A Contribution towards Graph Embedding in Symmetric Spaces

Submitted by:
Clemens FRUBÖSE
born in Berlin

Supervised by:
Prof. Dr. Ullrich KÖTHE
Prof. Dr. Anna WIENHARD
Prof. Dr. Christoph SCHNÖRR

Abstract

In so-called information age many big data sets have the form of graphs. Using a meaningful embedding into a matching geometric space, it is possible to infer information about the graph from its structure within the embedding space. Due to the natural connection of hierarchical structure and spaces of negative sectional curvature, embedding in hyperbolic space has recently received much attention. Since a graph very often also possesses non-hierarchical structure at the same time, it is promising to consider embedding spaces of non-constant sectional curvature. For this purpose symmetric spaces of higher rank come into one's mind.

In the first part of this thesis an introduction to symmetric spaces is given. As simple cases of symmetric spaces with non-constant sectional curvature, Cartesian products are examined thoroughly with respect to their sectional curvature and totally geodesic submanifolds.

By means of Siegel's upper half-space, the consideration of symmetric spaces is motivated with the versatility of totally geodesic submanifolds contained therein. An implementation of the Riemann *gradient descent* method is developed.

In the second part, different variants of this optimization algorithm are examined and evaluated using test graphs. Since previous embedding algorithms have considered working with a preprocessed embedding, but have not yet implemented that, this procedure is investigated experimentally.

In order to be interpretable, it is essential that the embedding maintains the structure of the graph. Since current optimization methods are not (yet) able to point out this structure reliably, it is suggested that the "entanglement" of the embedded graph is addressed at first such that further optimization can be successful.

Finally, with regard to these results, suggestions are made for a further improvement of these algorithms in order to enable a future implementation of a *structure-preserving* algorithm to embed graphs in general symmetric spaces.

ZUSAMMENFASSUNG

Im sogenannten Informationszeitalter liegen große Datensätze in Form von Graphen vor. Durch eine gute Einbettung in einen passenden geometrischen Raum ist es möglich, aus der Struktur des Graphen in diesem Raum Informationen über den Graphen selbst zu erhalten. Da ein natürlicher Zusammenhang zwischen hierarchischen Strukturen und Räumen negativer Schnittkrümmung besteht, wird insbesondere der Einbettung in den hyperbolischen Raum in jüngster Zeit viel Beachtung geschenkt. Da ein Graph sehr häufig auch nicht-hierarchische Strukturen enthält, erscheint es vielversprechend, als Einbettungsraum Räume nicht-konstanter Schnittkrümmung zu betrachten. Dafür bieten sich symmetrische Räume höheren Rangs an.

Im ersten Teil dieser Arbeit wird daher eine Einführung in die Theorie der symmetrischen Räume gegeben. Als einfacher Fall symmetrischer Räume mit nicht-konstanter Schnittkrümmung werden ausführlich Cartesische Produkträume hinsichtlich der Schnittkrümmung und enthaltenen total-geodätischen Untermannigfaltigkeiten untersucht.

Anhand des Siegel'schen Halbraums wird die Betrachtung symmetrischer Räume mit der Vielseitigkeit der darin enthaltenen total-geodätischer Unterräume motiviert und eine Implementierung des Riemannschen *gradient descent* Verfahrens erarbeitet.

Im zweiten Teil werden anhand von Testgraphen unterschiedliche Varianten dieses Optimierungsalgorithmus' untersucht und bewertet. Da bisherige Einbettungsalgorithmen eine Starteinbettung zwar in Betracht gezogen, jedoch noch nicht umgesetzt haben, wird dieses Vorgehen experimentell untersucht.

Um eine Interpretierbarkeit der Einbettung zu ermöglichen, ist die Erhaltung der Struktur des Graphen essenziell. Da aktuelle Optimierungsmethoden (noch) nicht in der Lage sind, diese Struktur zuverlässig herauszuarbeiten, wird vorgeschlagen, die "Verknäuelung" des eingebetteten Graphen zunächst anzugehen, damit eine weitere Optimierung erfolgreich sein kann.

Abschließend werden hinsichtlich dieser Resultate Vorschläge für eine weitere Verbesserung der Algorithmen gemacht, um eine zukünftige Implementierung eines *strukturerhaltenden* Algorithmus' zur Einbettung von Graphen in allgemeine symmetrische Räume zu ermöglichen.

