{i}}}{\sum_{j=0}^{N_{m}} e^{\omega_{j}}}, i=0,1, \ldots, N_{m}$, we have

$$
\begin{equation*}
\frac{\partial \gamma_{i}}{\partial \omega_{i}}=\gamma_{i}\left(1-\gamma_{i}\right), \frac{\partial \gamma_{j}}{\partial \omega_{i}}=-\gamma_{i} \gamma_{j}, j \neq i \tag{17}
\end{equation*}
$$

and it is easy to derive

$$
\begin{align*}
\frac{\partial L}{\partial \omega_{i}} & =\frac{\partial L}{\partial \gamma_{i}} \frac{\partial \gamma_{i}}{\partial \omega_{i}}+\sum_{j \neq i} \frac{\partial L}{\partial \gamma_{j}} \frac{\partial \gamma_{j}}{\partial \omega_{i}} \\
& =\frac{\partial L}{\partial \gamma_{i}} \gamma_{i}\left(1-\gamma_{i}\right)-\sum_{j \neq i} \frac{\partial L}{\partial \gamma_{j}} \gamma_{i} \gamma_{j} \\
& =\frac{\partial L}{\partial \gamma_{i}} \gamma_{i}-\sum_{j=1}^{N_{m}} \frac{\partial L}{\partial \gamma_{j}} \gamma_{i} \gamma_{j}  \tag{18}\\
& =\left(\frac{\partial L}{\partial \gamma_{i}}-\sum_{j=1}^{N_{m}} \frac{\partial L}{\partial \gamma_{j}} \gamma_{j}\right) \gamma_{i}, \quad i=0,1, \ldots, N_{m}
\end{align*}
$$

which leads to

$$
\begin{equation*}
\frac{\partial L}{\partial \Omega}=\left(\frac{\partial L}{\partial \Gamma}-\Gamma^{\top} \frac{\partial L}{\partial \Gamma}\right) \odot \Gamma \tag{19}
\end{equation*}
$$

where $\odot$ is the Hadamard product operator.

## III. Proof of Proposition 1

Proof. We first prove the stability of surface second-order average-pooling, then prove the stability of surface secondorder max-pooling.

1) For the surface second-order average-pooling, notice that the descretized forms approximate the continuous form when the number of points approximates infinity, so we just present the stability of the continuous form, and the descretized ones can be proved similarly. We have

$$
\begin{align*}
\|H-\tilde{H}\|_{F} & =\left\|\frac{1}{|\mathcal{S}|} \int_{\mathcal{S}} \mathbf{h}(s) \mathbf{h}(s)^{\top} d s-\frac{1}{|\mathcal{S}|} \int_{\mathcal{S}} \tilde{\mathbf{h}}(s) \tilde{\mathbf{h}}(s)^{\top} d s\right\|_{F} \\
& =\left\|\frac{1}{|\mathcal{S}|} \int_{\mathcal{S}}\left\{\mathbf{h}(s) \mathbf{h}(s)^{\top}-\tilde{\mathbf{h}}(s) \tilde{\mathbf{h}}(s)^{\top}\right\} d s\right\|_{F} \\
& \leq \frac{1}{|\mathcal{S}|} \int_{\mathcal{S}}\left\|\mathbf{h}(s) \mathbf{h}(s)^{\top}-\tilde{\mathbf{h}}(s) \tilde{\mathbf{h}}(s)^{\top}\right\|_{F} d s \tag{20}
\end{align*}
$$

Considering that

$$
\begin{align*}
& \left\|\mathbf{h}(s) \mathbf{h}(s)^{\top}-\tilde{\mathbf{h}}(s) \tilde{\mathbf{h}}(s)^{\top}\right\|_{F} \\
& =\left\|\mathbf{h}(s)\left\{\mathbf{h}(s)^{\top}-\tilde{\mathbf{h}}(s)^{\top}\right\}+\tilde{\mathbf{h}}(s)^{\top}\{\mathbf{h}(s)-\tilde{\mathbf{h}}(s)\}\right\|_{F} \\
& \leq\|\mathbf{h}(s)\|_{F}\|\mathbf{h}(s)-\tilde{\mathbf{h}}(s)\|_{F}+\|\mathbf{h}(s)\|_{F}\|\mathbf{h}(s)-\tilde{\mathbf{h}}(s)\|_{F} \\
& \leq 2 M \epsilon \tag{21}
\end{align*}
$$

we achieve

$$
\begin{align*}
\|H-\widetilde{H}\|_{F} & \leq \frac{1}{|\mathcal{S}|} \int_{\mathcal{S}}\left\|\mathbf{h}(s) \mathbf{h}(s)^{\top}-\tilde{\mathbf{h}}(s) \tilde{\mathbf{h}}(s)^{\top}\right\|_{F} d s \\
& \leq 2 M \epsilon \frac{1}{|\mathcal{S}|} \int_{\mathcal{S}} d s  \tag{22}\\
& =2 M \epsilon
\end{align*}
$$

