Self-Organizing Maps on non-euclidean Spaces

Helge Ritter
Faculty of Technology
Bielefeld University,
D-33501 Bielefeld, Germany

We propose a new type of Self-Organizing Map (SOM) that is based on discretizations of curved, non-euclidean spaces. As an introductory example, we briefly discuss "spherical SOMs" on tesselations of the sphere for the display of directional data. We then describe the construction of "hyperbolic SOMs", using regular tesselations of the hyperbolic plane, which is a non-euclidean space characterized by constant negative gaussian curvature. The approach is motivated by the recent observation that the geometry of hyperbolic spaces possesses very favourable properties for the mapping of hierarchical data. We conclude with some initial simulation results illustrating some properties of the hyperbolic SOM and discuss a number of issues for future research.

1. INTRODUCTION

The Self-Organizing Map, as introduced by Kohonen more than a decade ago, has stimulated an enormous body of work in a broad range of applied and theoretical fields, including pattern recognition, brain theory, biological modeling, mathematics, signal processing, data mining and many more [8]. Much of this impressive success is owed to the combination of elegant simplicity in the SOM's algorithmic formulation, together with a high ability to produce useful answers for a wide variety of applied data processing tasks and even to provide a good model of important aspects of structure formation processes in neural systems.

While the applications of the SOM are extremely wide-spread, the majority of uses still follow the original motivation of the SOM: to create dimension-reduced "feature maps" for various uses, most prominently perhaps for the purpose of data visualization.

The suitability of the SOM for this task has been analyzed in great detail and linked to earlier approaches, such as PCA, multidimensional scaling, principal surfaces and the like. At the same time, this property of the SOM has stimulated the design of a number of related algorithms, including hierarchical variants of the SOM [10], the ASSOM [9] introducing the idea of mapping onto subspaces, the PSOM [21], the GTM [1] and the stochastic SSOM for mapping of proximity data [5], to just name a few.

While these algorithms often differ from the SOM in many details, they share with it the idea of using a deformable template to translate *data similarities* into *spatial relationships*. While the way the template is implemented (e.g., as a discrete lattice or as a continuous, parametrized manifold) can affect important details of the generated mappings, a much larger role is played by its topological structure.

So far, the overwhelming majority of SOM approaches have taken it for granted to use (some subregion of) a *flat space* as their data model and, motivated by its convenience for visualization, have favored the (suitably discretized) *euclidean plane* as their chief "canvas" for the generated mappings, although a number of works also has considered higher-dimensional euclidean lattices to represent, e.g., the configuration manifold of a robot arm [16] and there have been a few approaches to use different lattice topologies, such as hierarchical trees [10] and subsequent work, or hypercubic lattices [2] to adapt the SOM to particular requirements in the data.

2. NON-EUCLIDEAN SOMS AND THEIR MOTIVATION

However, even if our thinking is deeply entrenched with the "obvious" structure of euclidean space, this is not necessarily honored by the multitude of data occurring around us. An obvious example is already given by *directional data*: in this case, the natural data model is the surface S_2 of a sphere; this is a *non-euclidean* space characterized by *uniform positive curvature* and it is a well-known result that there is no singularity-free continuous one-to-one mapping of this space onto a whatsoever shaped piece of flat $\mathbb{R}^2$ (which regularly bothers when we have to describe points on the sphere by a pair of coordinates, which always is a hidden attempt to identify S_2 with a part of $\mathbb{R}^2$). Therefore, many directional data may much more naturally admit a projection onto a SOM with a *spherical topology* (which in addition to its special symmetry also does not suffer from a boundary). However, to the knowledge of the present author, such attempts have not appeared in the literature so far [1]. Part of the reason for this may be the somewhat more difficult handling of a "spherical lattice" and in Sec. 3 we briefly comment on this point.

Another interesting type of data are *hierarchical structures*. The capability of the 2d "euclidean" SOM to map even an amazing amount of the semantic features of language [17], which gave rise to the ambitious WEBSOM project [6], is remarkable, but it also raises the question to what extent a "flat" $\mathbb{R}^2$ (or a higher-dimensional euclidean $\mathbb{R}^D$) provides sufficient freedom to map the neighborhood of an item from a more complex "information space" (such as language) into spatial relationships in the chosen euclidean projection space.

The obvious limiting factor is the rather restricted neighborhood that "fits" around a point on a 2d lattice. Recently, it has been observed that another type of non-euclidean spaces, the *hyperbolic spaces* that are characterized by uniform negative curvature, are very well suited to overcome this limitation [11] since their geometry is such that the size of a neighborhood around a point increases *exponentially* with its radius r (while in a D-dimensional euclidean space the growth follows the much slower power law r^D). This exponential scaling behavior fits very nicely with the scaling behavior within hierarchical, tree-like structures, where the number of items r steps away from the root grows as b^r where b is the (average) branching factor. This interesting property of hyperbolic spaces has been exploited for creating novel displays of large hierarchical structures that are more accessible to visual inspection than in previous approaches [12].

[1] to avoid boundary effects, periodic boundary conditions for 2d-lattices are a well-known and frequently used technique; the resulting topology, however, is that of a torus, which is of zero curvature and topologically different from a sphere.