ACKNOWLEDGEMENTS

Foremost, I thank my supervisors Ullrich Köthe, Anna Wienhard and Christoph Schnörr for giving me the opportunity to write this thesis, which I really enjoyed due to its balancing between Mathematics and Informatics. In particular, I want to thank Ullrich Köthe for guiding me through the process and for many discussions and stimuli.
Secondly, I want to thank Jonas Beyrer who spent hours of discussion and explanation of mathematical details with me.

I am grateful for the pleasant working atmosphere at the office at IWR and thus want to thank all group members for patiently answering questions about neural networks which I had to learn in.
I also want to thank Julian Chorus and Joachim Fruböse for kindly finding (lots of) typos.
My last thanks goes to Mikio Nakahara for his excellent textbook on Differential Geometry, Topology and Physics which taught me a lot of intuition behind concepts of differential geometry.

Contents

1 Introduction

Learning representations of graphs into embedding spaces has become a crucial task of modern data science. Applications include classical graph operations such as community detection, link prediction or finding the shortest path within the network:

A meaningful embedding mirrors similarity between objects by their distance in the embedding space: As a small distance in the embedding space shows that these data points are related, it is promising to consider link prediction and community detection within the embedding space. Furthermore, embedding a network into space transforms combinatorial problems to geometrical problems: The shortest path between two nodes in a graph can in principle only be found by means of combinatorics, whereas a particular embedding enables the application of a *greedy algorithm* based on the vertex positions in the embedding space.

Additionally, a representation of a graph can also deliver valuable information about the data itself. This idea dates back at least 60 years to the idea of multidimensional scaling introduced by [Torgerson, 1952].

Due to the growing power of machine-learning, representation learning has gained lots of attention in the past years. A prominent example is natural language processing performing similarity tasks, e.g. inferring the sentimental context of words.

The machine-learning community has also become interested in the topic, since a meaningful representation of data in the latent space improves the explainability of neural networks.

For a long time, the default embedding space has been Euclidean space. To capture properties of the graph more closely, one has to adapt the geometry of the embedding space to the problem, i.e. embed graphs into curved spaces.
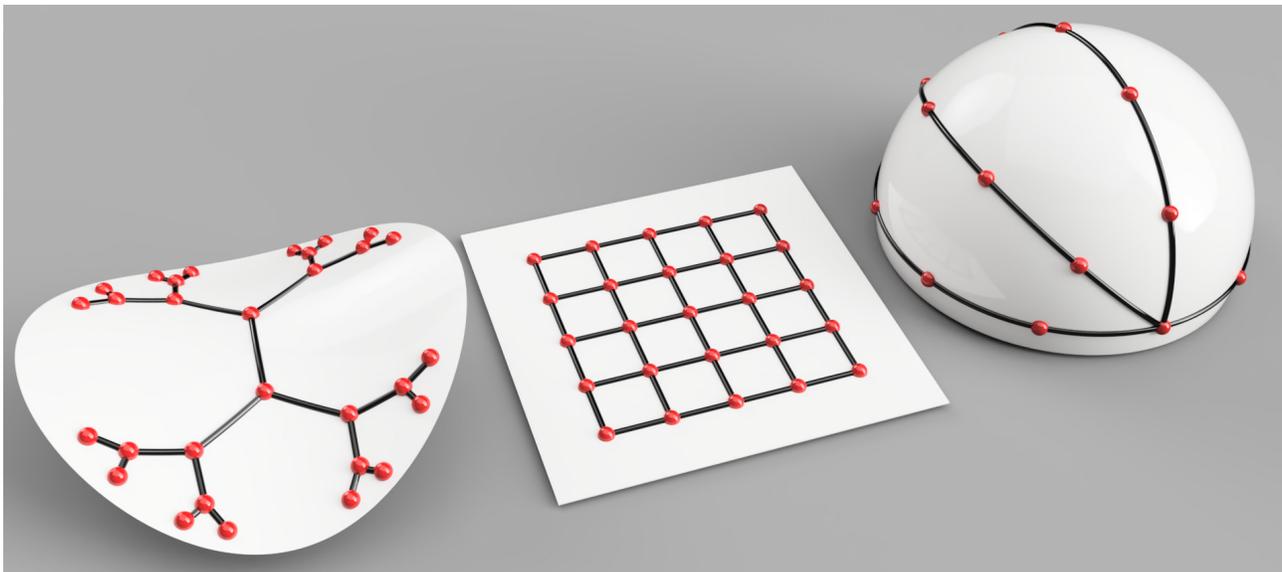
Figure 1: Illustration of the correspondence between a graph and its natural curved embedding space. On the left, a hierarchical structure is embedded on a negatively curved surface. In the middle, a grid-like structure is embedded on a flat surface. On the right, a graph with loops is embedded on a surface of positive curvature.