2) For the surface second-order max-pooling, we prove its stability by two steps.
Step 1. We first prove this: $\left|h_{i}(s) h_{j}(s)-\tilde{h}_{i}(s) \tilde{h}_{j}(s)\right| \leq 2 M \epsilon$. This is because

$$
\begin{equation*}
\|\mathbf{h}(s)-\tilde{\mathbf{h}}(s)\|_{F}=\left\{\sum_{i=1}^{d}\left|h_{i}(s)-\tilde{h}_{i}(s)\right|^{2}\right\}^{\frac{1}{2}} \leq \epsilon \tag{23}
\end{equation*}
$$

by which we derive

$$
\begin{align*}
& \sum_{i=1}^{d}\left|h_{i}(s)-\tilde{h}_{i}(s)\right|^{2} \leq \epsilon^{2}  \tag{24}\\
& \left|h_{i}(s)-\tilde{h}_{i}(s)\right| \leq \epsilon \\
& \left|h_{j}(s)-\tilde{h}_{j}(s)\right| \leq \epsilon
\end{align*}
$$

Recall that $M$ is the upper bound of input descriptors,

$$
\begin{equation*}
\|\mathbf{h}(s)\|_{F}=\left\{\sum_{j=1}^{d} h_{j}(s)^{2}\right\}^{\frac{1}{2}} \leq M \tag{25}
\end{equation*}
$$

$$
\left|h_{j}(s)\right| \leq M
$$

Similarly, we have

$$
\begin{equation*}
\left|\tilde{h}_{i}(s)\right| \leq M \tag{26}
\end{equation*}
$$

$$
\begin{align*}
& \left|h_{i}(s) h_{j}(s)-\tilde{h}_{i}(s) \tilde{h}_{j}(s)\right| \\
& =\left|h_{i}(s) h_{j}(s)-\tilde{h}_{i}(s) h_{j}(s)+\tilde{h}_{i}(s) h_{j}(s)-\tilde{h}_{i}(s) \tilde{h}_{j}(s)\right| \\
& =\left|\left(h_{i}(s)-\tilde{h}_{i}(s)\right) h_{j}(s)+\tilde{h}_{i}(s)\left(h_{j}(s)-\tilde{h}_{j}(s)\right)\right| \\
& \leq 2 M \epsilon \tag{27}
\end{align*}
$$

Step 2. Then, we prove this: $\|H-\widetilde{H}\|_{F} \leq 2 M d \epsilon$. Without loss of generality, we set $s_{1}=\arg \max _{s \in \mathcal{S}} h_{i}(s) h_{j}(s), s_{2}=$ $\arg \max _{s \in \mathcal{S}} \tilde{h}_{i}(s) \tilde{h}_{j}(s)$, then

$$
\begin{align*}
& h_{i}\left(s_{2}\right) h_{j}\left(s_{2}\right) \leq h_{i}\left(s_{1}\right) h_{j}\left(s_{1}\right) \\
& \tilde{h}_{i}\left(s_{1}\right) \tilde{h}_{j}\left(s_{1}\right) \leq \tilde{h}_{i}\left(s_{2}\right) \tilde{h}_{j}\left(s_{2}\right) \tag{28}
\end{align*}
$$

$h_{i}\left(s_{2}\right) h_{j}\left(s_{2}\right)-\tilde{h}_{i}\left(s_{2}\right) \tilde{h}_{j}\left(s_{2}\right) \leq h_{i}\left(s_{1}\right) h_{j}\left(s_{1}\right)-\tilde{h}_{i}\left(s_{2}\right) \tilde{h}_{j}\left(s_{2}\right)$, $h_{i}\left(s_{1}\right) h_{j}\left(s_{1}\right)-\tilde{h}_{i}\left(s_{2}\right) \tilde{h}_{j}\left(s_{2}\right) \leq h_{i}\left(s_{1}\right) h_{j}\left(s_{1}\right)-\tilde{h}_{i}\left(s_{1}\right) \tilde{h}_{j}\left(s_{1}\right)$.

