Machine Learning with Cayley-Klein metrics*

Laëtitia Shao Frank Nielsen†

September 2017

Abstract

The Cayley-Klein metrics have been recently introduced as a generalization of the Mahalanobis distances used in metric learning. A prior work [1] experimented with supervised classification using these Cayley-Klein metrics, and showed improved results over Mahalanobis distances. In this paper, we first study Hilbert metrics, which is a generalization of Cayley-Klein metrics. We explain the shape of balls and their properties in a polygonal Hilbert geometry. We show how the shapes of Hilbert polygonal balls depend on the center location and on the domain, but not on their radii and give an explicit description of the Hilbert ball for any center and radius. Then we study the intersection of two Hilbert balls, considering the cases of empty intersection and tangency. Last, we outline some possible applications of Cayley-Klein metrics in Machine Learning, describing how they have been used in metric-dependent algorithms such as the k-Nearest Neighbours (k-NN) and the Stochastic Neighbouring Embedding algorithms (t-SNE).

1 Introduction to Hilbert geometries

1.1 Cayley-Klein geometries

Cayley-Klein geometry is a geometry relying on the properties of the cross-ratio.

1.1.1 Cross-Ratio

Definition 1 (Cross-Ratio) *For four collinear points a, b, c, d, the cross-ratio is defined as follows:*

$$(a, b; c, d) = \frac{\|c - a\|\|b - d\|}{\|d - a\|\|c - b\|} \tag{1}$$

*A multimedia presentation including a video `https://www.youtube.com/watch?v=XE5x5rAK8Hk&t=88s` and interactive demos (with autoplay mode) has been presented in [7].See also `https://franknielsen.github.io/HSG/`

†E-mail: `Frank.Nielsen@acm.org`

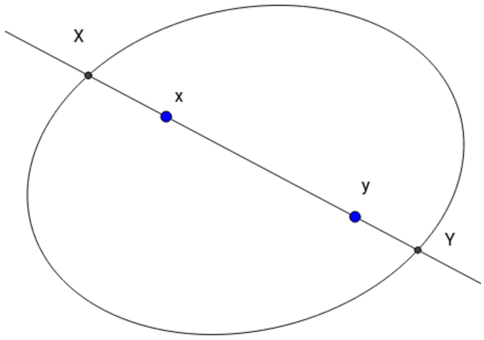

Figure 1: Distance in an elliptic Cayley-Klein geometry

The cross-ratio is an invariant measure under perspective transformation:

Property 1 (Projective invariance of the cross-ratio [10]) *Given four points a, b, c, d, define A, B, C, D their images through a projective transformation. The follow equality holds.*

$$(a, b; c, d) = (A, B; C, D) \tag{2}$$

1.1.2 Cayley-Klein geometry

In a projective space, a Cayley-Klein geometry is defined by a triple (F, c_{dist}, c_{angle}) where F is a fundamental conic and c_{dist} and c_{angle} are constant units for measuring distances and angles in the geometry.

In a Cayley-Klein geometry, the distance is defined by:

$$d(x, y) = c_{dist} \log\left((x, y; X, Y)\right) \tag{3}$$

where the points are defined as in figure 1 and $(x, y; X, Y)$ is the cross-ratio defined in equation 1.

1.2 Hilbert metrics

Hilbert geometries generalize Cayley-Klein geometries. Instead of being defined on a fundamental conic, Hilbert geometries are defined on a convex domain.

Definition 2 (Hilbert distance) *A Hilbert distance is defined in the interior of a convex bounded domain. Given two points, a and b of the domain, the distance is defined as follows:*

$$d_{HG}(a, b) = \log\left((a, b; A, B)\right) \tag{4}$$

$(a, b; A, B)$ is the cross-ratio where A and B denote the intersection points of line (a, b) with the domain. See figure 2.

Property 2 (Properties of the Hilbert distance) *Given two points a and b.*

- *The Hilbert distance is a signed distance: $d_{HG}(a, b) = -d_{HG}(b, a)$*

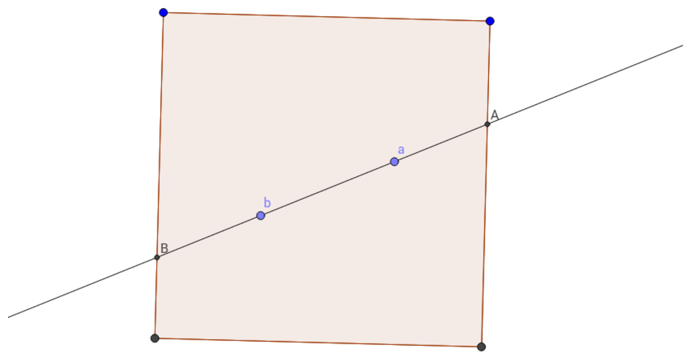

Figure 2: Metric in a Hilbert geometry defined by a square: the distance between points a and b is defined as $log((a, b; A, B))$

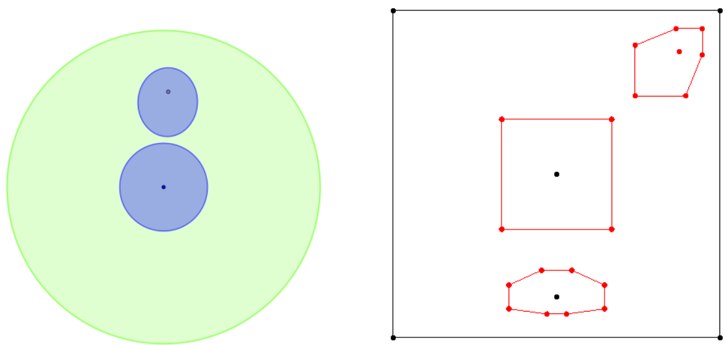

Figure 3: Left: In blue, two Hilbert balls in a circular domain. Right: In white, three Hilbert balls in a polygonal convex domain.

- $d_{HG}(a, a) = 0$ *(law of the indiscernibles).*
- *When a is on the boundary of the convex, $d_{HG}(a, b) = +\infty$.*
- $|d_{HG}|$ *respects the triangular inequality and therefore* $|d_{HG}|$ *is a metric distance.[5]*

Notations We will note $\mathcal{C}$ the convex domain. We first study Hilbert geometries when $\mathcal{C}$ is a simplex and note $e_1, ..e_s$ the s vertices of the convex.
The distance between two points p and q is noted $d_{\mathcal{C}}(p, q)$.
The ball of radius r and center c is noted $\mathcal{B}(c, r)$.

2 Properties of Hilbert balls

Having introduced Hilbert geometries and defined the metric, we now show some properties of balls in polygonal Hilbert geometries. Figure 3 shows that, contrasting with smooth balls that appear in smooth convex domains, Hilbert balls of a polygonal domain are polygonal with different shapes.

De La Harpe [3] reported an isometric embedding of the Hilbert simplex geometry [8] to a polytopal normed space (see also [9]). Later, it was shown that the only Hilbert geometries that can be isometrically embedded into a normed vector space are the polytopal Hilbert geometries [11].

2.1 Complexity of a Hilbert ball

Lemma 1 (Description of a Hilbert ball) *$\mathcal{B}(c,r)$ is a Euclidean polygon with at most $2s$ edges and at least s edges. Each vertex of $\mathcal{B}(c,r)$ lies on a line $(c, ei), i \in [s] = 1, ..., s$.*

Proof 1

Given a point c lying in the convex domain, we first partition the polygonal domain with s to $2s$ triangles, by tracing rays $(c, e_i), i \in [s]$. We will show that each triangle induces a linear edge of the Hilbert ball.