Choosing a curved embedding space can be motivated by the following observation: The number of vertices in a tree graph grows exponentially with the depth of that tree. Moreover, circles in hyperbolic space also grow exponentially with their radius. This connection between tree-like structures and hyperbolic space has been investigated by mathematicians for a long time - most notably by GROMOV in [Gromov, 1987]. Theoretical work (such as [Kleinberg, 2007]) introduced that concept to the data science community and suggested to choose hyperbolic space as representation space for tree-like structures. In particular, an embedding in hyperbolic space can preserve the distances within nodes of a tree very well [Sarkar, 2011]. This idea has led to great success of embedding real world networks.

The reason for that great success is that lots of real world networks are essentially scale-free[1] networks [Adcock et al., 2013]. These graphs posses an underlying hierarchical structure [Ravasz and Barabási, 2003] and are hence well-suited to hyperbolic space. The connection between scale-free networks and hyperbolic space has also been shown by proving that randomly connected nodes in hyperbolic space naturally form scale-free graphs [Gugelmann et al., 2012].

Despite the great success of embeddings into hyperbolic space, there are graph properties which cannot be represented accurately in hyperbolic space: Hierarchical data need not be belonging to *one hierarchy only*. There can be different hierarchies at once to which a node might belong. An example in terms of a social network could be the president of the fire department who is at the same time just an ordinary member of a bowling club without special responsibility. The underlying space should therefore be able to embed different hierarchies at once, i.e. it should possess independent hyperbolic subspaces. Additionally, a network does not have to be only hierarchical. In particular there are networks which are known to fit best into flat spaces - such as grids. Of course, one should also not be constrained to flat and hyperbolic spaces only: A three-dimensional tilting vector for example can be represented best on a two-dimensional spherical subspace.

Rather than again improving the methods how to embed networks in hyperbolic space, the idea leading to this work is to take the next logical step and to further adapt the geometry of the embedding space to the problem. This should done by choosing a different, more general manifold from the class of symmetric spaces.

Figure 2: Illustration of a graph which does not fit perfectly in hyperbolic space due to its loops. Every vertex belongs to two structures at the same time, namely the hierarchical order (indicated by black edges) and the cyclical order (indicated by blue edges).

Symmetric spaces can have subspaces of different curvature at the same time which addresses the problems stated above. Additionally, these spaces are as smooth manifolds usable for operations within the embedding space.

Embedding into a symmetric space can be a very powerful tool, because a unified embedding in differently curved subspaces simplifies the interpretation of the network significantly: A promising application of network interpretability is a network in the latent space of a neural network. This Euclidean space of high dimension could be transformed to a symmetric space of fewer dimensions, such that the curvature of the subspaces provides information about the meaning of features. For instance, one might think about image recognition: The orientation of an object could be represented in a spherical subspace, the position in an Euclidean subspace and the size (i.e. the amount of the picture the object occupies) in a hyperbolic subspace.

---

[1]A scale-free network is defined as network whose degree distribution follows a power law: $P(k) \sim k^{-\gamma}$. Scale-free networks possess so-called "hub nodes" which dominate the graph structure due to their enormous amount of edges. They serve as hubs for travelling within the graph and illustrate the hierarchical structure of the network.

Outline of the thesis

This thesis is split in two parts:

1. A general treatment of symmetric spaces

2. An experimental examination of graph embedding in curved manifolds.

Both parts can be read independently from each other. However, they are conceptually related: The survey on symmetric spaces focuses on sectional curvature and totally geodesic submanifolds. These concepts are important for network embedding.
The embedding part points out the connection between graphs and their natural embedding space.

Symmetric spaces (in their full generality) have not received much attention from data science yet. This justifies the introduction to symmetric spaces aiming at a broad readership and hence trying to give an intuitive approach. For this reason, details which might hamper intuition are omitted while being sufficiently mathematically sound. To simplify reading, table 1 serves as an index of notation.

