# Beyond Mahalanobis Metric: Cayley-Klein Metric Learning
## Supplementary Material

Yanhong Bi, Bin Fan, Fuchao Wu
Institute of Automation, Chinese Academy of Sciences
Beijing, 100190, China
{yanhong.bi, bfan, fcwu}@nlpr.ia.ac.cn

In this supplementary material, we first give the proofs of the three propositions given in our paper, then show more experimental results of similarity search on the OSR and PubFig datasets.

## 1. Proofs of the propositions

**Proposition 1**: Given two points $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n(\mathbb{B}^n)$, let $\mathbf{z}_+$ and $\mathbf{z}_-$ be the points at which the straight line determined by $\mathbf{x}$ and $\mathbf{y}$ intersects the quadric surface $\Omega = \{\mathbf{z} | \psi(\mathbf{z}, \mathbf{z}) = 0\}$, then:

$$\rho(\mathbf{x}, \mathbf{y}) = \frac{k}{2} \left|\log r(\mathbf{xy}, \mathbf{z}_+\mathbf{z}_-)\right| \tag{1}$$

where $r(\mathbf{xy}, \mathbf{z}_+\mathbf{z}_-)$ is the cross-ratio of this quadruple of points $\{\mathbf{x}, \mathbf{y}, \mathbf{z}_+, \mathbf{z}_-\}$.

***Proof.*** Let $\mathbf{z} = s\mathbf{x} + (1-s)\mathbf{y}, (s \in \mathbb{R})$ be the line determined by $\mathbf{x}$ and $\mathbf{y}$, then the two intersections of $\mathbf{z}$ and $\Omega$ satisfy:

$$\psi(s\mathbf{x} + (1-s)\mathbf{y}, s\mathbf{x} + (1-s)\mathbf{y}) = 0 \tag{2}$$

According to definition, $\psi$ is the bilinear form of matrix $\Psi$, then we have

$$\begin{aligned}
&\psi(s\mathbf{x} + (1-s)\mathbf{y}, s\mathbf{x} + (1-s)\mathbf{y}) \\
&= (s\mathbf{x}^T + (1-s)\mathbf{y}^T, 1)\Psi \begin{pmatrix} s\mathbf{x} + (1-s)\mathbf{y} \\ 1 \end{pmatrix} \\
&= s^2(\mathbf{x}^T, 1)\Psi \begin{pmatrix} \mathbf{x} \\ 1 \end{pmatrix} + 2s(1-s)(\mathbf{x}^T, 1)\Psi \begin{pmatrix} \mathbf{y} \\ 1 \end{pmatrix} \\
&\quad + (1-s)^2(\mathbf{y}^T, 1)\Psi \begin{pmatrix} \mathbf{y} \\ 1 \end{pmatrix} \\
&= s^2\psi(\mathbf{x}, \mathbf{x}) + 2s(1-s)\psi(\mathbf{x}, \mathbf{y}) + (1-s)^2\psi(\mathbf{y}, \mathbf{y})
\end{aligned} \tag{3}$$

Hence, two solutions $s_\pm$ of Eq. (2) satisfy:

$$\frac{s_\pm}{1 - s_\pm} = \frac{-\psi(\mathbf{x}, \mathbf{y}) \pm \sqrt{\psi^2(\mathbf{x}, \mathbf{y}) - \psi(\mathbf{x}, \mathbf{x}) \cdot \psi(\mathbf{y}, \mathbf{y})}}{\psi(\mathbf{x}, \mathbf{x})} \tag{4}$$

Accordingly, the two intersections are:

$$\mathbf{z}_\pm = s_\pm \mathbf{x} + (1 - s_\pm)\mathbf{y} \tag{5}$$

Note that in elliptic geometry space, $\mathbf{z}_\pm$ is a pair of conjugate complex points.

The cross-ratio of the quadruple of points $\{\mathbf{x}, \mathbf{y}, \mathbf{z}_+, \mathbf{z}_-\}$ is:

$$\begin{aligned}
r(\mathbf{xy}, \mathbf{z}_+\mathbf{z}_-) &= \frac{s_-}{1 - s_-} : \frac{s_+}{1 - s_+} \\
&= \frac{\psi(\mathbf{x}, \mathbf{y}) + \sqrt{\psi^2(\mathbf{x}, \mathbf{y}) - \psi(\mathbf{x}, \mathbf{x}) \cdot \psi(\mathbf{y}, \mathbf{y})}}{\psi(\mathbf{x}, \mathbf{y}) - \sqrt{\psi^2(\mathbf{x}, \mathbf{y}) - \psi(\mathbf{x}, \mathbf{x}) \cdot \psi(\mathbf{y}, \mathbf{y})}}
\end{aligned} \tag{6}$$