Therefore, it appears very promising to use hyperbolic spaces also in conjunction with the SOM, and in the present contribution we want to present some initial results along these lines. The resulting *hyperbolic SOMs* are based on a tesselation of the hyperbolic plane (or some higher-dimensional hyperbolic space) and their lattice neighborhood reflects the hyperbolic distance metric that is responsible for the non-intuitive properties of hyperbolic spaces. Together with the spherical SOM, they provide an important special case of *"non-euclidean" SOMs*, which are the generalization of the SOM to non-euclidean spaces. By making this generalization we obtain an important new degree of freedom for choosing our "template" such that it may better support the desired translation of data similarities into spatial neighborhood relationships. Specifically, for hyperbolic SOMs, we argue that the nature of the chose "space template" is particularly adequate for mapping high-dimensional and hierarchical structures.

Since the notion of non-euclidean spaces may be unfamiliar to many readers, we below first make a few remarks on spherical SOMs which may help to form an intuitive idea of the general concept. Then we give a brief account of some basic properties of hyperbolic spaces that are exploited for hyperbolic SOMs, in particular *regular tesselations* of the hyperbolic plane. Finally, we provide some hints on the actual implementation of hyperbolic SOMs after which we conclude with some computer experiments and a discussion.

3. SPHERICAL SOMS

The surface of the sphere provides us with the simplest model of a non-euclidean, *curved space*. It is curved, since no finite patch of it can be mapped onto $\mathbb{R}^2$ without distorting local distances. Like in the plane, the distance between two points on a sphere is given by the shortest path connecting the two points, but this path is now an arc, since it is confined to run within the surface. As a result, no two points are separated by more than the maximal distance of π and the area in a circular neighborhood of radius r around a point grows slower than the usual πr^2 law in the plane.

Since we are quite familiar with the sphere, we find these properties not too surprising, but we see the striking differences as compared to the geometry on flat $\mathbb{R}^2$. By virtue of our geometric picture of the sphere we easily can imagine that the sphere will be very suitable to create mappings of data with an underlying *directional structure*. If we want to create a topology-preserving mapping for such data, we should provide a *spherical SOM*, i.e., we must use node lattice whose topology mimicks that of a sphere with a corresponding distance measure between nodes.

Unfortunately, there exist only five regular tesselations of the sphere (the five platonic bodies *tetrahedron*, *octahedron*, *cube*, *pentagondodecahedron* and *icosahedron*) which are all rather coarse. However, if we are willing to make a slight departure from perfect uniformity we can recursively subdivide the triangles of the icosahedron (at each step subdividing each triangle into four smaller ones as indicated in Fig.1 and "lifting" the newly introduced midpoints to the surface of the bounding sphere), leading to spherical triangle tesselations with arbitrarily fine resolution (these are only "almost" regular, since

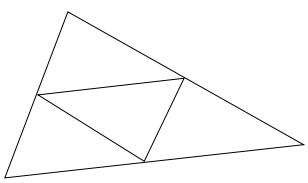

Figure 1: Recursive triangle split

the original icosahedron and the newly created vertices are surrounded by 5 and 6 triangles, resp.).

At first sight, it may seem cumbersome to evaluate the required distance function on the icosaeder-based tesselations. However, it is fairly easy to compute along with the recursive tesselation the radial normals $\vec{n}_r$ towards the newly created lattice nodes r. Then, the required distance is simply given by $d_{r,r'} = \arccos(\vec{n}_r \cdot \vec{n}_{r'})$, which we can plug into the standard SOM adaptation rule. For lack of space, we will not go into further details about spherical SOMs and instead focus now on their less intuitive counterparts, the hyperbolic SOMs.

4. HYPERBOLIC SPACES

While the sphere is characterized by a constant positive gaussian curvature (i.e., the surface "bends" into the same perpendicular direction, as one moves along two orthogonal directions), there are also surfaces that possess *negative gaussian curvature*. The counterpart of the sphere is achieved by requiring a constant negative curvature everywhere, and the resulting space is known as the *hyperbolic plane H2* (with analogous generalizations to higher dimensions)[3,20] [2]. Unfortunately, it is no longer possible to embed H2 in a (distance-preserving) way in our euclidean $\mathbb{R}^3$, but we still can do so with sufficiently small "patches" of H2. The embedded patches then resemble the shape of a "saddle", i.e., the negative curvature shows up as a local bending into opposite normal directions, as we move on orthogonal lines along the patch. This may make it intuitively plausible that on such surfaces the area (and also the circumference) of a circular neighborhood around a point now grows faster than in the uncurved case. The geometry of H2 is a standard topic in *Riemannian geometry* (see, e.g. [19,15], and the relationships for the area A and the circumference C of a circle of radius r are given by

$$A = 4\pi \sinh^2(r/2) \quad (1)$$

$$C = 2\pi \sinh(r) \quad (2)$$

These formulae exhibit the highly remarkable property that both quantities grow *exponentially* with the radius r (whereas in the limit $r \to 0$ the curvature becomes insignificant and we recover the familiar laws for flat $\mathbb{R}^2$). It is this property that was observed in [11] to make hyperbolic spaces extremely useful for accommodating hierarchical structures: their neighborhoods are in a sense "much larger" than in the non-curved euclidean (or in the even "smaller" positively curved) spaces (it makes also intuitively understandable why no finite-dimensional euclidean space allows an (isometric) embedding: it only has space for a power-law-growth).