We consider a pair of triangles (A, B, c) and (c, C, D) such that A, c, D and B, c, C are respectively collinear. Let $P \in [A, c] \cap \mathcal{B}(c, r)$ and $O = (A, B) \cap (C, D)$, we will show that line (O, P) clipped to the triangle ABc is an edge of $\mathcal{B}(c, r)$.

Let U be a point on the clipped line, and M, N the intersections points of line (U, c) with the domain such that $M \in [A, B]$ and $N \in [C, D]$. Then M, U, c, N and A, P, c, D are related by the same projective transformation. (Refer to Figure 4 for an illustration of the setup).

Using the invariance property of the cross-ratio, we conclude that:

$$d_{\mathcal{C}}(c, P) = d_{\mathcal{C}}(c, U) = r$$

Thus, we proved Lemma 1. It is remarkable that depending on the position of the center, the number of triangles (and hence the complexity of the ball) varies.

This explains how the complexity of a Hilbert ball depends not only on the convex domain it lies in but also on the position of the center c, as shown in Figure 5. When the center c is placed in a way that two rays become one- this typically happens on the diagonals of the domain- the number of edges decreases. Moreover, as triangles are always paired, the number of edges is always even.

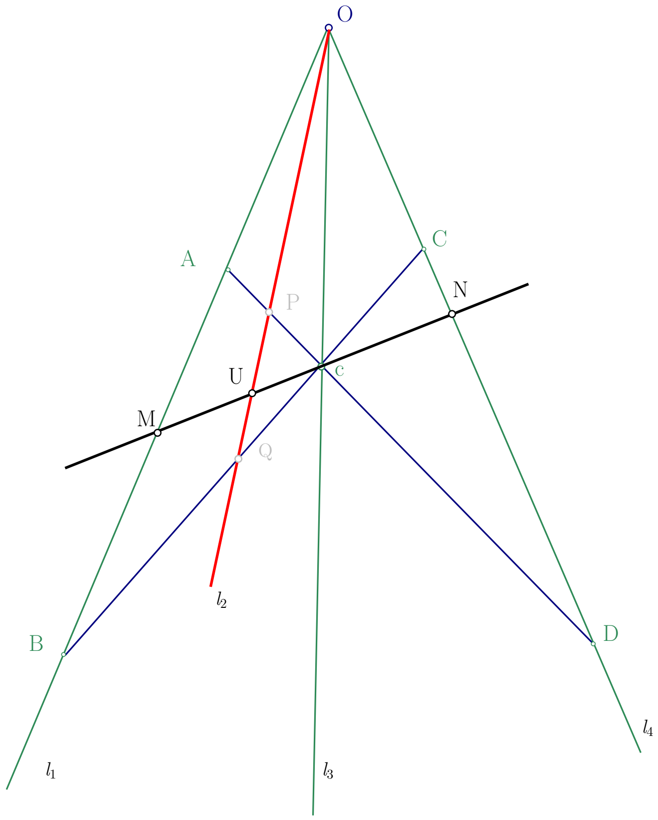

Figure 4: Configuration for proof of lemma 1

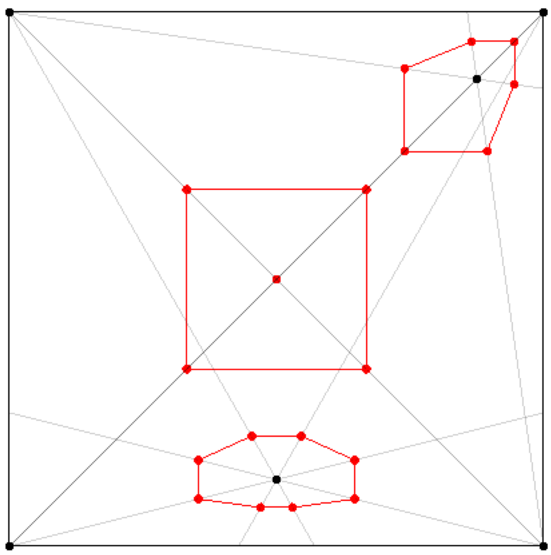

Figure 5: Unit balls in a square domain and with different number of edges. A unit ball typically has 8 edges (bottom ball), but when we move the center in the domain, critical configurations appear. At the center of the square the unit ball has 4 edges. On a diagonal, unit balls have 6 edges.

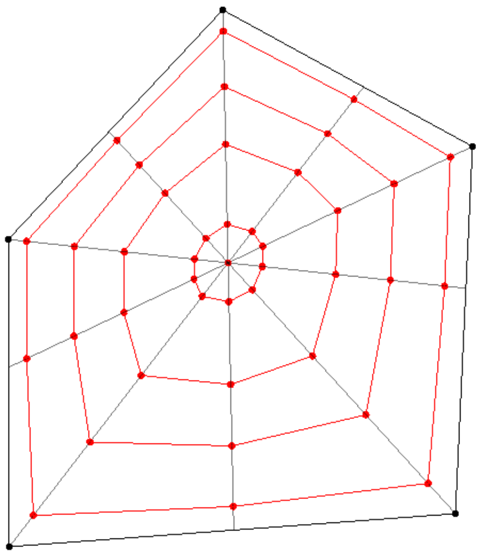

Figure 6: Several balls with the same center and an increasing radius. The bigger the radius is, the more the ball's shape covers the full domain. The black lines represent the rays passing through the center of the ball and a vertex of the domain as discussed in the previous subsection.

2.2 Shape changes of a Hilbert ball with radius

We observe that as the radius tends toward infinity, the Hilbert ball's shape covers the full domain, as seen in figure 6. As we showed previously, a Hilbert ball has at most $2s$ linear edges and as the radius tends towards infinity, the Hilbert ball tends to have s linear edges. We are therefore interested in finding a correlation between the number of edges and the radius of a Hilbert ball.

We first study the simple case of triangular domains and show that in such domains, Hilbert balls have constant shapes.

Lemma 2 (Property of Hilbert ball edges in a triangular domain) *If $\mathcal{B}$ is a Hilbert ball in a triangular domain (A, B, C), then for each edge (P, Q) of $\mathcal{B}$, either A, B or C is collinear with P and Q. -See Figure 7.*

Proof 2 *The lemma is easily proven with the property of the cross-ratio used to prove Lemma 1.*

Property 3 (Invariance of Hilbert ball shapes in a triangular domain) *In a triangular domain:*

$$\forall r > 0, \forall c \in \mathcal{C} : \mathcal{B}(c, r) \text{ has exactly 6 edges}$$