The second part gives a short introduction to network embedding. The prevalent embedding procedure -using the so-called gradient descent algorithm- is revisited. Starting from the implementation [Nickel and Kiela, 2017] improvements are proposed and examined using own experimental results.
An outlook to further research is given. This in particular includes using algorithms developed in the first part about symmetric spaces.

An Introduction to Symmetric Spaces

What makes a space a symmetric space?

A symmetric space is a Riemannian manifold M which possesses for every point $p \in M$ an isometric map s_p -called symmetry- with special property of being equivalent to the geodesic[2] reflection at p. A geodesic reflection at p is basically the "normal" point reflection at p: Every geodesic γ with $\gamma(0) = p$ is mapped by s_p such that for a neighbourhood of p $s_p(\gamma(t)) = \gamma(-t)$ holds.
A geodesic reflection exists in any Riemannian manifold, but in general it is not an isometry, i.e. it does not leave distances and angles invariant.
To simplify visualization of the following concepts, the well-known symmetric spaces $\mathbb{R}^n$, S^n and H^n are revisited in section 3.

2 Getting a feel for curvature

Curvature is going to play an important role for the study of symmetric spaces. For this reason, a summary of important tools and their application is given with emphasis on their illustration. This is covered in much more detail by the great illustrative source [Nakahara, 2003, Chapter 7]. Readers familiar with these notions may skip this section.

2.1 Tangent vectors

Since we are familiar with Euclidean space $\mathbb{R}^n$ -which is a vector space- one might be sloppy in differentiating vectors from points, since their correspondence is obvious. However, a manifold is in principle only a set of points which looks like $\mathbb{R}^m$ in a very tiny neighbourhood of an arbitrary point.
For every point we can define the tangent space as an ordinary flat vector space $\mathbb{R}^n$. The tangent space represents the local flatness of M. At a point p the tangent space of the manifold M is written T_pM. Since this tangent space is a vector space, one can find a basis for it. A vector v can hence be expressed as:

$$v = \sum_{\alpha}^{\dim(M)} v_\alpha e(p)_\alpha, \ \ v \in T_pM. \tag{1}$$

The basis vectors e_i are dependent on the point to which the space is tangent, as the tangent spaces of different points are different spaces.
The following will use the abstract index notation instead of the notation above, which is used in differential geometry and Physics. It makes use of the Einstein sum convention, i.e. the same indices up and down are implicitly summed over.[3] If the components of a vector have superscripts, the vector belongs to the tangent space, a vector with subscript on its components belongs to the cotangent space[4]. The basis vectors of the tangent space are denoted by ∂_i which is shorthand for ∂_{x^i}. This basis vector is pointing in the direction of the coordinate x^i [5]. The notation for a vector v becomes:

$$v = v^\alpha \partial_\alpha := \sum_{\alpha}^{\dim(M)} v^\alpha \partial_\alpha. \tag{2}$$

---

[2]A geodesic is a path on a manifold which is a path whose velocity vector does not change along the path. It is the generalization of a straight path in a curved space. More details below in section 2.6.

[3]The superscript must not be confused with an exponent. To avoid ambiguity brackets are imposed if necessary.

[4]the dual space of the tangent space, see section 2.3

[5]A vector acts as a directional derivative at its base point. Magnitude and direction of that vector determine its action on a smooth scalar function on that manifold.