Combining above two inequalities, we derive

$$
\begin{align*}
& h_{i}\left(s_{2}\right) h_{j}\left(s_{2}\right)-\tilde{h}_{i}\left(s_{2}\right) \tilde{h}_{j}\left(s_{2}\right) \leq h_{i}\left(s_{1}\right) h_{j}\left(s_{1}\right)-\tilde{h}_{i}\left(s_{2}\right) \tilde{h}_{j}\left(s_{2}\right) \\
& \leq h_{i}\left(s_{1}\right) h_{j}\left(s_{1}\right)-\tilde{h}_{i}\left(s_{1}\right) \tilde{h}_{j}\left(s_{1}\right) \tag{30}
\end{align*}
$$

Recall that for an inequality $a \leq b \leq c$, we have $|b| \leq$ $\max \{|a|,|c|\}$, and by step 1 ,

$$
\begin{align*}
& \left|h_{i}\left(s_{2}\right) h_{j}\left(s_{2}\right)-\tilde{h}_{i}\left(s_{2}\right) \tilde{h}_{j}\left(s_{2}\right)\right| \leq 2 M \epsilon  \tag{31}\\
& \left|h_{i}\left(s_{1}\right) h_{j}\left(s_{1}\right)-\tilde{h}_{i}\left(s_{1}\right) \tilde{h}_{j}\left(s_{1}\right)\right| \leq 2 M \epsilon
\end{align*}
$$

Combining with Equation (30)

$$
\begin{equation*}
\left|h_{i}\left(s_{1}\right) h_{j}\left(s_{1}\right)-\tilde{h}_{i}\left(s_{2}\right) \tilde{h}_{j}\left(s_{2}\right)\right| \leq 2 M \epsilon \tag{32}
\end{equation*}
$$

Finally, we have

$$
\begin{align*}
& \|H-\widetilde{H}\|_{F} \\
& =\left\{\sum_{i, j=1}^{d}\left|h_{i}\left(s_{1}^{i j}\right) h_{j}\left(s_{1}^{i j}\right)-\tilde{h}_{i}\left(s_{2}^{i j}\right) \tilde{h}_{j}\left(s_{2}^{i j}\right)\right|^{2}\right\}^{\frac{1}{2}} \\
& \leq\left\{\sum_{i, j}^{d}(2 M \epsilon)^{2}\right\}^{\frac{1}{2}}  \tag{33}\\
& =2 M d \epsilon
\end{align*}
$$

Notice that in above inequality, we use $s_{1}^{i j}$ and $s_{2}^{i j}$ to denote the maximum points where $H_{i j}$ and $\widetilde{H}_{i j}$ reach the maximum values.

## IV. Proof of Proposition 2

Before the proof of Proposition 2, we first present a lemma:
Lemma 1: Given two matrices $X, Y$, if $\|X-Y\|_{F} \leq \epsilon$, then $\left\|\frac{X}{\|X\|_{F}}-\frac{Y}{\|Y\|_{F}}\right\|_{F} \leq \frac{2}{\|X\|_{F}} \epsilon$. Moreover, let $\widehat{X}=$ $Q_{x} X P_{x}^{T}, \widehat{Y}=Q_{y} Y P_{y}^{T}$ be the singular value decompositions of symmetric matrices $\widehat{X}, \widehat{Y}$, then if $\|\widehat{X}-\widehat{Y}\|_{F} \leq \epsilon$, we have $\left\|Q_{x} \frac{X}{\|Y\|_{F}} P_{x}^{T}-Q_{y} \frac{Y}{\|Y\|_{F}} P_{y}^{T}\right\|_{F} \leq \frac{2}{\|\widehat{X}\|_{F}} \epsilon$.
Proof. We first prove the first part of this lemma, i.e., if $\| X-$ $Y \|_{F} \leq \epsilon$, then $\left\|\frac{X}{\|X\|_{F}}-\frac{Y}{\|Y\|_{F}}\right\|_{F} \leq \frac{2}{\|X\|_{F}} \epsilon$, in following Step 1. Then we give the proof of the second part by Step 2.

