Beyond Mahalanobis Metric: Cayley-Klein Metric Learning

Yanhong Bi, Bin Fan, Fuchao Wu
Institute of Automation, Chinese Academy of Sciences.

Distance metric plays an important role in many computer vision and pattern recognition tasks. Metric learning aims to learn a specific distance function for a particular task, and is proven to be very useful when dealing with problems that rely on distances. In the Mahalanobis metric learning framework, the central task is to learn a positive semidefinite matrix $\mathbf{M}$ to fit the squared Mahalanobis distance $d^2(\mathbf{x},\mathbf{y}) = (\mathbf{x}-\mathbf{y})^T\mathbf{M}(\mathbf{x}-\mathbf{y})$. Basically, Mahalanobis metric can be viewed as Euclidean metric on a global linear transformed input space. How to estimate such linear transformations on the input space is at core of Mahalanobis metric learning, which aims to obtain a distance metric better modeling the underlying relationship among input data [2, 3]. While except for learning a positive semidefinite matrix for Mahalanobis metric, few attempts have been made for a proper non-Euclidean metric.

In this paper, we present a non-Euclidean metric beyond Mahalanobis framework. The core idea lies in a novel distance metric defined based on the non-Euclidean geometry discovered by A. Cayley and F. Klein in the 19th century. The so called Cayley-Klein metric, induced by an invertible symmetric matrix, is a metric in projective space defined using a cross-ratio. Compared to existing metrics, Cayley-Klein metric has several advantages. First, the Cayley-Klein space is a special case of Riemannian space with fixed curvature. Second, the Cayley-Klein metric has an explicit definition while a general Riemannian metric does not have. Finally, it is a generalization of Mahalanobis metric by extending the metric definition based on a linear transformation to a fractional transformation. As a result, Cayley-Klein metric is more general compared to Euclidean and Mahalanobis metrics.

Given an invertible symmetric matrix $\Psi \in \mathbb{R}^{(n+1)\times(n+1)}$, its bilinear representation of $\mathbf{x},\mathbf{y} \in \mathbb{R}^n$ can be denoted by $\psi(\mathbf{x},\mathbf{y})$:

$$\psi(\mathbf{x},\mathbf{y}) = (\mathbf{x}^T,1)\Psi\begin{pmatrix}\mathbf{y}\\1\end{pmatrix} \triangleq \psi_{\mathbf{xy}},\ \forall \mathbf{x},\mathbf{y} \in \mathbb{R}^n \quad (1)$$

If matrix Ψ is positive definite, then $\psi_{\mathbf{xx}} > 0$, we can define $\rho_E(\mathbf{x},\mathbf{y}) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^+$ as:

$$\rho_E(\mathbf{x},\mathbf{y}) = \frac{k}{2i}\log\left(\frac{\psi_{\mathbf{xy}} + \sqrt{\psi_{\mathbf{xy}}^2 - \psi_{\mathbf{xx}}\psi_{\mathbf{yy}}}}{\psi_{\mathbf{xy}} - \sqrt{\psi_{\mathbf{xy}}^2 - \psi_{\mathbf{xx}}\psi_{\mathbf{yy}}}}\right)(k>0) \quad (2)$$

If matrix Ψ is indefinite, set $\mathbb{B}^n = \{\mathbf{x} \in \mathbb{R}^n | \psi_{\mathbf{xx}} < 0\}$, we can define $\rho_H(\mathbf{x},\mathbf{y}) : \mathbb{B}^n \times \mathbb{B}^n \to \mathbb{R}^+$ as:

$$\rho_H(\mathbf{x},\mathbf{y}) = -\frac{k}{2}\log\left(\frac{\psi_{\mathbf{xy}} + \sqrt{\psi_{\mathbf{xy}}^2 - \psi_{\mathbf{xx}}\psi_{\mathbf{yy}}}}{\psi_{\mathbf{xy}} - \sqrt{\psi_{\mathbf{xy}}^2 - \psi_{\mathbf{xx}}\psi_{\mathbf{yy}}}}\right)(k>0) \quad (3)$$

$(\mathbb{R}^n,\rho_E)$ is called elliptic geometry space and $(\mathbb{B}^n,\rho_H)$ is called hyperbolic geometry space. ρ_E and ρ_H together constitute the Cayley-Klein metric.