Proof 3

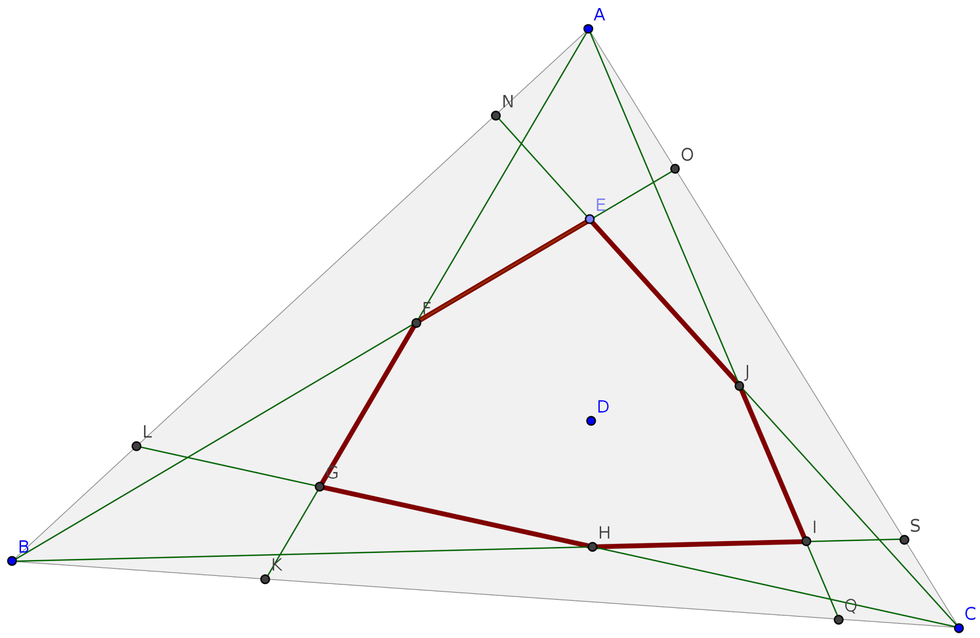

Figure 7: A Hilbert Ball in a triangular domain: the extension of each edge intersects one vertex of the domain.

We use the previous lemma to prove the property. Consider c the center of a Hilbert Ball in a triangle A, B, C. We consider the two edges of $\mathcal{B}(c, r)$: (P, Q) and (Q, R) that are closer to edge (A, B) (Figure 8).

Then P, Q, B are collinear and A, Q, R too. So, P, Q, R are collinear, if and only if, all points belong to $[A, B]$ ie. when $r = +\infty$.

Thus, for any $r > 0$, there are always two distinct edges induced by (A, B).

By extending this reasoning to the two other edges of the triangle we prove the property.

Definition 3 *Given an edge $[P, Q]$ of a Hilbert ball that belongs to a pair of triangles (A, B, c) and (c, D, E), we say that $[P, Q]$ is induced by edges E_i and E_j of the domain, if $[A, B] \subset E_i$ and $[D, E] \subset E_j$.*

Property 4 (Shape invariance with varying radius) *For c a fixed center point, and r a varying radius, $\mathcal{B}(c, r)$ has the same number of edges.*

Proof 4

Let $[P, Q]$ and $[Q, R]$ be two adjacent edges of a Hilbert ball such that E_i and E_j induce $[P, Q]$ and E_k and E_l induce $[Q, R]$. We show that P, Q, R cannot be collinear.

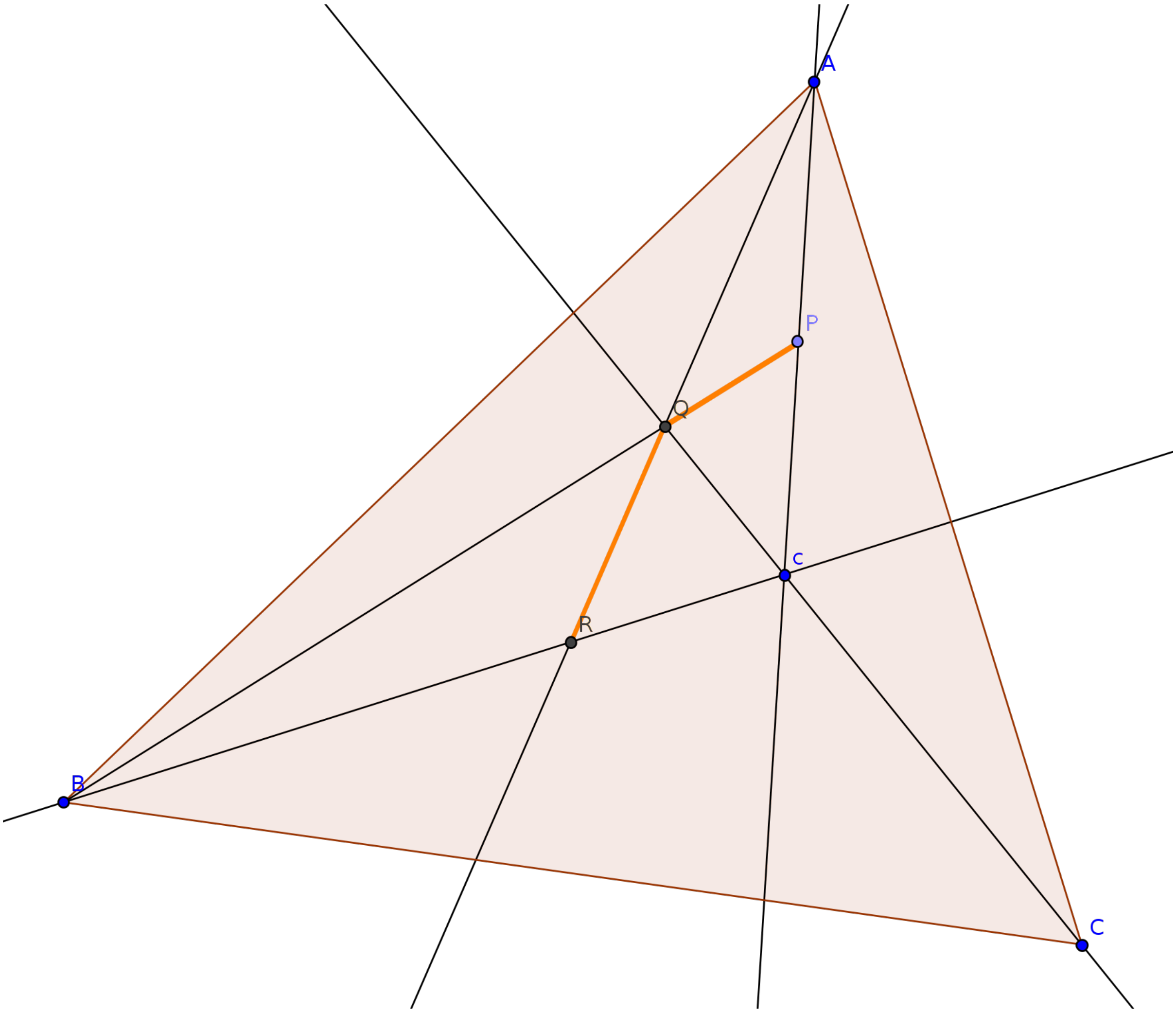

Figure 8: In triangle A, B, C we consider the two edges (P, Q) and (Q, R) induced by (A, B) of a Hilbert Ball of center c.

We note M the intersection points of the lines supported by E_i and E_j and N the intersection points of the lines supported by E_k and E_l. According to the previous proof, P, Q, M and Q, R, N are respectively collinear. We proceed by tackling two cases:

- *Case 1 E_i, E_j, E_k, E_l are distinct edges. Because $[P, Q]$ and $[Q, R]$ are adjacent, we can assume without loss of generality that E_i is adjacent to E_k and E_j is adjacent to E_l. Assuming that P, Q, R are collinear, then $E_i = E_k$ or $E_j = E_l$, which contradicts the previous assumption.*

- *Case 2, we can assume that $E_i = E_k$. In this case, assuming that P, Q, R are collinear, then they belong to line $(M, N) \subset E_i$. This is impossible unless $r = +\infty$.*

Therefore, as the radius varies but stay finite, the number of edges remains constant. See Figure 10 for a visualization of the proof. For infinite radius, all balls fully cover the polygonal domain.

2.3 Intersection of two Hilbert balls

We note that, in some configurations, Hilbert balls can share an infinite number of points without being equal, as seen in figure 10. The goal of this section is to characterize the intersection of two given Hilbert balls.