Step 1. Assuming $\|X\|_{F}=a,\|Y\|_{F}=b$, then

$$
\begin{align*}
\left\|\frac{X}{a}-\frac{Y}{b}\right\|_{F} & =\frac{1}{a}\left\|X-\frac{a}{b} Y\right\|_{F} \\
& =\frac{1}{a}\left\|X-\frac{a+b-b}{b} Y\right\|_{F} \\
& =\frac{1}{a}\left\|X-Y-\frac{a-b}{b} Y\right\|_{F}  \tag{34}\\
& \leq \frac{1}{a}\|X-Y\|_{F}+\frac{1}{a} \frac{|a-b|}{b}\|Y\|_{F} \\
& =\frac{1}{a}\|X-Y\|_{F}+\frac{|a-b|}{a}
\end{align*}
$$

because $|a-b|=\left|\|X\|_{F}-\|Y\|_{F}\right| \leq\|X-Y\|_{F}$, and $\| X-$ $Y \|_{F} \leq \epsilon$,

$$
\begin{equation*}
\left\|\frac{X}{a}-\frac{Y}{b}\right\|_{F} \leq \frac{1}{a}\|X-Y\|_{F}+\frac{|a-b|}{a} \leq \frac{2}{a} \epsilon \tag{35}
\end{equation*}
$$

That is, $\left\|\frac{X}{\|\left. X\right|_{F}}-\frac{Y}{\|Y\|_{F}}\right\|_{F} \leq \frac{2}{\|\left. X\right|_{F}} \epsilon$.
Step 2. In this step, we prove the second part, i.e., let $\widehat{X}=$ $Q_{x} X P_{x}^{T}, \widehat{Y}=Q_{y} Y P_{y}^{T}$ be the singular value decompositions of symmetric matrices $\widehat{X}, \widehat{Y}$, then if $\|\widehat{X}-\widehat{Y}\|_{F} \leq \epsilon$, we have $\left\|Q_{x} \frac{X}{\|X\|_{F}} P_{x}^{T}-Q_{y} \frac{Y}{\|Y\|_{F}} P_{y}^{T}\right\|_{F} \leq \frac{2}{\|\widehat{X}\|} \epsilon$.

At first, we give a property of the symmetric matrix: for any symmetric matrix $\widehat{X}$ with the singular values decomposition as $\widehat{X}=Q_{x} X P_{x}^{T},\|\widehat{X}\|_{F}=\|X\|_{F}$. This is because

$$
\begin{equation*}
\|\widehat{X}\|_{F}=\left\|Q_{x} X P_{x}^{T}\right\|_{F}=\left\|Q_{x}\right\|_{F}\|X\|_{F}\left\|P_{x}^{T}\right\|_{F} \tag{36}
\end{equation*}
$$

$Q_{x}, P_{x}$ are singular vector matrix, $\left\|Q_{x}\right\|_{F}=1,\left\|P_{x}^{T}\right\|_{F}=1$, so we have

$$
\begin{equation*}
\|\widehat{X}\|_{F}=\|X\|_{F} \tag{37}
\end{equation*}
$$

The same is for $\widehat{Y}$, i.e., $\|\widehat{Y}\|_{F}=\|Y\|_{F}$.
Then, we derive

$$
\begin{align*}
& \left\|Q_{x} \frac{X}{\|X\|_{F}} P_{x}^{T}-Q_{y} \frac{Y}{\|Y\|_{F}} P_{y}^{T}\right\|_{F} \\
& =\left\|Q_{x} \frac{X}{\|\widehat{X}\|_{F}} P_{x}^{T}-Q_{y} \frac{Y}{\|\widehat{Y}\|_{F}} P_{y}^{T}\right\|_{F} \\
& =\left\|\frac{Q_{x} X P_{y}^{T}}{\|\widehat{X}\|_{F}}-\frac{Q_{y} Y P_{y}^{T}}{\|\widehat{Y}\|_{F}}\right\|_{F}  \tag{38}\\
& =\left\|\frac{\widehat{X}}{\|\widehat{X}\|_{F}}-\frac{\widehat{Y}}{\|\widehat{Y}\|_{F}}\right\|_{F}
\end{align*}
$$

by step 1 , if $\|\widehat{X}-\widehat{Y}\|_{F} \leq \epsilon$,

$$
\begin{equation*}
\left\|Q_{x} \frac{X}{\|X\|_{F}} P_{x}^{T}-Q_{y} \frac{Y}{\|Y\|_{F}} P_{y}^{T}\right\|_{F} \leq \frac{2}{\|\widehat{X}\|} \epsilon \tag{39}
\end{equation*}
$$

Lemma 1 demonstrates the stability of the normalization operation, either on the matrices themselves or on the singular values of the symmetric matrices.