So, for each point there is a Euclidean vector space $\mathbb{R}^m$ with m corresponding to the dimension of M. Although the vector spaces at p and q are isomorphic, it does not make sense to compare vectors from T_pM and T_qM to each other, since the coordinates of the tangent vector spaces need not fit together. The set of all tangent spaces at M is called TM. It is of dimension $2\dim(M)$, since it consists of pairs (p,v), $p \in M$, $v \in T_pM$ which are both of dimension $\dim(M) = m$.
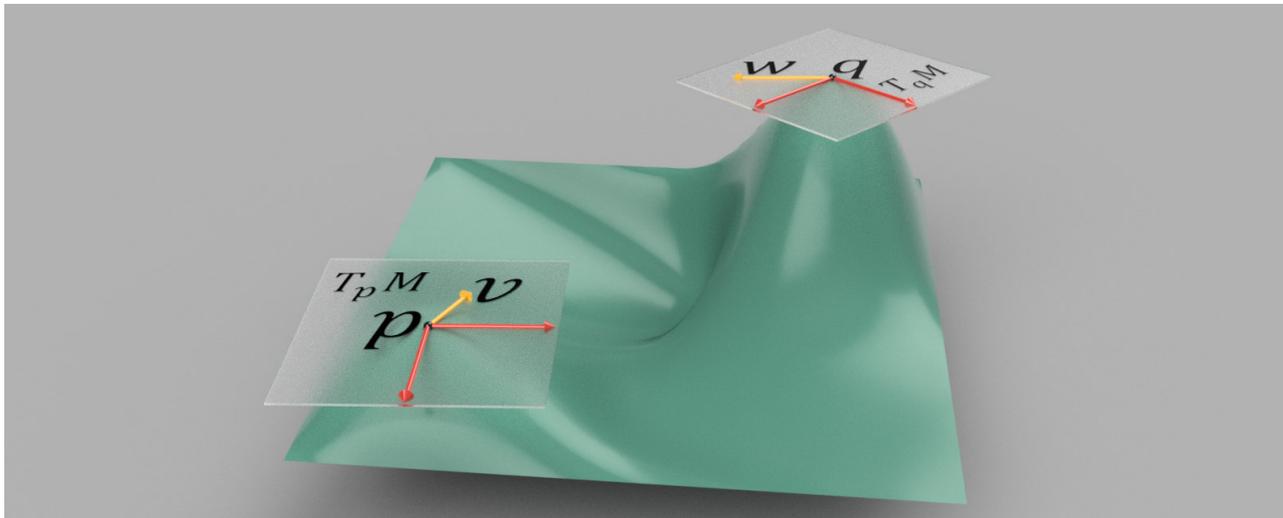
Figure 3: Two tangent spaces T_pM and T_qM with vectors v and w on a manifold M. The two tangent spaces are each isomorphic to $\mathbb{R}^2$, but since they are located at different points, the term $v + w$ does not make sense because the orientation of the basis vectors (symbolized by red arrows) in general does not fit.

2.2 The metric tensor g

A Riemannian manifold (and hence also a symmetric space) possesses a metric tensor, often denoted by g. This tensor is defined on all of the manifold and indicates the way to calculate the scalar product of two vectors. In contrast to Euclidean space, the scalar product is in general dependent on the point p where it is evaluated. Both vectors need to be tangent to the manifold at p. The scalar product of two vectors, the first a tangent vector at p, the second tangent to q is not defined.

The tensor field g is the central quantity to characterize the curvature, since all curvature tensors are derived from g.

Recap:

Let us recall the default metric on Euclidean space $\mathbb{R}^3$:

$$g(p)(\cdot, \cdot) = 1 \cdot dx \otimes dx + 1 \cdot dy \otimes dy + 1 \cdot dz \otimes dz \tag{3}$$

which means that the scalar product of two vectors v, w in e.g. $\mathbb{R}^3$ is calculated as:

$$\langle v, w \rangle = \langle v^1 \partial_x + v^2 \partial_y + v^3 \partial_z, w^1 \partial_x + w^2 \partial_y + w^3 \partial_z \rangle = g(p)(v, w) = 1 \cdot v^1 \cdot w^1 + 1 \cdot v^2 \cdot w^2 + 1 \cdot v^3 \cdot w^3. \tag{4}$$

Note that the Euclidean metric tensor is constant on the manifold because the coefficients in eq. 3 are constant. The scalar product is hence not dependent on the evaluation point p. The metric tensor can also be written as a matrix scheme, i.e. the entries correspond to the coefficients in eq. 3:

$$g = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{5}$$

A general metric tensor field is defined as map:

$$g : M \times TM \times TM \to \mathbb{R} \tag{6}$$

$$(p, v, w) \mapsto g(p)(v, w), \quad v, w \in T_p M. \tag{7}$$

The metric is a $(0, 2)$-tensor field which means that one needs to plug in zero covectors[6] and two tangent vectors to obtain a real number. A metric tensor field $\tilde{g}(p)$ could be:

$$\tilde{g}(p_0) = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 0 & 0 \\ 0 & 0 & f^2(p_0) \end{pmatrix} = 1 \cdot dx \otimes dx + 2 \cdot dx \otimes dy + 2 \cdot dy \otimes dx + f^2(p_0) \cdot dz \otimes dz \tag{8}$$

with

$$\tilde{g}(p_0)(v, w) = 1 \cdot v^1 \cdot w^1 + 2 \cdot v^1 \cdot w^2 + 2 \cdot v^2 \cdot w^1 + f^2(p) \cdot v^3 \cdot w^3, \quad \text{because } dx^i \partial_{x^j} = \delta^i_j. \tag{9}$$