Lemma 3 (Condition for empty intersection) *Given two points $c_1, c_2 \in \mathcal{C}$ and two reals $r_1, r_2 > 0$, with $r_2 \geq r_1$:*

$$d_{\mathcal{C}}(c_1, c_2) < |r_1 - r_2| \vee d_{\mathcal{C}}(c_1, c_2) > r_1 + r_2 \implies \mathcal{B}(c_1, r_1) \cap \mathcal{B}(c_2, r_2) = \emptyset \tag{5}$$

Proof 5 *This follows straightforwardly from the triangular inequality. [5]*

Lemma 4 (Inner tangency- Figure 11) *If $r_2 \geq r_1$ and $c_1 \in \mathcal{B}(c_2, r_2 - r_1)$, $\mathcal{B}(c_1, r_1)$ is innerly tangent to $\mathcal{B}(c_2, r_2)$. We distinguish two cases.*

- *if c_1 is a vertex of $\mathcal{B}(c_2, r_2 - r_1)$, the two balls share part of two edges.*

- *otherwise, the two balls share part of one edge.*

Lemma 5 (Outer tangency- Fig 12) *If $r_2 \geq r_1$ and $c_1 \in \mathcal{B}(c_2, r_2 + r_1)$, $\mathcal{B}(c_1, r_1)$ is externally tangent to $\mathcal{B}(c_2, r_2)$. We distinguish two cases:*

- *if c_1 is a vertex of $\mathcal{B}(c_2, r_2 + r_1)$, the two balls share one vertex.*

- *otherwise, the two balls share part of one edge.*

2.4 Cone intersection

Definition 4 (Cone intersection - Figure 13) *Given two points $p, q \in \mathcal{C}$, we define the following set:*

$$\mathcal{I}(p, q) = \{t \in \mathcal{C}, d_{\mathcal{C}}(p, t) + d_{\mathcal{C}}(t, q) = d_{\mathcal{C}}(p, q)\} \tag{6}$$

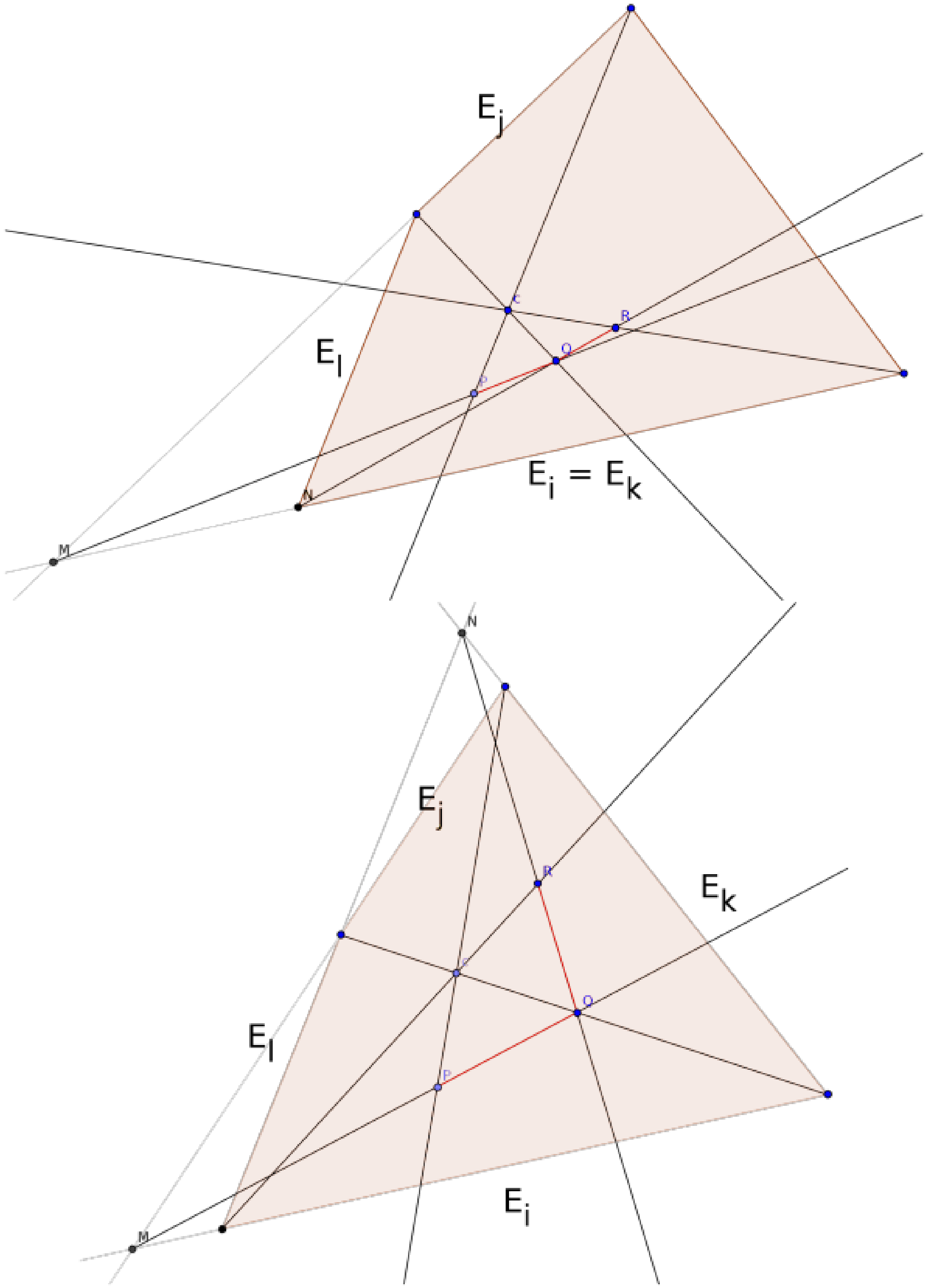

Figure 9: Configuration for proof 4. Top: Case where $E_i = E_k$. Bottom: Case where alre edges are distinct