To use this potential for the SOM, we must solve two problems: (*i*) we must find suitable discretization lattices on H2 to which we can "attach" the SOM prototype vectors. (*ii*) after having constructed the SOM, we must somehow project the (hyperbolic!) lattice into "flat space" in order to be able to inspect the generated maps (since we believe that it will remain difficult to buy hyperbolic computer screens within the foreseable future).

[2] We gloss over many mathematical details here; the mathematical treatment of curvature is based on the *Riemann curvature tensor* and requires the formalism of differential geometry, see e.g., [15] for an accessible introduction.

Fortunately, both problems have already been solved more than a century ago by the mathematicians exploring the properties of non-euclidean spaces (the historic references are [4,7], for newer accounts see, e.g., [13,3]), and the next two subsections summarize some of the results that we will need.

4.1. Projections of hyperbolic spaces

While the structure of hyperbolic spaces makes it impossible to find an isometric (ie., distance-preserving) embedding into a euclidean space, there are several ways to construct very useful embeddings under weaker conditions (in the following, we will restrict the discussion to embeddings of the hyperbolic plane H2). The closest analogy to the sphere-embedding occurs if we sacrifice the euclidean structure of the embedding space and instead endow it with a *Minkowski-metric* [14]. In such a space, the squared distance d^2 between two points (x, y, u) and (x', y', u') is given by

$$d^2 = (x - x')^2 + (y - y')^2 - (u - u')^2 \tag{3}$$

i.e., it ceases to be positive definite. Still, this is a space with zero curvature and its somewhat strange distance measure allows to construct an *isometric* embedding of the hyperbolic plane, given by

$$x = \sinh(\rho)\cos(\phi) \tag{4}$$

$$y = \sinh(\rho)\sin(\phi) \tag{5}$$

$$u = \cosh(\rho) \tag{6}$$

where (ρ, ϕ) are polar coordinates on the hyperbolic plane (note the close analogy of (4)–(6) with the formulas for the embedding of a sphere by means of spherical polar coordinates in $\mathbb{R}^3$!). Under this embedding, the hyperbolic plane appears as the surface M swept out by rotating the curve $u^2 = 1 + x^2 + y^2$ about the u-axis[3].

From this embedding, we can construct two further ones, the so-called *Klein model* and the *Poincaré model* (the latter will be used to visualize hyperbolic SOMs below). Both achieve a projection of the infinite H2 into the unit disk, however, at the price of distorting distances. The Klein model is obtained by projecting the points of M onto the plane $u = 1$ along rays passing through the origin O (see Fig.2). Obviously, this projects all points of M into the "flat" unit disk $x^2 + y^2 < 1$ of $\mathbb{R}^2$. (e.g., $A \mapsto B$). The Poincaré Model results if we add two further steps: first a perpendicular projection of the Klein Model (e.g., a point B) onto the ("northern") surface of the unit sphere centered at the origin (point C), and then a stereographic projection of the "northern" hemisphere onto the unit circle about the origin in the ground plane $u = 0$ (point D). It turns out that the resulting projection of H2 has a number of pleasant properties, among them the preservation of

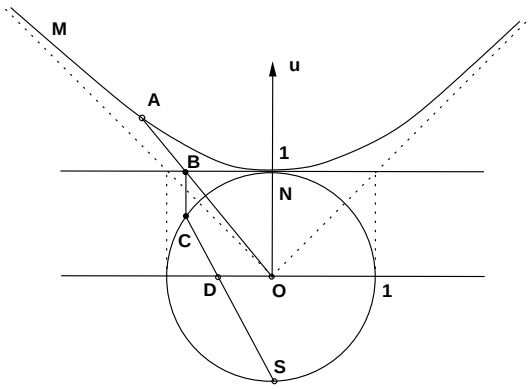

Figure 2: Construction steps underlying *Klein* and *Poincaré*-models of the hyperbolic space H2

[3]the alert reader may notice the absence of the previously described local saddle structure; this is a consequence of the use of a Minkowski metric for the embedding space, which is not completely compatible with our "euclidean" expectations.

angles and the mapping of shortest paths onto circular arcs belonging to circles that intersect the unit disk at right angles. Distances in the original H2 are strongly distorted in its Poincaré (and also in the Klein) image (cf. Eq. (10)), however, in a rather useful way: the mapping exhibits a strong "fisheye"-effect. The neighborhood of the H2 origin is mapped almost faithfully (up to a linear shrinkage factor of 2), while more distant regions become increasingly "squeezed". Since asymptotically the radial distances and the circumference grow both according to the same exponential law, the squeezing is "conformal", ie., (sufficiently small) shapes painted onto H2 are not deformed, only their size shrinks with increasing distance from the origin. By translating the original H2 [4]. the fisheye-fovea can be moved to any other part of H2, allowing to selectively zoom-in on interesting portions of a map painted on H2 while still keeping a coarser view of its surrounding context.