Therefore, we obtain:

$$\rho(\mathbf{x}, \mathbf{y}) = \frac{k}{2} \left|\log r(\mathbf{xy}, \mathbf{z}_+\mathbf{z}_-)\right| \qquad \square$$

**Proposition 2**: For any $\mathbf{G} \in \mathbb{G}(\Psi)$, there exists a $(n+1)$-dimensional antisymmetric matrix $\mathbf{W}$ satisfying:

$$\mathbf{G} = (\Psi + \mathbf{W})^{-1}(\Psi - \mathbf{W}) \tag{7}$$

***Proof.*** Based on the definition of $\mathbb{G}(\Psi)$, it holds $\mathbf{G}^{-T}\Psi\mathbf{G}^{-1} = \Psi$, *i.e.* $\Psi = \mathbf{G}^T\Psi\mathbf{G}$. Since $\mathbb{G}(\Psi)$ is a linear matrix group, for any $\mathbf{G} \in \mathbb{G}(\Psi)$, its Cayley transform can be represented as:

$$\mathbf{B} = (\mathbf{I} - \mathbf{G})(\mathbf{I} + \mathbf{G})^{-1} \tag{8}$$

Similarly, $\mathbf{G}$ is the Cayley transform of $\mathbf{B}$:

$$\mathbf{G} = (\mathbf{I} - \mathbf{B})(\mathbf{I} + \mathbf{B})^{-1} \tag{9}$$

Then:

$$\Psi = (\mathbf{I} + \mathbf{B})^{-T}(\mathbf{I} - \mathbf{B})^T\Psi(\mathbf{I} - \mathbf{B})(\mathbf{I} + \mathbf{B})^{-1} \tag{10}$$

Or, equivalently:

$$(\mathbf{I} + \mathbf{B})^T\Psi(\mathbf{I} + \mathbf{B}) = (\mathbf{I} - \mathbf{B})^T\Psi(\mathbf{I} - \mathbf{B}) \tag{11}$$



978-1-4673-6964-0/15/$31.00 ©2015 IEEE

Thus, we obtain $\mathbf{B}^T\Psi + \Psi\mathbf{B} = 0$, which means $\Psi\mathbf{B}$ is an antisymmetric matrix. That is, there exists $\mathbf{W} \in \mathbb{A}$ with $\Psi\mathbf{B} = \mathbf{W}$, thus $\mathbf{B} = \Psi^{-1}\mathbf{W}$, and

$$\begin{aligned}\mathbf{G} &= (\mathbf{I} - \Psi^{-1}\mathbf{W})(\mathbf{I} + \Psi^{-1}\mathbf{W})^{-1} \\ &= (\mathbf{I} + \Psi^{-1}\mathbf{W})^{-1}(\mathbf{I} - \Psi^{-1}\mathbf{W}) \\ &= (\Psi + \mathbf{W})^{-1}(\Psi - \mathbf{W})\end{aligned} \tag{12}$$

That is

$$\mathbf{G} = (\Psi + \mathbf{W})^{-1}(\Psi - \mathbf{W}) \qquad \square$$

**Proposition 3**: The Cayley-Klein metrics $d_E(\mathbf{x}_i, \mathbf{x}_j)$, $d_H(\mathbf{x}_i, \mathbf{x}_j)$ and Mahalanobis metric $d_\Sigma(\mathbf{x}_i, \mathbf{x}_j)$ have the following relationship:

$$\lim_{k\to+\infty} d_E(\mathbf{x}_i, \mathbf{x}_j) = d_\Sigma(\mathbf{x}_i, \mathbf{x}_j) = \lim_{k\to+\infty} d_H(\mathbf{x}_i, \mathbf{x}_j) \tag{13}$$

***Proof.*** First we prove

$$\lim_{k\to+\infty} d_E(\mathbf{x}_i, \mathbf{x}_j) = d_\Sigma(\mathbf{x}_i, \mathbf{x}_j)$$

Since $\mathbf{G}^+$ is a positive definite matrix, we have:

$$\sigma^{+2}_{\mathbf{x}_i\mathbf{x}_j} \le \sigma^+_{\mathbf{x}_i\mathbf{x}_i} \cdot \sigma^+_{\mathbf{x}_j\mathbf{x}_j} \tag{14}$$