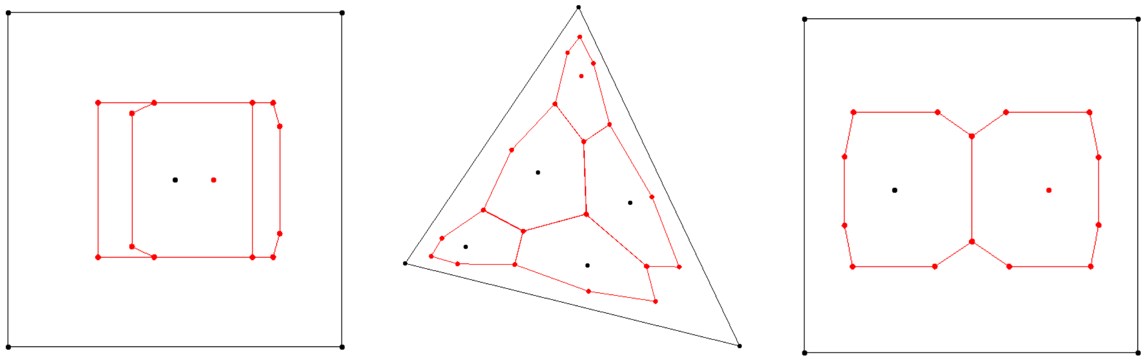

Figure 10: Unit balls in various domains that share part of an edge

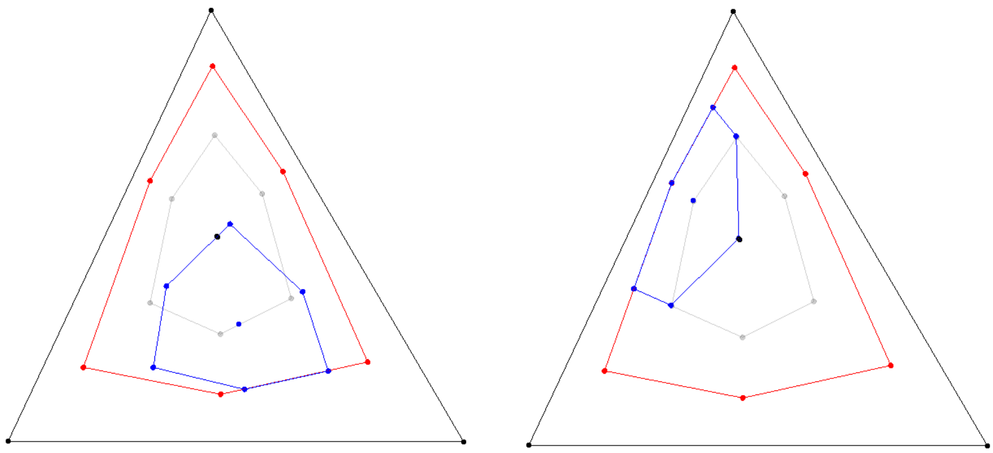

Figure 11: Two cases of inner tangency between the red ball and the blue ball. Left: The two balls share one edge. Right: the two balls share two edges.

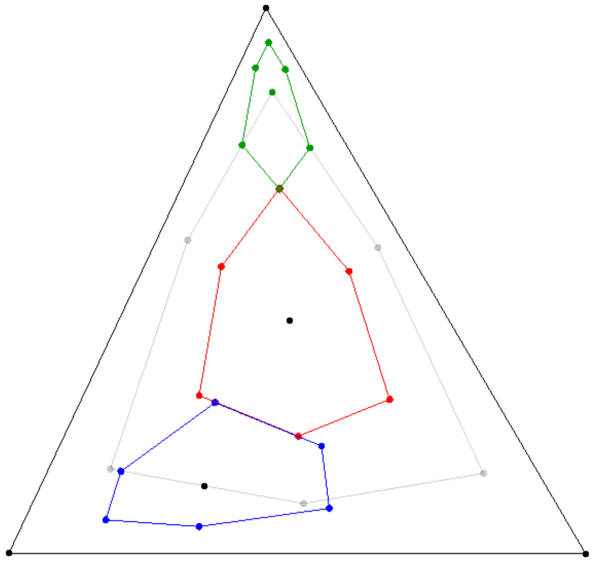

Figure 12: In a triangular domain, two cases of outer tangency. The blue ball and the red ball are tangent and share part of an edge. The red ball and the green ball share only one vertex.

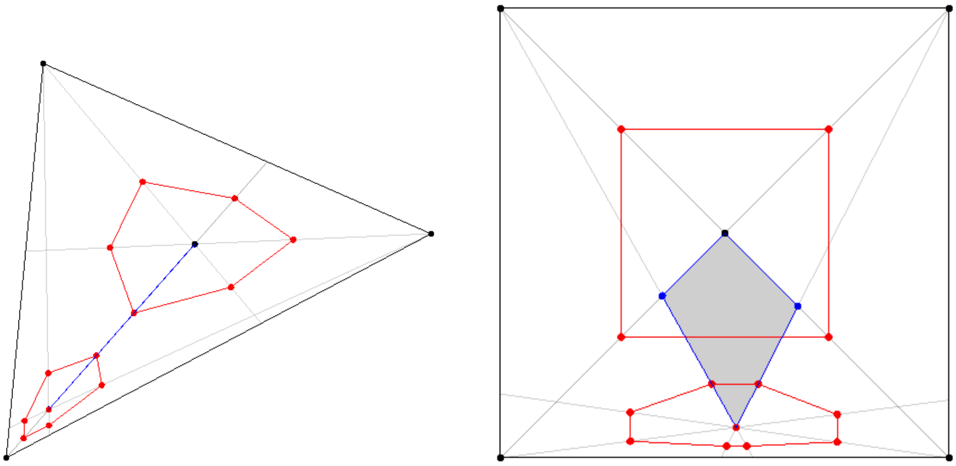

Figure 13: Visualization of cone intersection. Left: The intersection is reduced to a segment (blue). Right: The cone intersection is a quadrangle.

Using the previous claims on tangency between two Hilbert balls, it is possible to describe the complexity of the set $\mathcal{I}(p,q)$.

Property 5 (Complexity of cone intersection) *Given any convex simplex $\mathcal{C}$ and any two points $p, q \in \mathcal{C}$, the cone intersection $\mathcal{I}(p,q)$ is either the segment $[p,q]$, or a quadrangle. The complexity of the cone intersection depends on the positions of p and q in $\mathcal{C}$.*

More precisely there are two cases:

- *p, q are collinear with a vertex of $\mathcal{C}$. In that case, $\mathcal{I}(p,q)$ is reduced to a segment.*
- *Otherwise, p lies strictly in one of the triangles obtained by tracing rays from q. Similarly, q also lies in a triangle obtained by tracing rays from p. Let's note T_p and T_q such triangles. Then the cone intersection is $I(p,q) = T_p \cap T_q$.*

3 Applications to Machine Learning algorithms

We outline here some algorithms that could be generalized using Cayley-Klein or Hilbert metrics. We first talk about Cayley-Klein metric learning which was introduced in [10] then discuss how the Stochastic Neighbouring Embedding algorithm [4] could be adapted to Cayley-Klein metrics as well. For this we first remind that in elliptic and hyperbolic Cayley-Klein geometries, the distance has a closed expression, as shown in [10].

3.1 Distances in elliptic and hyperbolic Cayley-Klein geometries

In $\mathbb{RP}^d$, the projective space of dimension d, the fundamental conic can be represented by a $(d+1)\times(d+1)$ symmetric positive matrix S. Depending on the eigenvalues of S, there are 9 different types of Cayley-Klein geometries [10]. We'll focus on the elliptic and hyperbolic geometry.

We introduce the bilinear form associated to the matrix S.

$$S_{x,y} = \tilde{x}^T S \tilde{y} \tag{7}$$

where $\tilde{x}$ is the expression of x in homogeneous coordinates.

The cross-ratio can then be expressed as follows:

$$(x, y; X, Y) = \frac{S_{x,y} + \sqrt{S_{x,y}^2 - S_{x,x}S_{y,y}}}{S_{x,y} - \sqrt{S_{x,y}^2 - S_{x,x}S_{y,y}}} \tag{8}$$

Elliptic distance In order to have real-value distances, we chose $c_{dist} = \frac{\kappa}{2i}, \kappa \in \mathbb{R}$. The distance in an elliptic Cayley-Klein geometry can be defined as follows, according to [6].

$$d_E(x,y) = \frac{\kappa}{2i} log(\frac{S_{x,y} + \sqrt{S_{x,y}^2 - S_{x,x}S_{y,y}}}{S_{x,y} - \sqrt{S_{x,y}^2 - S_{x,x}S_{y,y}}}) \tag{9}$$

$$d_E(x,y) = \kappa \arccos(\frac{S_{x,y}}{\sqrt{S_{x,x}S_{y,y}}}) \tag{10}$$

Hyperbolic distance We here choose $c_{dist} = \frac{-\kappa}{2}$. Furthermore, the distance is only defined on the domain $\{p, S_{p,p} \not\succcurlyeq 0\}$. According to [6], the distance is then defined by the formula:

$$d_H(x,y) = \kappa \text{ arccosh} \left(\frac{S_{x,y}}{\sqrt{S_{x,x}S_{y,y}}} \right) \tag{11}$$

These expressions make it easier to use Cayley-Klein metrics in the Machine Learning algorithms presented below.