The metric is written in index notation as:

$$g = g_{\alpha\beta} dx^\alpha dx^\beta. \tag{10}$$

The two lower indices on $g_{\alpha\beta}$ indicate that it is a $(0, 2)$-tensor. By convention, the tensor product sign "$\otimes$" as in eq. 8 is dropped. The values of $g_{\alpha\beta}$ can be read off from the matrix form as e.g. in eq. 8. The differentials dx^i are basis vectors of the cotangent space $T_p^* M$ introduced below.

In index notation, the scalar product has the form:

$$g(p)(v, w) = g_{\alpha\beta} dx^\alpha dx^\beta v^\mu \partial_\mu w^\nu \partial_n u = g_{\alpha\beta} v^\mu w^\nu (dx^\alpha \partial_\mu)(dx^\beta \partial_n u) = g_{\alpha\beta} v^\mu w^\nu \delta^\alpha_\mu \delta^\nu_\beta = g_{\alpha\beta} v^\alpha w^\beta. \tag{11}$$

with usual Kronecker delta and the dependence on p dropped for simplicity.

2.3 Covectors

As each of the tangent vector spaces of a manifold of dimension m is isomorphic to the vector space $\mathbb{R}^m$, linear algebra states that there is a dual vector space of dimension m. The dual vector space is the space of linear maps sending a vector to a scalar. If we consider the space $T_p M$ as vector space, its dual space is written $T_p^* M \cong \mathcal{L}in\mathcal{M}aps(T_p M; \mathbb{R})$. In abstract index notation, a covector w is written:

$$w = w_\alpha dx^\alpha := \sum_\alpha^{\dim(M)} w_\alpha dx^\alpha. \tag{12}$$

The covector's coefficient has a lower index, in contrast to the upper index of a vector as in eq. 2.
An intuitive understanding of a covector is given in the figure below:

---

[6]See section 2.3 below.
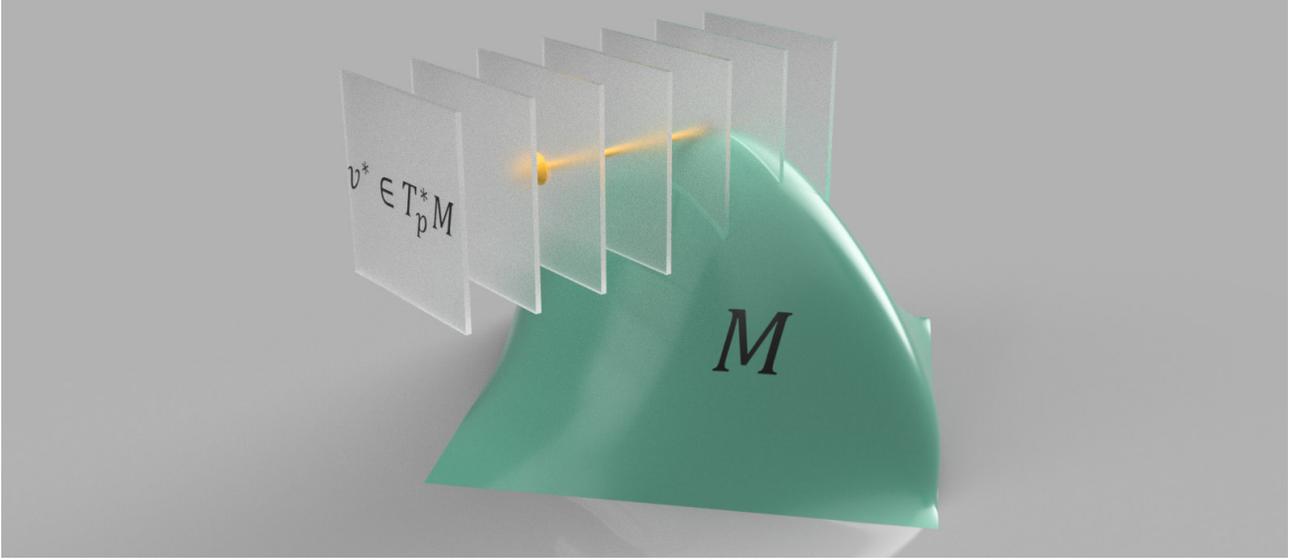
Figure 4: The concept of a covector v^* can be understood as planes with normal vector v. The density of those planes corresponds to the magnitude of the covector v^* living in T_p^*M. The product of a covector with a vector determines "how many" planes the vector intersects. In fact, since a covector is a linear map from T_pM to $\mathbb{R}$, their product is a real number. In this sense, one should understand the planes as densities, since the result does not have to be a natural number.