4.2. Tesselations of the hyperbolic plane

To complete the set-up for a hyperbolic SOM we still need an equivalent of a regular grid in the hyperbolic plane. Fortunately, the tesselation of hyperbolic spaces is a well developed subject in the mathematics literature, and we can borrow the following results: while the choices for tesselations with congruent polygons on the sphere and even in the plane such that each grid point is surrounded by the same number n of neighbors are severely limited (the only possible values for n being 3,4,5 on the sphere, and 3,4,6 in the plane), there is an infinite set of choices for the hyperbolic plane. In the following, we will restrict ourselves to lattices consisting of equilateral triangles only. In this case, there is for each $n \geq 7$ a regular tesselation such that each vertex is surrounded by n congruent equilateral triangles. Fig.3 shows two example tesselations (for the minimal value of $n = 7$ and for $n = 10$), using the Poincaré model for their visualization. While in Fig.3 these tesselations appear non-uniform, this is only due to the fisheye effect of the Poincaré projection. In the original H2, each triangle has the same size, and this can be checked by re-projecting any distant part of the tesselation into the center of the Poincaré disk, after which it looks identical (up to a possible rotation) to the center of Fig.3.

As already evident from inspection of Fig.3, the resulting lattices share with the underlying space H2 the property that the number of vertices that fall within a given distance r of a node increase (asymptotically) exponentially with r. Thus, they correctly mimick the essential property of providing "more neighborhood space" to each node than tesselations derived from a euclidean space can do.

One way to generate these tesselations algorithmically is by repeated application of a suitable set of generators of their symmetry group to a (suitably sized, cf. below) "starting triangle". For the here considered case of a uniform mesh of equilateral triangles, a suitable set of generators are the hyperbolic rotations around (the "peripheral" subset of) the already generated triangle vertices. These are special cases of (orientation-preserving) isometries on H2 and it can be shown that in the Poincaré model any such isometry can

[4] the readers with physics background will observe that the translations of H2 correspond to a subgroup of the Lorentz group for the chosen Minkowski embedding space; in the Poincaré model such mappings can be most elegantly described by identifying each point (x, y) with the complex number $z = x + iy$. Then a H2 translation is described by the "disk-automorphism" $z' = (z - a)/(1 - \bar{a}z)$ where a denotes the point that becomes translated into the origin.

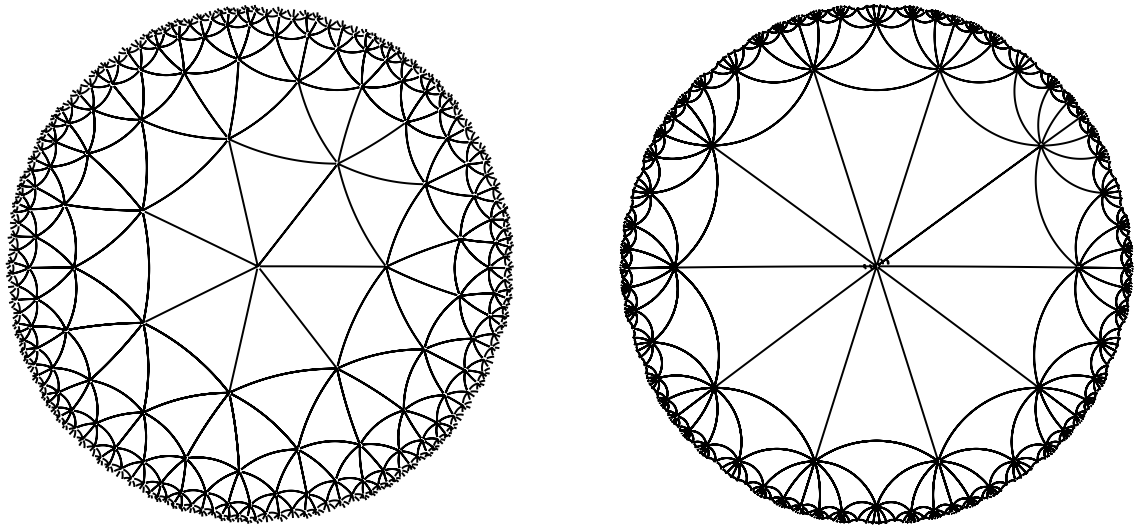

Figure 3. Regular triangle tesselations of the hyperbolic plane, projected into the unit disk using the Poincaré mapping. The leftmost tesselation shows the case where the minimal number ($n = 7$) of equilateral triangles meet at each vertex and is best suited for the hyperbolic SOM, since tesselations for larger values of n (right: $n = 10$) lead to bigger triangles. In the Poincaré projection, only sides passing through the origin appear straight, all other sides appear as circular arcs, although in the original space all triangles are congruent.

be described as a complex mapping

$$z \mapsto z' = T(\phi, \zeta)(z) = e^{i\phi} \frac{z - \zeta}{1 - \bar{\zeta} z} \tag{7}$$

where ϕ is an angle and $\zeta \in C$ is some complex number with $|\zeta| < 1$. In particular, a hyperbolic rotation $D(\alpha, z_0) = T(\phi, \zeta)$ of angle α about a point z_0 can be decomposed into two translations and a rotation about the origin

$$D(\alpha, z_0) = T(0, -z_0) \circ T(\alpha, 0) \circ T(0, z_0) \tag{8}$$

from which ϕ and ζ can be obtained analytically.