3.2 Metric learning with Cayley-Klein metrics

3.2.1 Metric learning for the $k-$NN algorithm

We consider the machine learning problem of supervised classification. Given a set of points with their labels, the problem amounts to tagging a new point according to the known data set. The classic method to solve this problem is to use the k-Nearest Neighbour classifier which associates to a point the dominant label of its k nearest neighbours. The k-NN classifier depends of the chosen metric. When this metric depends on the training set, we first have to learn the relevant metric. Learning the relevant topic is called metric learning.

Up until recently, the Mahalanobis metric was the main metric that was targeted by metric learning. A Mahalanobis distance is defined for a symmetric positive definite matrix Q by:

$$d_Q(x,y) = \sqrt{(x-y)^T Q (x-y)} \tag{12}$$

Xing et al. proposed an algorithm to learn a Mahalanobis distance (ie. the symmetric positive definite matrix) in [12].

3.2.2 The MMC algorithm

The MMC algorithm presented in [12] is a metric learning algorithm that uses convex optimization to learn the metric. We define D the set of couples with different labels and E the set of couples with the same labels. The objective function for the optimization problem is:

$$\epsilon(S) = \sum_{(x,y)\in D} d_S(x,y) \tag{13}$$

The optimization problem can be defined as:

$$\begin{array}{rl} & \max \epsilon(S) \\ \text{with constraints} & \left\{ \begin{array}{c} S \succcurlyeq 0 \\ \sum_{(x,y)\in E} d_S(x,y) < 1 \end{array} \right. \end{array} \tag{14}$$

The optimization problem is solved by the gradient ascent algorithm. At each step, the matrix S is modified using the formula below, then by projection on the two constraint spaces.

$$S = S + \eta \frac{\partial \epsilon(S)}{\partial S} \tag{15}$$

$$\frac{\partial \epsilon(S)}{\partial S} = \sum_{(x,y)\in D} \frac{\partial d_S(x,y)}{\partial S} \tag{16}$$

For an elliptic distance We derive equation 10.

$$\begin{aligned} \frac{\partial d^E_{CK}(x,y)}{\partial S} &= \kappa \frac{\sqrt{S_{x,x}S_{y,y}}}{\sqrt{S_{x,x}S_{y,y}-S^2_{x,y}}} \frac{\partial \frac{S_{x,y}}{\sqrt{S_{x,x}S_{y,y}}}}{\partial S} \\ &= \frac{\kappa}{\sqrt{S_{x,x}S_{y,y}-S^2_{x,y}}}\left(\frac{\partial S_{x,y}}{\partial S} - \frac{S_{x,y}}{2S_{x,x}S_{y,y}}\frac{\partial S_{x,x}S_{y,y}}{\partial S}\right) \\ &= \frac{\kappa}{2\sqrt{S_{x,x}S_{y,y}-S^2_{x,y}}}\left(2\frac{\partial S_{x,y}}{\partial S} - \frac{S_{x,y}}{S_{y,y}}\frac{\partial S_{y,y}}{\partial S} - \frac{S_{x,y}}{S_{x,x}}\frac{\partial S_{x,x}}{\partial S}\right) \end{aligned} \tag{17}$$

As in [1], we note $C_{x,y} = \tilde{x}\tilde{y}^T$ then we can express $S_{x,y}$ as:

$$S_{x,y} = Tr(C_{x,y}S) \tag{18}$$

Because S is symmetric and using equation 18:

$$\frac{\partial S_{x,y}}{\partial S} = C_{x,y} \tag{19}$$

Therefore, we find the follwing gradient.

$$\boxed{\frac{\partial d^E_{CK}(x,y)}{\partial S} = \frac{\kappa}{2\sqrt{S_{x,x}S_{y,y}-S^2_{x,y}}}\left(2C_{x,y} - \frac{S_{x,y}}{S_{y,y}}C_{y,y} - \frac{S_{x,y}}{S_{x,x}}C_{x,x}\right)} \tag{20}$$

We note that the same line of reasoning can be applied to the hyperbolic distance, by deriving equation 11.

3.3 Data Visualization with t-SNE

3.3.1 The Stochastic Neighbouring Method

Given a data set $X = \{x_1, x_2, ..x_n\}$ of high-dimension d. The aim of the Stochastic Neighbouring Method is to find a map $Y = \{y_1, y_2..y_n\}$ of dimension 2 or 3 that will allow a better vizualisation of the data set.

The SNE method starts by defining the conditional probabilities $p_{j|i}$ which represents the probability of picking x_j as a neighbour of x_i if the probability density is a Gaussian centered around x_i. In the classic SNE method, this probability decreases as the Euclidean distance between x_j and x_i increases and is defined by the following formula.

$$p_{j|i} = \begin{cases} \frac{\exp - \frac{\|x_i - x_j\|^2}{2\sigma_i^2}}{\sum_{k \neq i} \exp - \frac{\|x_i - x_j\|^2}{2\sigma_i^2}} \\ 0 \text{ if } i = j \end{cases} \tag{21}$$

Similarly, in the classic SNE method, a conditional probability $q_{j|i}$ is defined on the data set Y.

$$q_{j|i} = \begin{cases} \frac{\exp - \|y_i - y_j\|^2}{\sum_{k \neq i} \exp - \|y_i - y_j\|^2} \\ 0 \text{ if } i = j \end{cases} \tag{22}$$

The aim of the method is to find the points $y_1, ..y_n$ that will minimize the difference between the probability distributions p and q. Using gradient-descent, we need to minimize the cost C which corresponds to the sum of the KL divergences between the conditional probabilities.

$$C(Y) = \sum_{i \neq j} p_{j|i} \log(\frac{p_{j|i}}{q_{j|i}}) \tag{23}$$

3.3.2 The t-SNE method

The t-SNE method was introduced in [4] as a way to overcome some of the shortcomings of the original Stochastic Neighbouring Method by minimizing the KL-divergence between the joint probabilities instead. As such the cost function becomes

$$C(Y) = KL(P\|Q) = \sum_{i,j} p_{i,j} \log \frac{p_{i,j}}{q_{i,j}} \tag{24}$$

First, the joint probability p is obtained by symmetrizing the conditional probability expression from the SNE algorithm.

$$p_{i,j} = \frac{p_{j|i} + p_{i|j}}{2n} \tag{25}$$

Second, probability q is defined as a Student distribution and symmetrized.

$$q_{i,j} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq l}(1 + \|y_k - y_l\|^2)^{-1}} \tag{26}$$

3.3.3 t-SNE method in a Cayley-Klein metric

Adapting the t-SNE method to a Cayley-Klein metric means considering the Cayley-Klein distance between two points instead of their Euclidean distances. As a reminder, a Cayley-Klein geometry can be defined by a symmetric matrix S of dimension $(d + 1) \times (d + 1)$.

Given N data points $X \in \mathbb{R}^{N\times D}$, the initial symmetric matrix S can be defined as the following [2]:

$$S = \begin{pmatrix} \Sigma & -\Sigma\mu \\ -\mu^T\Sigma & \mu^T\Sigma\mu + \kappa^2 \end{pmatrix} \tag{27}$$

where Σ is the inverse covariance matrix of the data, and μ is the empirical mean vector.