2.4 The duality of tangent space and cotangent space

Since $(T_p^*M)^* \cong T_pM$, the spaces are closely related to each other. In fact, the metric tensor provides the natural map between them: Recall that the metric tensor g maps two vectors to a scalar. This allows to map a vector $v \in T_pM$ to its dual $v^* \in T_p^*M$:

$$\flat : T_pM \to T_p^*M = \mathcal{L}in\mathcal{M}aps(T_pM; \mathbb{R})$$
$$v \mapsto g(v, \cdot) =: v^*. \tag{13}$$
In coordinate form: $v^\alpha \partial_\alpha \mapsto g_{\alpha\beta} v^\alpha dx^\beta =: v_\beta dx^\beta.$

In fact, v^* [7] is a covector, since $\flat(v) = g(v, \cdot)$ linearly maps a vector to a scalar. In index notation:

$$v_\beta dx^\beta w^\gamma \partial_\gamma = v_\beta w^\gamma \delta_\gamma^\beta = v_\beta w^\beta = g(v, w) = w_\beta v^\beta. \tag{14}$$

The inverse metric tensor g^{-1} maps two covectors to a scalar. Using this, the inverse operation of $\flat$ is also possible:

$$\sharp : T_p^*M \to T_pM = \mathcal{L}in\mathcal{M}aps(T_p^*M; \mathbb{R})$$
$$v \mapsto g^{-1}(v, \cdot) \tag{15}$$
in coordinate form: $v_\alpha dx^\alpha \mapsto g^{\alpha\beta} v_\alpha \partial_\beta =: v^\beta \partial_\beta.$[8]

In fact, since $g^{-1}(v, \cdot) \in \mathcal{L}in\mathcal{M}aps(T_p^*M; \mathbb{R})$, $\sharp(v)$ is a vector.

---

[7]To avoid confusion with complex conjugate numbers, we will not use the "$*$" symbol. Instead, we use the index notation with lower index on the component and upper index on the basis covector.

[8]Note that $g^{\alpha\beta}$ is the inverse metric tensor $g^{-1}(\cdot, \cdot)$, since it is a $(2, 0)$-tensor, whereas $g_{\alpha\beta}$ is a $(0, 2)$-tensor.

Cotangent space in the language of Physics:
The function of the metric tensor becomes clearer in "bra/ket" notation: Let us use a vector space V and its dual V^*: A regular vector is a ket element and a covector is a bra element. The metric tensor then has the form:

$$g(|v\rangle, |w\rangle) = \langle v| |w\rangle = \langle v, w\rangle \text{ and } g^{-1}(\langle v|, \langle w|) = \langle v| |w\rangle = \langle v, w\rangle. \tag{16}$$

The metric tensor thus sends a vector to its covector by:

$$NV \ni |v\rangle \mapsto g(|v\rangle, \cdot) = \langle v| \in V^* \text{ which is a covector.} \tag{17}$$

The opposite direction is also possible via:

$$V^* \ni \langle v| \mapsto g^{-1}(\langle v|, \cdot) = |v\rangle \in V \text{ which is a regular vector.} \tag{18}$$