For a triangle tesselation with vertex order n (ie. $\alpha = 2\pi/n$), we would apply these hyperbolic rotations to an equilateral starting triangle with equal corner angles α. In a hyperbolic space, specifying the angles of a triangle automatically fixes its size: unlike the euclidean case, the angles no longer sum up to π, instead, the sum is always less and the "deficit" equals the area of the triangle. Therefore, the smallest triangles are obtained for the minimal value of $n = 7$, which, therefore, leads to the "finest" tesselation and will be our choice for the simulations reported below. Since the resulting lattice structure is very different from a rectangular array, there is no straighforward indexing scheme

and an efficient implementation has to store pointers to the topological neighbors of each newly generated vertex in order to allow iterating the SOM update step over the lattice neighborhood of each winner node.

5. HYPERBOLIC SOM ALGORITHM

We have now all ingredients required for a "hyperbolic SOM". It employs a (finite patch) of a hyperbolic lattice, e.g., the regular triangle tesselation with vertex order $n = 7$. Using the construction scheme sketched in the previous section, we can organize the nodes of such a lattice as "rings" around an origin node (ie., it is simplest to build approximately "circular" lattices). The numbers of nodes of such a lattice grows very rapidly (asymptotically exponentially) with the chosen lattice radius R (its number of rings). For instance, for $n = 7$, Table 1 shows the total number N_R of nodes of the resulting regular hyperbolic lattices with different radii ranging from $R = 1$ to $R = 10$. Each lattice node r carries a prototype vector $\vec{w}_r \in I\!R^D$ from some D-dimensional feature space (if we wish to make any non-standard assumptions about the metric structure of this space, we would build this into the distance metric that is used for determining the best-match node). The SOM is then formed in the usual way, e.g., in on-line mode by repeatedly determining the winner node s and adjusting all nodes $r \in N(s,t)$ in a radial lattice neighborhood $N(s,t)$ around s according to the familiar rule

$$\Delta\vec{w}_r = \eta h_{rs}(\vec{x} - \vec{w}_r) \tag{9}$$

with $h_{rs} = \exp(-d^2(r,s)/2\sigma^2)$. However, since we now work on a hyperbolic lattice, we have to determine both the neighborhood $N(s,t)$ and the (squared) node distance $d^2(r,s)$ according to the natural metric that is inherited by the hyperbolic lattice.

R	1	2	3	4	5	6	7	8	9	10
N_R	8	29	85	232	617	1625	4264	11173	29261	76616

Table 1: node numbers N_R of hyperbolic triangle lattices with vertex order 7 for different numbers R of "node rings" around the origin

The simplest way to do this is to keep with each node r a complex number z_r to identify its position in the hyperbolic space (it is convenient to use the Poincaré model; with correspondingly changed distance formulas any of the other models would do as well). The node distance is then given (using the Poincaré model, see e.g. [19]) as

$$d = 2\text{arctanh}\left(\left|\frac{z_r - z_s}{1 - \bar{z}_s \cdot z_r}\right|\right) \tag{10}$$

The neighborhood $N(t,s)$ can be defined as the subset of nodes within a certain graph distance (which is chosen as a small multiple of the neighborhood radius σ) around s. In our current implementation, we use a recursive node traversal scheme which is more

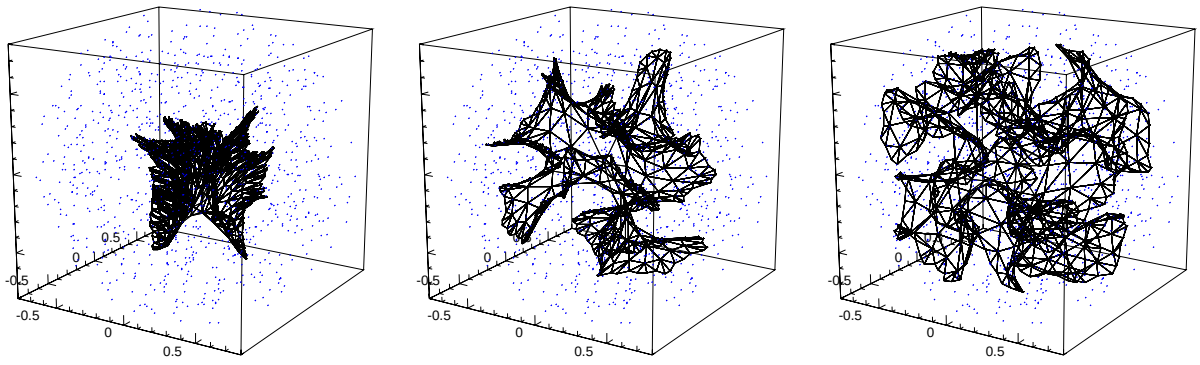

Figure 4. Unfolding of a 2d-hyperbolic SOM lattice in a uniform data density inside the 3-dimensional unit sphere. *(a)(left)*: early ordering phase, *(b)(center)* emergence of local "saddle structure", *(c)(right)* final configuration: the node-rich periphery of the hyperbolic lattice allows the realization of a highly contorted "space-filling" surface to approximate the 3d-topology of the spherical data distribution.

costly than a scheme optimized for a particular lattice type; however, it still is reasonably fast and has the virtue that it can be used for arbitrary lattice topologies.