For Proposition 2, we prove it by following three steps.

Proof. Step 1. For the input symmetric matrix $H \in R^{d \times d}$ and output $H^{\prime} \in R^{d \times d}$ with their singular value decomposition as $H=U \Lambda V^{\top}, H^{\prime}=U \Lambda^{\prime} V^{\top}$ as in our paper, we have

$$
\begin{equation*}
\|H\|_{F}=\sqrt{\sum_{i=1}^{d} \Lambda_{i}^{2}}, \quad\left\|H^{\prime}\right\|_{F}=\sqrt{\sum_{i=1}^{d} \Lambda_{i}^{\prime 2}} \tag{40}
\end{equation*}
$$

Note that in the spectral transform operation with $H$ as input, $\underset{\sim}{\text { we }}$ first normalize the singular value matrix $\Lambda$, achieving $\widetilde{\Lambda}$ such that $\sum_{i=1}^{d} \widetilde{\Lambda}_{i}^{2}=1$, with $\left|\widetilde{\Lambda}_{i}\right| \leq 1$. After function $f_{M P F}(\cdot)$, we get $\Lambda^{\prime}$ with its elements as $\Lambda_{i}^{\prime}=f_{M P F}\left(\widetilde{\Lambda}_{i}\right)$. By the definition of $f_{M P F}(\cdot)$, it is easy to achieve that: $\Lambda_{i}^{\prime} \leq 1$, and combining it with Equation (40), we finally get

$$
\begin{equation*}
\left\|H^{\prime}\right\|_{F}=\sqrt{\sum_{i=1}^{d} \Lambda_{i}^{\prime 2}} \leq \sqrt{d} \tag{41}
\end{equation*}
$$

For another input symmetric matrix $\widetilde{H}$ with its output as $\widetilde{H}^{\prime}$, we have

$$
\begin{equation*}
\left\|H^{\prime}-\widetilde{H}^{\prime}\right\|_{F} \leq\left\|H^{\prime}\right\|_{F}+\left\|\tilde{H}^{\prime}\right\|_{F} \leq 2 \sqrt{d} \tag{42}
\end{equation*}
$$

Step 2. For the input symmetric matrix (with its singular values are already normalized), denoting the spectral transform as $f_{S T}(\cdot)$, and it is in fact following function that works on eigenvalues:
$f_{S T}(x)=\operatorname{sign}(x) f_{M P F}(\operatorname{abs}(x)), x \in[-1,-\delta] \cup\{0\} \cup[\delta, 1]$,
where $\delta=\inf \{\bar{\lambda} \mid \bar{\lambda} \neq 0\}, \bar{\lambda}$ is $l_{2}$-normalized $\lambda$ which is singular value of input matrix. We now prove this:

$$
\begin{equation*}
\left|f_{S T}\left(x_{1}\right)-f_{S T}\left(x_{2}\right)\right| \leq \delta^{\alpha_{0}-1}\left|x_{1}-x_{2}\right| \tag{44}
\end{equation*}
$$

where $\alpha_{0} \in[0,1]$ is the minimum power index utilized in $f_{M P F}(\cdot)$ and we discuss it as follows.

1) when $x_{i} \neq 0, \operatorname{abs}\left(x_{i}\right) \geq \delta, i=1,2$, we have following situations:
a) $x_{i} \geq \delta, i=1,2$, then

$$
\begin{equation*}
\left|f_{S T}\left(x_{1}\right)-f_{S T}\left(x_{2}\right)\right|=\left|f_{M P F}\left(x_{1}\right)-f_{M P F}\left(x_{2}\right)\right| \tag{45}
\end{equation*}
$$