In this form, it can be shown that $S_{x,y}$ can be expressed as a simple function of μ and Σ:

$$S_{x,y} = (x-\mu)^T\Sigma(y-\mu) + \kappa^2 \tag{28}$$

Probabilities in a CK space Because the two set of points considered don't have the same dimensions, we define first S the symmetric matrix of dimension $(d+1)\times(d+1)$ that will define the Cayley-Klein space in which the data set X will lie. Since points of the map are of lower dimension than the original data set, we can't directly compute a Cayley-Klein metric using S, but we can consider y to be the orthogonal projection of a point of dimension d.

We note $d_{CK}(x,y)$ the Cayley-Klein distance between two points.

In this case, the probability p is defined as follows:

$$p^{CK}_{j|i} = \begin{cases} \frac{\exp - \frac{d_{CK}(x_i,x_j)^2}{2\sigma_i^2}}{\sum_{k\neq i}\exp - \frac{d_{CK}(x_i,x_j)^2}{2\sigma_i^2}} \\ 0 \text{ if } i=j \end{cases} \tag{29}$$

And the probability q is defined by:

$$q^{CK}_{i,j} = \begin{cases} \frac{(1+d_{CK}(y_i,y_j)^2)^{-1}}{\sum_{k\neq l}(1+d_{CK}(y_k,y_l)^2)^{-1}} \\ 0 \text{ if } i=j \end{cases} \tag{30}$$

Gradient of the cost We follow the same reasoning as presented in Appendix A of [4] to compute the gradient of the cost.

$$C(Y) = \sum_{i\neq j} p_{ij}\log(p_{ij}) - p_{ij}\log(q_{ij}) \tag{31}$$

We define the following intermediary variables.

$$d_{ij} = d_{CK}(y_i,y_j)^2$$

$$Z = \textstyle\sum_{k\neq l}(1+d_{kl})^{-1}$$

If y_i changes, only d_{ij} and d_{ji} will vary, for any j given.

$$\begin{aligned} \frac{\partial C}{\partial y_i} &= \textstyle\sum_{j\neq i} \frac{\partial C}{\partial d_{ij}}\frac{\partial d_{ij}}{\partial y_i} + \frac{\partial C}{\partial d_{ji}}\frac{\partial d_{ji}}{\partial y_i} \\ &= 2\textstyle\sum_{j\neq i}\frac{\partial C}{\partial d_{ij}}\frac{\partial d_{ij}}{\partial y_i} \end{aligned} \tag{32}$$

According to [4],

$$\frac{\partial C}{\partial d_{ij}} = \frac{p_{ij} - q_{ij}}{1 + d_{ij}} \tag{33}$$

Derivative of the Cayley-Klein distance

$$\frac{\partial d_{ij}}{\partial y_i} = 2 d_{CK}(y_i, y_j) \frac{\partial d_{CK}(y_i, y_j)}{\partial y_i} \tag{34}$$

$$\frac{\partial C}{\partial y_i} = 4 \sum_{j \neq i} \frac{p_{ij} - q_{ij}}{1 + d_{CK}(y_i, y_j)^2} d_{CK}(y_i, y_j) \frac{\partial d_{CK}(y_i, y_j)}{\partial y_i} \tag{35}$$

For an elliptic Cayley-Klein distance We derive equation 10.

$$\begin{aligned} \frac{\partial d^E_{CK}(y_i,y_j)}{\partial y_i} &= \kappa \frac{\partial}{\partial y_i}\left(\frac{S_{ij}}{\sqrt{S_{jj} S_{ii}}}\right) \frac{-1}{\sqrt{1 - \frac{S_{ij}^2}{S_{ii} S_{jj}}}} \\ &= \frac{\kappa}{\sqrt{S_{jj}}} \frac{\frac{\partial S_{ij}}{\partial y_i} \sqrt{S_{ii}} - \frac{1}{2\sqrt{S_{ii}}} \frac{\partial S_{ii}}{\partial y_i} S_{ij}}{S_{ii}} \frac{-\sqrt{S_{ii} S_{jj}}}{\sqrt{S_{ii} S_{jj} - S_{ij}^2}} \end{aligned} \tag{36}$$

The derivative of the expressions $S_{i,j}$ are given by the following formulae.

$$\begin{aligned} \frac{\partial S_{ij}}{\partial y_i} &= S y_j \quad \text{if } i \neq j \\ \frac{\partial S_{ii}}{\partial y_i} &= 2 S y_i \end{aligned} \tag{37}$$

$$\begin{aligned} \frac{\partial d^E_{CK}(y_i,y_j)}{\partial y_i} &= \frac{\kappa}{\sqrt{S_{jj}}} \frac{S y_j \sqrt{S_{ii}} - \frac{S y_i S_{ij}}{\sqrt{S_{ii}}}}{S_{ii}} \frac{-\sqrt{S_{ii} S_{jj}}}{\sqrt{S_{ii} S_{jj} - S_{ij}^2}} \\ &= -\frac{\kappa}{\sqrt{S_{ii} S_{jj} - S_{ij}^2}} S\left(y_j - \frac{S_{ij}}{S_{ii}} y_i\right) \end{aligned} \tag{38}$$

$$\boxed{\frac{\partial C}{\partial y_i} = -4\kappa \sum_{j \neq i} \frac{p_{ij} - q_{ij}}{1 + d_{CK}(y_i,y_j)^2} d_{CK}(y_i, y_j) \frac{S}{\sqrt{S_{ii} S_{jj} - S_{ij}^2}} \left(y_j - \frac{S_{ij}}{S_{ii}} y_i\right)} \tag{39}$$

For a hyperbolic Cayley-Klein distance

$$\begin{aligned} \frac{\partial d^H_{CK}(y_i,y_j)}{\partial y_i} &= \kappa \frac{\partial}{\partial y_i}\left(\frac{S_{ij}}{\sqrt{S_{jj} S_{ii}}}\right) \frac{1}{\sqrt{\frac{S_{ij}^2}{S_{ii} S_{jj}} - 1}} \\ &= \frac{\kappa}{\sqrt{S_{jj}}} \frac{\frac{\partial S_{ij}}{\partial y_i} \sqrt{S_{ii}} - \frac{1}{2\sqrt{S_{ii}}} \frac{\partial S_{ii}}{\partial y_i} S_{ij}}{S_{ii}} \frac{\sqrt{S_{ii} S_{jj}}}{\sqrt{S_{ij}^2 - S_{ii} S_{jj}}} \end{aligned} \tag{40}$$

$$\boxed{\frac{\partial C}{\partial y_i} = 4\kappa \sum_{j \neq i} \frac{p_{ij} - q_{ij}}{1 + d_{CK}(y_i,y_j)^2} d_{CK}(y_i, y_j) \frac{S}{\sqrt{S_{ij}^2 - S_{ii} S_{jj}}} \left(y_j - \frac{S_{ij}}{S_{ii}} y_i\right)} \tag{41}$$

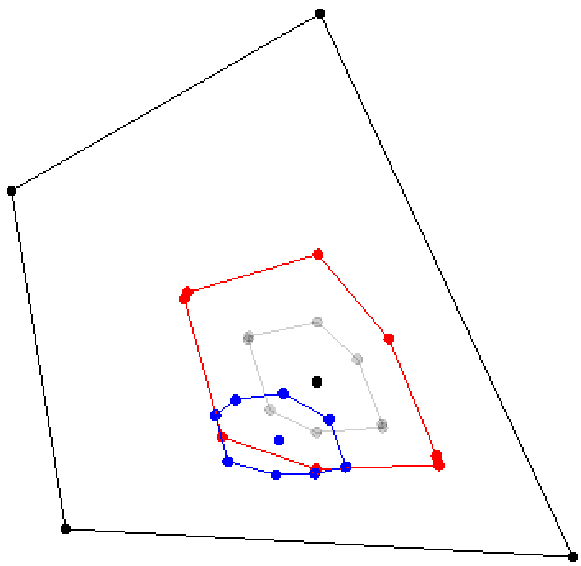

Figure 14: In a quadrangle domain, the blue ball and the red ball are not tangent but they still share part of two non-neighboring edges.