Like the standard SOM, also the hyperbolic SOM can become trapped in topological defects. Therefore, it is also important here to control the neighborhood radius $\sigma(t)$ from an initially large to a final small value (we use a simple exponential decrease, $\sigma(t) = (\sigma_{T_{\max}}/\sigma(0))^{t/T_{\max}}$). In addition, we found two measures very effective to support the ordering process further: (i), the "conscience mechanism" introduced by deSieno [18], which equalizes the winning rates of the nodes by adding for each node to the computation of its distance to the input a "distance penalty" which is raised whenever a node becomes winner, and which is lowered otherwise, and (ii) "radial learning", which initially lets only the subset of nodes closer than a distance R to the origin participate in the winner determination, where R gradually increases from 0 to the final, maximal value during training. While this is also useful for the euclidean SOM, it is particularly effective in the hyperbolic case, since it ensures a proper centering on the data distribution before the node-rich periphery with its many degrees of freedom gets involved.

6. EXPERIMENTS

In the following, we give some examples of hyperbolic SOMs ("HSOMs") generated on simple data sets and compare the resulting maps with those of a standard 2d SOM.

As a first data set we use a uniform point distribution in the unit sphere of $I\!R^3$. This is not yet a high-dimensional data set, but in three dimensions we still can follow the unfolding of the map using perspective drawings of the sequence of "virtual nets" in their 3d-embedding space.

Figs.4a-c show three development stages of this process for a HSOM with $R = 4$ (232 nodes). The pictures give some insight into how the HSOM takes advantage of its "larger

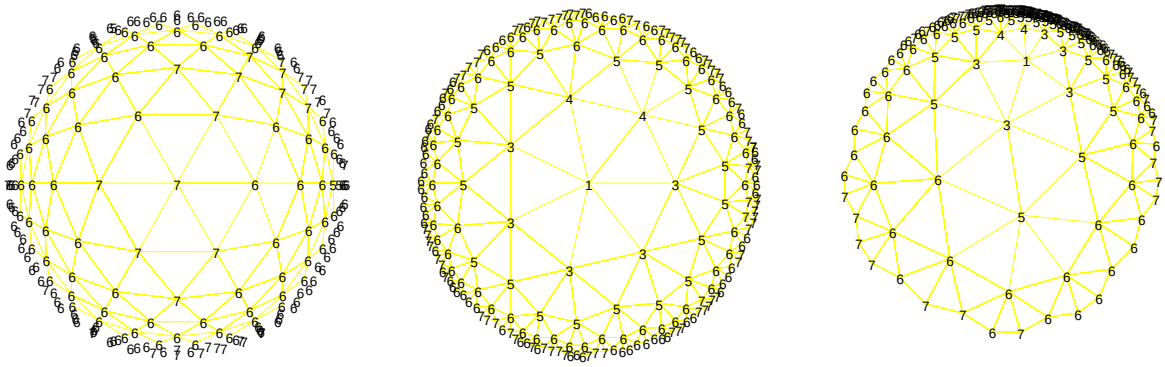

Figure 5. Projection of 5d-sphere data on euclidean SOM *(a)(left)* and on hyperbolic SOM *(b)(center)*. The right diagram *(c)* shows that the hyperbolic lattice allows to obtain a detailed, magnified view of the border region while still keeping the central region (now upper right) well visible (compare sizes of edges emanating from `'1'`-node). Note that circular arcs in the hyperbolic image (cf. Fig.3) have been simplified to straight lines.

neighborhood" to map the 3d space onto its 2d hyperbolic lattice. Initially, with $\sigma(t)$ still large (Fig. 4a, left), the map contracts and centers itself in the data distribution in the usual way, as familiar from the standard SOM. Then, as $\sigma(t)$ gets lower (Fig. 4b, center), the map starts to "feel" the dimension conflict (which here still is rather mild and only $3d$ vs. $2d$). This is accompanied by developing protrusions to better fill the higher-dimensional volume with the $2d$ lattice. Now the advantage of the hyperbolic $2d$-lattice over its euclidean cousin becomes apparent: due to its hyperbolic nature the growth of the number of its nodes with radial distance r can much better cope with the required r^3 law of the data distribution (and, due to the exponential growth law, this would remain so for data of *any* dimensionality D)! This makes its periphery much more flexible to fold into the missing space dimensions (of course, this still can be done only by way of making the boundary produce a "meandering" space filling curve). The final result is visible in Fig 4c (right), which is reminiscent of a 2d-saddle whose (1d-) periphery has become deformed to follow a space filling curve in the (2d) surface of a sphere.