Considering $\delta \geq 0, \delta \leq 1, \delta^{\alpha_{0}-1} \geq 1$, and for function $f_{M P F}(\cdot)$ that being defined on $[\delta, 1]$, it is Lipschitz continuous with Lipschitz constant as $L$,

$$
\begin{align*}
L & =\sup _{x \in[\delta, 1]} f_{M P F}^{\prime}(x) \\
& =\sum_{i=1}^{N_{m}} \gamma_{i} \alpha_{i} \delta^{\alpha_{i}-1}  \tag{46}\\
& \leq \sum_{i=1}^{N_{m}} \gamma_{i} \delta^{\alpha_{0}-1} \\
& =\delta^{\alpha_{0}-1}
\end{align*}
$$

So, we derive

$$
\begin{equation*}
\left|f_{S T}\left(x_{1}\right)-f_{S T}\left(x_{2}\right)\right| \leq \delta^{\alpha_{0}-1}\left|x_{1}-x_{2}\right| \tag{47}
\end{equation*}
$$

b) $x_{i} \leq-\delta, i=1,2$, then

$$
\begin{align*}
& \left|f_{S T}\left(x_{1}\right)-f_{S T}\left(x_{2}\right)\right| \\
= & \left|-f_{M P F}\left(\operatorname{abs}\left(x_{1}\right)\right)+f_{M P F}\left(\operatorname{abs}\left(x_{2}\right)\right)\right| \\
\leq & \delta^{\alpha_{0}-1}\left|\operatorname{abs}\left(x_{1}\right)-\operatorname{abs}\left(x_{2}\right)\right|  \tag{48}\\
= & \delta^{\alpha_{0}-1}\left|x_{1}-x_{2}\right|
\end{align*}
$$

which also derives

$$
\begin{equation*}
\left|f_{S T}\left(x_{1}\right)-f_{S T}\left(x_{2}\right)\right| \leq \delta^{\alpha_{0}-1}\left|x_{1}-x_{2}\right| \tag{49}
\end{equation*}
$$

c) $x_{1} \geq \delta, x_{2} \leq-\delta$, then

$$
\begin{align*}
& \left|f_{S T}\left(x_{1}\right)-f_{S T}\left(x_{2}\right)\right| \\
= & \left|f_{M P F}\left(x_{1}\right)+f_{M P F}\left(\operatorname{abs}\left(x_{2}\right)\right)\right| \\
= & \left|f_{M P F}\left(x_{1}\right)\right|+\left|f_{M P F}\left(\operatorname{abs}\left(x_{2}\right)\right)\right|  \tag{50}\\
\leq & \delta^{\alpha_{0}-1}\left|x_{1}\right|+\delta^{\alpha_{0}-1}\left|x_{2}\right| \\
= & \delta^{\alpha_{0}-1}\left|x_{1}-x_{2}\right|,
\end{align*}
$$

which is

$$
\begin{equation*}
\left|f_{S T}\left(x_{1}\right)-f_{S T}\left(x_{2}\right)\right| \leq \delta^{\alpha_{0}-1}\left|x_{1}-x_{2}\right| \tag{51}
\end{equation*}
$$

d) $x_{1} \leq-\delta, x_{2} \geq \delta$, it is the same as $x_{1}<-\delta, x_{2}>$ $\delta$ as above, and we do not repeat it.
2) $x_{1}=0, x_{2} \neq 0$ or $x_{2}=0, x_{1} \neq 0$. Without loss of generality, we set $x_{1} \neq 0, x_{2}=0$. So we have

$$
\begin{align*}
\left|f_{S T}\left(x_{1}\right)-f_{S T}\left(x_{2}\right)\right| & =\left|f_{M P F}\left(\left|x_{1}\right|\right)\right| \\
& =\sum_{i=0}^{N_{m}} \gamma_{i}\left|x_{1}\right|^{\alpha_{i}}  \tag{52}\\
& \leq \sum_{i=0}^{N_{m}} \gamma_{i}\left|x_{1}\right|^{\alpha_{0}} \\
& =\left|x_{1}\right|^{\alpha_{0}} .
\end{align*}
$$

On the other hand, we have

$$
\begin{align*}
\delta^{\alpha_{0}-1}\left|x_{1}-x_{2}\right| & =\delta^{\alpha_{0}-1}\left|x_{1}\right| \\
& \geq\left|x_{1}\right|^{\alpha_{0}-1}\left|x_{1}\right|  \tag{53}\\
& =\left|x_{1}\right|^{\alpha_{0}}
\end{align*}
$$