4 Conclusion

Knowing that Cayley-Klein metrics have been used to improve metric learning algorithms, we first studied a generalization of these geometries. In a polygonal Hilbert geometry, we gave an explicit description of balls in the geometry, proving that they are polygonal with a number of edges that depends on the geometry of the domain and the position of the center in it, but not on the radius of the ball. We also gave some characterization of the intersection of two balls, highlighting the existence of infinite intersection in non-trivial cases.

We then studied how Cayley-Klein metrics could be used to tackle Machine Learning problems such as Metric learning for unsupervised classification and Stochastic Neighbouring Method for high-dimensional data embedding. For the first one we analyzed the algorithm presented in [1] and [6]. For the SNE method, we showed how Cayley-Klein distances could be used to replace the Euclidean distance.

The work on Hilbert geometries can be extended to better characterizing the intersection of two balls. Namely, we observed that tangency is not the only non-trivial case of intersection - see Figure 14. Lastly, better understanding the intersection of two balls could help understand Voronoi diagrams in this geometry, as the shape of a bisector also seems to depend on the geometry and the position of center points. (Fig 15) . Moreover, we observe than in extreme cases, Voronoi diagrams are not piecewise-linear (Figure 16.)

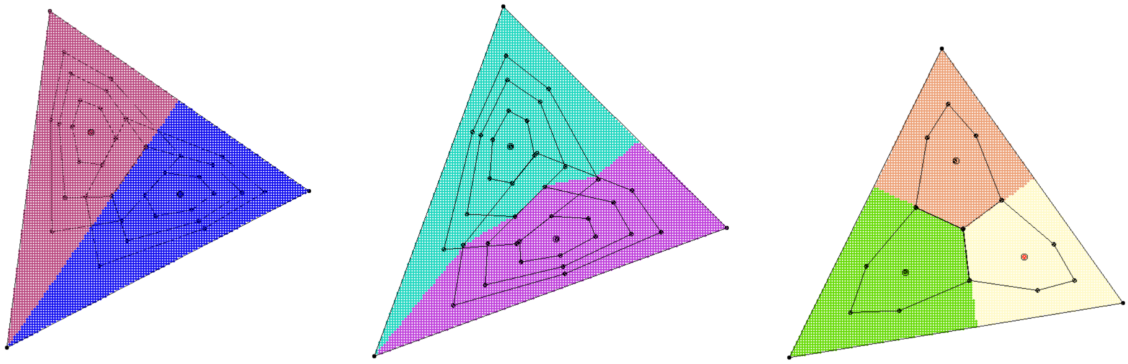

Figure 15: Voronoi diagrams in a triangular domain. Left: two balls of same radius share an edge. The bisector is a straight line. Middle: two balls of same radius share part of an edge. The bisector is piecewise linear. Right: Voronoi diagram for three center points.

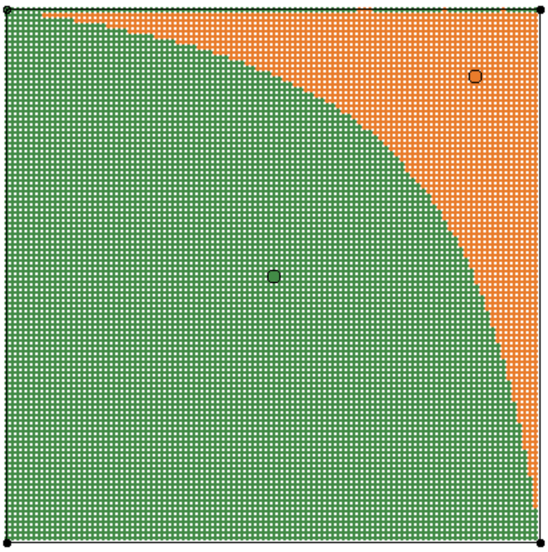

Figure 16: In a square domain, two points on the diagonals and their corresponding Voronoi diagram which is non linear.

A Notations

- d_{CK} is a generic Cayley-Klein distance.
- $d_{CK}^E(p,q)$ is an elliptic Cayley-Klein distance.
- $d_{CK}^H(p,q)$ is a hyperbolic Cayley-Klein distance.
- If S is a symmetric matrix, the bilinear form associated to it is noted $S_{x,y}$ and the distance in the associated Cayley-Klein space is noted d_S.
- d is the dimension of the data.
- n is the number of elements in the data set.
- if x is a vector $\in \mathbb{R}^d$, we note $\tilde{x} \in \mathbb{PR}^d$, its expression in homogeneous coordinates.
- $\mathcal{C}$ is a generic convex domain for a Hilbert geometry.
- When $\mathcal{C}$ is a simplex, we note n its number of vertices and $e_1, ..e_n$ said vertices.
- $d_{\mathcal{C}}(p,q)$ is the distance between two points p and q in a Hilbert geometry defined on the convex domain $\mathcal{C}$.

References

[1] Yanhong Bi, Bin Fan, and Fuchao Wu. Beyond Mahalanobis metric: Cayley-Klein metric learning. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[2] Yanhong Bi, Bin Fan, and Fuchao Wu. Beyond Mahalanobis metric: Cayley-Klein metric learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2339–2347, 2015.

[3] Pierre de La Harpe. On Hilbert's metric for simplices. *Geometric group theory*, 1:97–119, 1993.

[4] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(Nov):2579–2605, 2008.

[5] Curtis T. McMullen. Coxeter groups, Salem numbers and the Hilbert metric. *Publications mathématiques de l'IHÉS*, 95(1):151–183, 2002.

[6] Frank Nielsen, Boris Muzellec, and Richard Nock. Classification with mixtures of curved Mahalanobis metrics. In *Image Processing (ICIP), 2016 IEEE International Conference on*, pages 241–245. IEEE, 2016.

[7] Frank Nielsen and Laetitia Shao. On balls in a Hilbert polygonal geometry. In *Symposium on Computational Geometry (SoCG Multimedia Contribution)*, volume 77 of *LIPIcs*, pages 67:1–67:4. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2017.

[8] Frank Nielsen and Ke Sun. Clustering in Hilbert simplex geometry. *CoRR*, abs/1704.00454, 2017.

[9] Victor Pambuccian. The elementary geometry of a triangular world with hexagonal circles. *Contributions to Algebra and Geometry*, 49(1):165–175, 2008.

[10] Jürgen Richter-Gebert. *Perspectives on Projective Geometry: A Guided Tour Through Real and Complex Geometry*. Springer Publishing Company, Incorporated, 1st edition, 2011.

[11] Constantin Vernicos. On the Hilbert geometry of convex polytopes. *preprint arXiv:1406.0733*, 2014.

[12] Eric P. Xing, Andrew Y. Ng, Michael I. Jordan, and Stuart Russell. Distance metric learning with application to clustering with side-information. *Advances in neural information processing systems*, 15:505–512, 2003.