This behavior becomes even more pronounced as we proceed towards higher data space dimensionalities. We then can no longer visualize the corresponding "virtual" nets, but we still can show the distribution of the data points on the 2d SOM lattice. Fig.5 compares the results for the same experiment, but now with points distributed with uniform density in the 5-dimensional unit ball $\|\vec{r}\| < 1$ of $\mathbb{R}^5$. To compare the result for the hyperbolic SOM with that of a "flat" euclidean SOM, we need some comparable displays of the underlying, very different lattice topologies. For the hyperbolic SOM, we use the Poincaré projection into the unit disk (as already familiar from Fig.3). To create a roughly comparable display for the "flat" SOM (for which we used a hexagonal lattice with size adjusted for about the same number of nodes as were used for the HSOM), we performed a central projection of its lattice hexagon onto a hemisphere that we then view from above. To visualize at least one dimension of the underlying data distribution, we use the radial coordinate

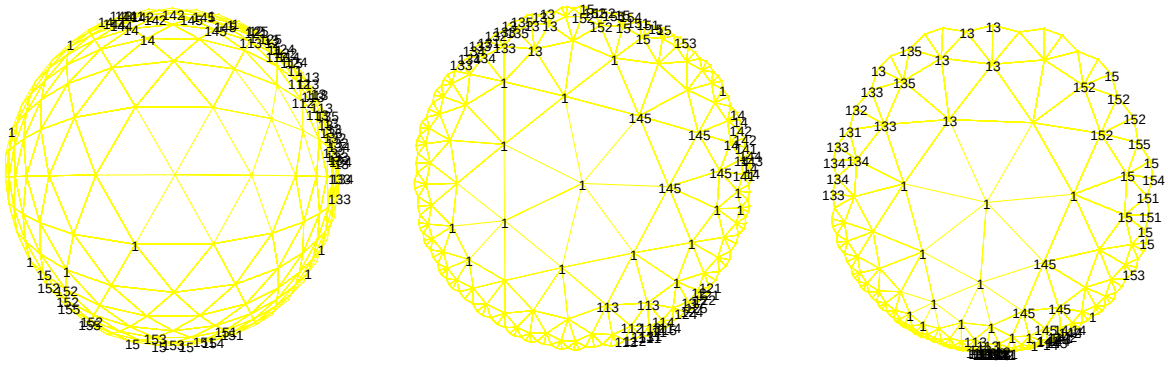

Figure 6. Projections obtained for synthetic, hierarchically clustered data set in 5d space. *(a)(left)* euclidean SOM, *(b), center* and *(c), right*: two different views of the hyperbolic SOM (again with connection arcs simplified to straight lines).

and label in both images each node by the (scaled) radial distance $|\vec{w}_r|$ of its prototype vector $\vec{w}_r$ (which, for a continuous distribution, represents the best-match data point) from the origin. Fig.5a (left) depicts the resulting map for the standard SOM, while the central diagram shows the HSOM. The difference is striking: while the euclidean SOM has primarily mapped the "superficial" structure of the sphere (all radius values are similar and close to the maximal value), the HSOM gives a much clearer impression of the radial structure, arranging spherical shells of increasing radius into concentric rings around its origin. In Fig.5c (right) we have moved the "focus" into the border region of the hyperbolic SOM to illustrate the exponential "scale compression" of the hyperbolic lattice: we can now see a magnified view of a rather small part of the total circumference of the lattice, but the still the central portion is sufficiently near (only a few links in the hyperbolic lattice) to be kept recognizably within view in the upper right portion of the display.

Figs.6a-c illustrate the capability to display the structure of hierarchically clustered data. In this case, we have created an artificial data set with gaussian clusters centered at the node points of a "random tree" in $\mathbb{R}^5$. Each tree node had a fixed branching factor of $M = 5$, and the directions of the branches were chosen randomly, while the branching lengths and the cluster diameters were diminished by a constant factor of $f = 0.5$. The HSOM training was performed by randomly sampling the nodes of this tree (we used $L = 3$ levels, yielding 31 clusters, identified by labels *1*, *11* ... *15*, *111* ... *155*). Again we compare a standard SOM (hexagonal lattice, leftmost picture (a)) with a hyperbolic SOM (middle and right pictures (b) and (c)), using the projections explained previously. Apparently, the periphery of the HSOM (middle) provides a somewhat better resolution of the cluster structure of the data. In addition, parts of the tree that are closer to the root (the more "general items") tend to become mapped centrally (although they only occupy relatively few nodes, since they are comparably small in number), with the branches of the tree arranged towards the periphery. Fig. 6c (right) finally provides a zoomed view of a subregion in the periphery of the HSOM. We clearly see that the globally visible

arrangement of the hierarchical structure is continued here at a smaller scale, providing a spatial layout of the *"13"*- and the *"15"*-subtrees adjacent to the *1*-root node.

7. DISCUSSION

Much of the enormous success of the SOM is based on its remarkable ability to translate high-dimensional similarity relationships among data items that in their raw form may be rather inaccessible to us into a format that we are highly familiar with: into spatial neighborhoods within a low-dimensional space. So far, we have been accustomed to use only euclidean spaces for that purpose. But mathematicians have prepared everything that we need to extend the realm of SOMs into non-euclidean spaces. With this contribution, we hope to have demonstrated that this may offer more than just an exotic exercise: besides the more obvious uses of spherical SOMs the newly proposed hyperbolic SOMs are characterized by very interesting scaling properties of their lattice neighborhood that makes them promising candidates to better capture the structure of very high-dimensional or hierarchical data. We have shown how regular tesselations of the hyperbolic plane H2 can be used to construct a SOM that projects onto discretized H2 and we have presented initial simulation results demonstrating that this "hyperbolic SOM" shares with the standard SOM the robust capability to unfold into an ordered structure. We have compared its properties with the standard SOM and found that the exponential node growth towards its periphery favors the formation of mappings that are structured differently than for the euclidean SOM. Our initial simulations indicate that the faster increasing hyperbolic neighborhood can indeed facilitate the construction of space-filling map configurations that underly the dimension reduction in SOMs and that hyperbolic SOMs can visualize hierarchical data, benefitting at the same time from techniques introduced in previous work on the exploitation of hyperbolic spaces for creating good displays of tree-like structures.