According to the definition of $d_E(\mathbf{x}_i, \mathbf{x}_j)$, we have:

$$\begin{aligned}\frac{d_E(\mathbf{x}_i, \mathbf{x}_j)}{k} &= \frac{1}{2i}\log\left(\frac{\sigma^+_{\mathbf{x}_i\mathbf{x}_j} + i\sqrt{\sigma^+_{\mathbf{x}_i\mathbf{x}_i}\cdot\sigma^+_{\mathbf{x}_j\mathbf{x}_j} - \sigma^{+2}_{\mathbf{x}_i\mathbf{x}_j}}}{\sigma^+_{\mathbf{x}_i\mathbf{x}_j} - i\sqrt{\sigma^+_{\mathbf{x}_i\mathbf{x}_i}\cdot\sigma^+_{\mathbf{x}_j\mathbf{x}_j} - \sigma^{+2}_{\mathbf{x}_i\mathbf{x}_j}}}\right) \\ &= \frac{1}{2i}\log\left(\frac{\sigma^+_{\mathbf{x}_i\mathbf{x}_j} + i\sqrt{\sigma^+_{\mathbf{x}_i\mathbf{x}_i}\cdot\sigma^+_{\mathbf{x}_j\mathbf{x}_j} - \sigma^{+2}_{\mathbf{x}_i\mathbf{x}_j}}}{\sqrt{\sigma^+_{\mathbf{x}_i\mathbf{x}_i}\cdot\sigma^+_{\mathbf{x}_j\mathbf{x}_j}}}\right)^2 \\ &= \frac{1}{i}\log\left(\frac{\sigma^+_{\mathbf{x}_i\mathbf{x}_j} + i\sqrt{\sigma^+_{\mathbf{x}_i\mathbf{x}_i}\cdot\sigma^+_{\mathbf{x}_j\mathbf{x}_j} - \sigma^{+2}_{\mathbf{x}_i\mathbf{x}_j}}}{\sqrt{\sigma^+_{\mathbf{x}_i\mathbf{x}_i}\cdot\sigma^+_{\mathbf{x}_j\mathbf{x}_j}}}\right)\end{aligned} \tag{15}$$

According to Euler formula:

$$\cos\left(\frac{d_E(\mathbf{x}_i, \mathbf{x}_j)}{k}\right) = \frac{\sigma^+_{\mathbf{x}_i\mathbf{x}_j}}{\sqrt{\sigma^+_{\mathbf{x}_i\mathbf{x}_i}\cdot\sigma^+_{\mathbf{x}_j\mathbf{x}_j}}} \tag{16}$$

$$\sin\left(\frac{d_E(\mathbf{x}_i, \mathbf{x}_j)}{k}\right) = \sqrt{\frac{\sigma^+_{\mathbf{x}_i\mathbf{x}_i}\cdot\sigma^+_{\mathbf{x}_j\mathbf{x}_j} - \sigma^{+2}_{\mathbf{x}_i\mathbf{x}_j}}{\sigma^+_{\mathbf{x}_i\mathbf{x}_i}\cdot\sigma^+_{\mathbf{x}_j\mathbf{x}_j}}} \tag{17}$$

Let $\tilde{\mathbf{x}}_i = \mathbf{x}_i - \mathbf{m}$ and $\tilde{\mathbf{x}}_j = \mathbf{x}_j - \mathbf{m}$, then

$$\sin\left(\frac{d_E(\mathbf{x}_i, \mathbf{x}_j)}{k}\right) = \sqrt{\frac{(\tilde{\mathbf{x}}_i^T\Sigma\tilde{\mathbf{x}}_i + k^2)\cdot(\tilde{\mathbf{x}}_j^T\Sigma\tilde{\mathbf{x}}_j + k^2) - (\tilde{\mathbf{x}}_i^T\Sigma\tilde{\mathbf{x}}_j + k^2)^2}{(\tilde{\mathbf{x}}_i^T\Sigma\tilde{\mathbf{x}}_i + k^2)\cdot(\tilde{\mathbf{x}}_j^T\Sigma\tilde{\mathbf{x}}_j + k^2)}} \tag{18}$$

For the squared Mahalanobis distance, we can rewrite it as:

$$\begin{aligned}d^2_\Sigma(\mathbf{x}_i, \mathbf{x}_j) &= (\mathbf{x}_i - \mathbf{x}_j)^T\Sigma(\mathbf{x}_i - \mathbf{x}_j) \\ &= (\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j)^T\Sigma(\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j) \\ &= \tilde{\mathbf{x}}_i^T\Sigma\tilde{\mathbf{x}}_i + \tilde{\mathbf{x}}_j^T\Sigma\tilde{\mathbf{x}}_j - 2\tilde{\mathbf{x}}_i^T\Sigma\tilde{\mathbf{x}}_j\end{aligned} \tag{19}$$