Based on the Cayley-Klein metric, we propose a special form of it which we call *generalized Mahalanobis* metric since it approaches Mahalanobis metric in an extreme case. Firstly, we define two reversible symmetric matrices $\mathbf{G}^{\pm}$ as:

$$\mathbf{G}^{\pm} = \begin{pmatrix}\Sigma & -\Sigma\mathbf{m}\\ -\mathbf{m}^T\Sigma & \mathbf{m}^T\Sigma\mathbf{m} \pm k^2\end{pmatrix}(k>0) \quad (4)$$

where $\mathbf{m}$ and Σ are mean and inverse covariance of a set of data points. $\mathbf{G}^+$ is positive definite while $\mathbf{G}^-$ is indefinite. Then, according to Eq. (1)(2)(3), we can obtain two specific Cayley-Klein metrics $d_E(\mathbf{x}_i,\mathbf{x}_j)$ and $d_H(\mathbf{x}_i,\mathbf{x}_j)$.

Figure 1 illustrates the difference between Cayley-Klein metric and Mahalanobis metric in 2-dimensional space. Under Mahalanobis metric, the equidistant distribution of a fixed point is an ellipse and stays unchanged when the fixed point changes its location. On the contrary, in case of Cayley-Klein metric, the shape and scale of this equidistant distribution would change, depending on the location of the fixed point.

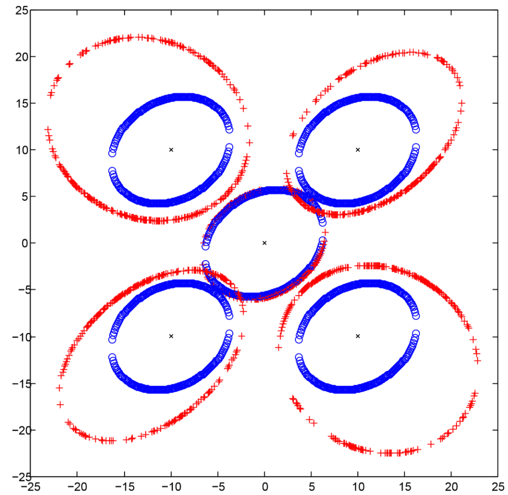

Figure 1: Equidistant distribution of five fixed points for Cayley-Klein metric and Mahalanobis metric. Under Mahalanobis metric (marked as "o" in blue), all points with unit distance to a fixed point (marked as "x" in black) form an ellipse, whose center is the fixed point. This ellipse is identical for any fixed point, wherever its location. On the contrary, under Cayley-Klein metric (marked as "+" in red), all points with unit distance to the origin form a shape similar to ellipse. However, this shape differs when the fixed point moves.

Method	OSR	PubFig
MMC [3]	64.0 ± 1.6	80.3 ± 1.0
CK-MMC	$\mathbf{67.1 \pm 1.6}$	$\mathbf{82.5 \pm 1.0}$
LMNN [2]	67.3 ± 1.3	78.8 ± 1.4
GB-LMNN [1]	69.3 ± 1.3	79.9 ± 1.6
CK-LMNN	$\mathbf{70.1 \pm 1.1}$	$\mathbf{81.3 \pm 1.7}$

Table 1: Classification accuracies (mean and standard deviation in %) obtained on OSR and PubFig. CK-MMC and CK-LMNN have a clear improvement over MMC and LMNN respectively.

In metric learning, two of the most typical methods are MMC [3] and LMNN [2]. In the paper, we integrate Cayley-Klein metric into these typical supervised metric learning paradigms to obtain two Cayley-Klein metric learning methods, which we called CK-MMC and CK-LMNN.

Implementation of these two methods are described in the paper, with details on optimization objectives and solvers. Our conclusion is that the proposed two Cayley-Klein metric learning methods can use labeled training data to learn an appropriate metric for a given task. Experiments on image classification have shown that better performance can be obtained by using Cayley-Klein metric.

[1] Dor Kedem, Stephen Tyree, Kilian Q. Weinberger, and Fei Sha. Nonlinear metric learning. In *NIPS*, 2012.

[2] K. Weinberger and L. Saul. Distance metric learning for large margin nearest neighbor classification. *Journal of Machine Learning Research*, 10:207–244, 2009.

[3] E. Xing, A. Ng, M. Jordan, and S. Russell. Distance metric learning, with application to clustering with side-information. In *NIPS*, pages 505–512, 2002.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.