So far, this is only a very modest beginning and many challenging questions for hyperbolic SOMs are still open. The next steps will require to study the properties of HSOM for larger, real world data sets. Here, interesting candidates should be data resulting from various *branching processes* (which includes textual data), since they are usually described by an exponentially diverging flow in their state space, which might fit better onto a hyperbolic space than onto the euclidean plane. From a mathematical point of view, it will be necessary to revisit many of the questions that have been addressed for the euclidean SOM in the past, in particular the computation of density laws. An important point in this regard is the study of edge effects, which are much stronger for hyperbolic SOMs since they are by construction such that from any point there is always a (logarithmically) short path leading to the border. Of course, there will also be the ordering issue, but there is little reason to assume simpler answers when the underlying lattice is a H2-tesselation (this might be different, however, for the spherical SOMs; in this case, the closedness of the surface may provide a strong ordering drive). Finally, the introduction of a hyperbolic map space for SOMs suggests similar generalizations for related types of mapping algorithms, an issue which we intend to address in future work.

REFERENCES

1. Christopher M. Bishop, Markus Svenson, and Christopher K. I. Williams. GTM: The generative topographic mapping. *Neural Computation*, 10(1):215–234, 1998.
2. D.S. Bradburn. Reducing transmission error effects using a self-organizing network. In *Proc. of the IJCNN89*, volume II, pages 531–538, San Diego, CA, 1989.
3. H. S. M. Coxeter. *Non Euclidean Geometry.* Univ. of Toronto Press, Toronto, 1957.
4. R. Fricke and F. Klein. *Vorlesungen über die Theorie der automorphen Funktionen*, volume 1. Teubner, Leipzig, 1897. Reprinted by Johnson Reprint, New York, 1965.
5. Thore Graepel and Klaus Obermayer. A stochastic self-organizing map for proximity data. *Neural Computation*, 11(1):139–155, 1999.
6. Timo Honkela, Samuel Kaski, Krista Lagus, and Teuvo Kohonen. WEBSOM—self-organizing maps of document collections. In *Proceedings of WSOM'97, Workshop on Self-Organizing Maps, Espoo, Finland, June 4-6*, pages 310–315. Helsinki University of Technology, Neural Networks Research Centre, Espoo, Finland, 1997.
7. F. Klein and R. Fricke. *Vorlesungen über die Theorie der elliptischen Modulfunktionen.* Teubner, Leipzig, 1890. Reprinted by Johnson Reprint, New York, 1965.
8. T. Kohonen. *Self-Organizing Maps.* Springer Series in Information Sciences. Springer, second edition edition, 1997.
9. Teuvo Kohonen, Samuel Kaski, and Harri Lappalainen. Self-organized formation of various invariant-feature filters in the adaptive-subspace som. *Neural Computation*, 9(6):1321–1344, 1997.
10. Pasi Koikkalainen and Erkki Oja. Self-organizing hierarchical feature maps. In *Proc. of the IJCNN 1990*, volume II, pages 279–285, 1990.
11. John Lamping and Ramana Rao. Laying out and visualizing large trees using a hyperbolic space. In *Proceedings of UIST'94*, pages 13–14, 1994.
12. John Lamping, Ramana Rao, and Peter Pirolli. A focus+content technique based on hyperbolic geometry for viewing large hierarchies. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, Denver, May 1995. ACM.
13. W. Magnus. *Noneuclidean Tesselations and Their Groups.* Academic Press, 1974.
14. Charles W. Misner, J. A. Wheeler, and Kip S. Thorne. *Gravitation.* Freeman, 1973.
15. Frank Morgan. *Riemannian Geometry: A Beginner's Guide.* Jones and Bartlett Publishers, Boston, London, 1993.
16. H. Ritter, T. Martinetz, and K. Schulten. *Neural Computation and Self-organizing Maps.* Addison Wesley Verlag, 1992.
17. Helge Ritter and Teuvo Kohonen. Self-organizing semantic maps. *Biol. Cybern.*, 61:241–254, 1989.
18. D. De Sieno. Adding a conscience to competitive learning. In *Proc. of the ICNN88*, volume I, pages 117–124, San Diego, CA, 1988.
19. Karl Strubecker. *Differentialgeometrie III: Theorie der Flächenkrümmung.* Walter de Gruyter & Co, Berlin, 1969.
20. J.A. Thorpe. *Elementary Topics in Differential Geometry.* Springer-Verlag, New York, Heidelberg, Berlin, 1979.
21. Jörg Walter and Helge Ritter. Rapid learning with parametrized self-organizing maps. *Neurocomputing*, pages 131–153, 1996.