Combining Eq. (19) and Eq. (18), we have:

$$\sin\left(\frac{d_E(\mathbf{x}_i, \mathbf{x}_j)}{k}\right) = \sqrt{\frac{k^2\cdot d^2_\Sigma(\mathbf{x}_i, \mathbf{x}_j) + \left(\tilde{\mathbf{x}}_i^T\Sigma\tilde{\mathbf{x}}_i\cdot\tilde{\mathbf{x}}_j^T\Sigma\tilde{\mathbf{x}}_j - (\tilde{\mathbf{x}}_i^T\Sigma\tilde{\mathbf{x}}_j)^2\right)}{(\tilde{\mathbf{x}}_i^T\Sigma\tilde{\mathbf{x}}_i + k^2)\cdot(\tilde{\mathbf{x}}_j^T\Sigma\tilde{\mathbf{x}}_j + k^2)}} \tag{20}$$

then

$$\lim_{k\to+\infty}\frac{\sin\left(\frac{d_E(\mathbf{x}_i,\mathbf{x}_j)}{k}\right)}{\frac{d_\Sigma(\mathbf{x}_i,\mathbf{x}_j)}{k}} = \lim_{k\to+\infty}\sqrt{\frac{1 + \frac{\tilde{\mathbf{x}}_i^T\Sigma\tilde{\mathbf{x}}_i\cdot\tilde{\mathbf{x}}_j^T\Sigma\tilde{\mathbf{x}}_j - (\tilde{\mathbf{x}}_i^T\Sigma\tilde{\mathbf{x}}_j)^2}{k^2\cdot d^2_\Sigma(\mathbf{x}_i,\mathbf{x}_j)}}{\left(1 + \frac{\tilde{\mathbf{x}}_i^T\Sigma\tilde{\mathbf{x}}_i}{k^2}\right)\cdot\left(1 + \frac{\tilde{\mathbf{x}}_j^T\Sigma\tilde{\mathbf{x}}_j}{k^2}\right)}} = 1 \tag{21}$$

Therefore, we obtain

$$\lim_{k\to+\infty} d_E(\mathbf{x}_i, \mathbf{x}_j) = d_\Sigma(\mathbf{x}_i, \mathbf{x}_j)$$

For the hyperbolic Mahalanobis metric, in a similar way we can obtain:

$$\sinh\left(\frac{d_H(\mathbf{x}_i, \mathbf{x}_j)}{k}\right) = \sqrt{\frac{k^2\cdot d^2_\Sigma(\mathbf{x}_i, \mathbf{x}_j) - \left(\tilde{\mathbf{x}}_i^T\Sigma\tilde{\mathbf{x}}_i\cdot\tilde{\mathbf{x}}_j^T\Sigma\tilde{\mathbf{x}}_j - (\tilde{\mathbf{x}}_i^T\Sigma\tilde{\mathbf{x}}_j)^2\right)}{(\tilde{\mathbf{x}}_i^T\Sigma\tilde{\mathbf{x}}_i - k^2)\cdot(\tilde{\mathbf{x}}_j^T\Sigma\tilde{\mathbf{x}}_j - k^2)}} \tag{22}$$

where $\sinh(x)$ is a hyperbolic sine function as:

$$\sinh(x) = \frac{e^x - e^{-x}}{2} \tag{23}$$

Thus

$$\lim_{k\to+\infty}\frac{\sinh\left(\frac{d_H(\mathbf{x}_i,\mathbf{x}_j)}{k}\right)}{\frac{d_\Sigma(\mathbf{x}_i,\mathbf{x}_j)}{k}} = \lim_{k\to+\infty}\sqrt{\frac{1 - \frac{\tilde{\mathbf{x}}_i^T\Sigma\tilde{\mathbf{x}}_i\cdot\tilde{\mathbf{x}}_j^T\Sigma\tilde{\mathbf{x}}_j - (\tilde{\mathbf{x}}_i^T\Sigma\tilde{\mathbf{x}}_j)^2}{k^2\cdot d^2_\Sigma(\mathbf{x}_i,\mathbf{x}_j)}}{\left(1 - \frac{\tilde{\mathbf{x}}_i^T\Sigma\tilde{\mathbf{x}}_i}{k^2}\right)\cdot\left(1 - \frac{\tilde{\mathbf{x}}_j^T\Sigma\tilde{\mathbf{x}}_j}{k^2}\right)}} = 1 \tag{24}$$