Combining with Eqn.(52), we now have

$$
\begin{equation*}
\left|f_{S T}\left(x_{1}\right)-f_{S T}\left(x_{2}\right)\right| \leq \delta^{\alpha_{0}-1}\left|x_{1}-x_{2}\right| \tag{54}
\end{equation*}
$$

3) $x_{i}=0, i=1,2$, then

$$
\begin{align*}
\left|f_{S T}\left(x_{1}\right)-f_{S T}\left(x_{2}\right)\right| & =|0-0| \\
& \leq \delta^{\alpha_{0}-1}|0-0|  \tag{55}\\
& =\delta^{\alpha_{0}-1}\left|x_{1}-x_{2}\right|
\end{align*}
$$

Till now, we have proved the property (Eqn. (44)) of $f_{S T}(\cdot)$. Step 3. For two input symmetric matrices $H, \widetilde{H}$, in the spectral transform operation, we first normalize their singular values, achieving $H^{*}, \widetilde{H}^{*}$. By Lemma 1 , if $\|H-\widetilde{H}\|_{F} \leq \epsilon$, we have

$$
\begin{equation*}
\left\|H^{*}-\widetilde{H}^{*}\right\|_{F} \leq \frac{2}{M} \epsilon \tag{56}
\end{equation*}
$$

with $M$ as the lower bound of the inputs. Denoting the SVD and eigenvalue decomposition of $H^{*}, \widetilde{H}^{*}$ as ${\underset{\sim}{H}}^{*}=$ $U_{1} \Lambda_{1} V_{1}^{\top}, \widetilde{H}^{*}=U_{2} \Lambda_{2} V_{2}^{\top}$ and $H^{*}=U_{1} D_{1} U_{1}^{\top}, \widetilde{H}^{*}=$
$U_{2} D_{2} U_{2}^{\top}$ respectively, where $D_{1}=\Lambda_{1} \odot S_{1}, D_{2}=\Lambda_{1} \odot S_{2}$ with $S_{1}, S_{2}$ indicating the sign of corresponding eigenvalues, after the spectral transform, we derive $H^{\prime}, \widetilde{H}^{\prime}$. Combining with Step 2, we have

$$
\begin{align*}
\left\|H^{\prime}-\widetilde{H}^{\prime}\right\|_{F} & =\left\|U_{1} f_{S T}\left(D_{1}\right) U_{1}^{\top}-U_{2} f_{S T}\left(D_{2}\right) U_{2}^{\top}\right\|_{F} \\
& =\left\|U_{1}^{\top} U_{1} f_{S T}\left(D_{1}\right) U_{1}^{\top} U_{2}-U_{1}^{\top} U_{2} f_{S T}\left(D_{2}\right) U_{2}^{\top} U_{2}\right\|_{F} \\
& =\left\|f_{S T}\left(D_{1}\right) U_{1}^{\top} U_{2}-U_{1}^{\top} U_{2} f_{S T}\left(D_{2}\right)\right\|_{F} \tag{57}
\end{align*}
$$

denoting $U_{1}^{\top} U_{2}$ as $Z$, then

$$
\begin{align*}
\left\|H^{\prime}-\widetilde{H}^{\prime}\right\|_{F} & =\sum_{i, j=1}^{d}\left|f_{S T}\left(D_{1}^{i}\right)-f_{S T}\left(D_{2}^{j}\right)\right| \cdot\left|Z_{i j}\right| \\
& \leq \sum_{i, j=1}^{d} \delta^{\alpha_{0}-1}\left|D_{1}^{i}-D_{2}^{j}\right| \cdot\left|Z_{i j}\right|  \tag{58}\\
& =\delta^{\alpha_{0}-1}| | H^{*}-\widetilde{H}^{*} \|_{F} \\
& \leq \delta^{\alpha_{0}-1} \frac{2}{M} \epsilon
\end{align*}
$$

Combining with Step 1, we finally get

$$
\begin{equation*}
\left\|H^{\prime}-\widetilde{H}^{\prime}\right\|_{F} \leq \min \left\{\delta^{\alpha_{0}-1} \frac{2}{M} \epsilon, 2 \sqrt{d}\right\} \tag{59}
\end{equation*}
$$

## REFERENCES

[1] C. Ionescu, O. Vantzos, and C. Sminchisescu, "Matrix backpropagation for deep networks with structured layers," in ICCV, 2015, pp. 2965-2973.