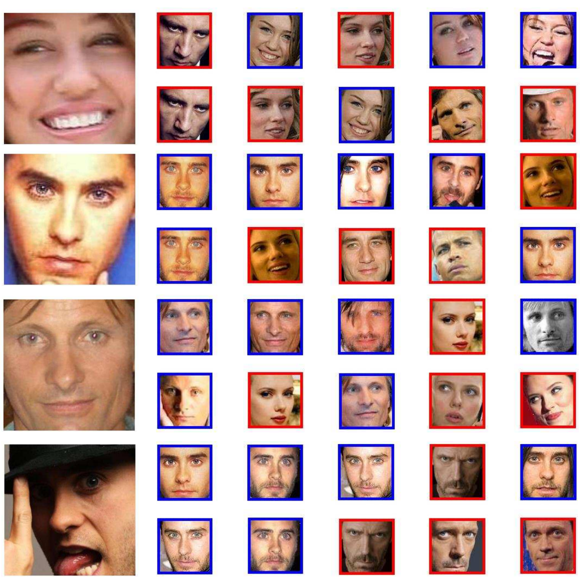
from which we can obtain

$$\lim_{k\to+\infty} d_H(\mathbf{x}_i, \mathbf{x}_j) = d_\Sigma(\mathbf{x}_i, \mathbf{x}_j) \qquad \square$$

## 2. Experimental results of similarity search

Figure 1 presents recognition results of CK-MMC and MMC on OSR and PubFig. We show for each query the 5 most similar images using the metric learned by CK-MMC (first row) and MMC (second row) respectively.

(a) OSR

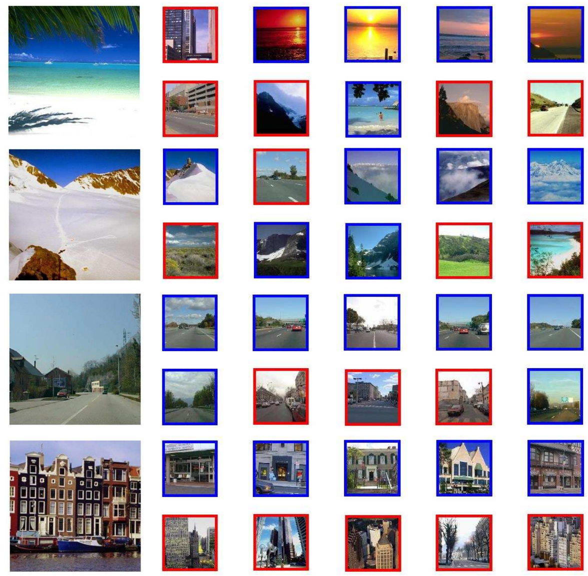


(b) PubFig

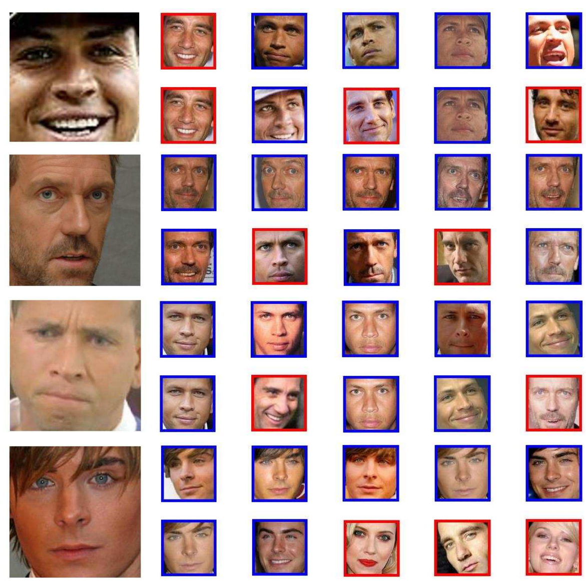
Figure 1. Similarity search results of CK-MMC and MMC on OSR and Pubfig. For each query from the OSR and PubFig datasets, we return the 5 most nearest neighbors according to the metric learned by CK-MMC (first row) and MMC (second row) respectively. The images with the same class of query image are marked in blue, and those from different classes are marked in red.

Figure 2 shows recognition results of CK-LMNN (first row) and LMNN (second row) on OSR and PubFig.

It is clear that our methods (CK-MMC and CK-LMNN) could return more semantically relevant images. Therefore, Cayley-Klein metric is more reliable than Mahalanobis metric.



(a) OSR



(b) PubFig

Figure 2. Similarity search results of CK-LMNN and LMNN on OSR and Pubfig. For each query from the OSR and PubFig datasets, we return the 5 most nearest neighbors according to the metric learned by CK-LMNN (first row) and LMNN (second row) respectively. The images with the same class of query image are marked in blue, and those from different classes are marked in red.