2.5   The covariant derivative

The covariant derivative is a directional derivative sending functions to functions and tensors to tensors. It is the natural generalization of the directional derivative from Euclidean space. It takes into account that not only the components of a tensor may change, but also that the basis vectors change because of the curvature of the manifold and thus can be understood as total derivative.
The direction of this derivative follows a vector. Since a vector v is always located at one point (e.g. the point p), the covariant derivative measures the change of a quantity in the direction of v at point p.
The covariant derivative of a vector field X along a vector field V is thus:

$$\nabla_V X = (V^\alpha \partial_\beta X^\beta)\partial_\alpha + X^\alpha V^\beta \Gamma^\gamma_{\alpha\beta}\partial_\gamma \tag{19}$$

which is implicitly dependent on the evaluation point p at which both V and X must be defined. The first part corresponds to the derivative of the components of X, the second part corresponds to the derivative of the basis vectors ∂_i. In Euclidean space, the second part vanishes. The symbol Γ is called Christoffel symbol and keeps track of the derivative of the basis vectors. The Christoffel symbols are also a field, i.e. they are dependent on the evaluation point.

2.6   Parallel transport

Using the covariant derivative, we can introduce the concept of parallel transport along a curve. Pictorially, a vector field X is said to be parallel along a curve $\mathcal{C}(t)$, if its magnitude and the angle between X and the velocitiy vector of the curve are constant at every point of the curve. This is the case, if the covariant derivative with respect to the velocity vector of the vector X is zero. More formally, consider a curve $\mathcal{C}$

$$\begin{aligned} \mathcal{C} : \mathbb{R} &\to M \\ t &\mapsto \mathcal{C}(t) \end{aligned} \tag{20}$$

and its velocity vector field[9]

$$\begin{aligned} \mathcal{C}_* : \mathbb{R} &\to TM \\ t &\mapsto (\partial_t \mathcal{C}^i(t))\partial_i =: \dot{\mathcal{C}}, \ \dot{\mathcal{C}} \in T_{\mathcal{C}(t)}M. \end{aligned} \tag{21}$$

One searches a vector field, which is parallel, i.e.

$$\nabla_{\dot{\mathcal{C}}} X = 0, \tag{22}$$

_____

[9]Note that this vector field is only defined at the image of $\mathcal{C}$, since the other points of M are not reached by the curve. Thus one cannot define a velocity vector at every point of M.

which is an ordinary differential equation. If one wants to parallel transport a vector $v \in T_p M$ to the point q along a curve γ, one solves the above equation uniquely with the initial condition $X(p) = v$. The vector $X(\gamma(t_1))$ with $\gamma(t_1) = q$ is then the parallel tranported vector at q.

Note that the outcome of the parallel transport is dependent on the curve. In particular, this dependency will be used to determine the curvature of a manifold in section 2.8.

2.7 Geodesics and the Riemann exponential map

The notion of parallel transport allows to define a geodesic curve as "straightest possible" path. More formally, the velocity vector of a geodesic curve is parallel along that very curve, i.e.

$$\nabla_{\dot{\mathcal{C}}} \dot{\mathcal{C}} = 0. \tag{23}$$

Translating this equation to coordinates as in eq. 19 with $V = X = \dot{\mathcal{C}}$ yields:

$$\left(\frac{\mathrm{d}^2 \mathcal{C}^\gamma}{\mathrm{d}t^2} + \frac{\mathrm{d}\mathcal{C}^\alpha}{\mathrm{d}t^2} \frac{\mathrm{d}\mathcal{C}^\beta}{\mathrm{d}t} \Gamma^\gamma_{\alpha\beta} \right) \partial_\gamma \tag{24}$$

This differential equation can (locally) be solved uniquely with initial conditions for a starting point and a starting velocity vector as a result of the Picard-Lindelöf-theorem for ordinary differential equations.

The Riemann exponential map is defined implicitly using the solution to the geodesic equation eq. 24:

$$\exp : M \times TM \to M$$
$$(p, v) \mapsto \exp_p(v) = \gamma_v(t = 1). \tag{25}$$

where γ_v is defined as the unique geodesic starting at p with initial tangent vector v. As a consequence from eq. 24, the Riemann exponential map is in principle difficult to compute explicitly.

2.8 The Riemann curvature tensor

Note that the parallel transport in flat space is trivial, i.e. path independent. The dependency of parallel transport on the chosen path thus serves as indicator of the intrinsic curvature of the manifold. To determine the curvature at a point p, one has to parallel transport a vector along a path following the first vector field and then following the second vector field and vice versa.

The Riemann curvature tensor maps three vector fields to another vector field and is hence a $(1, 3)$-tensor:

$$R : \mathcal{X}(M) \times \mathcal{X}(M) \times \mathcal{X}(M) \to \mathcal{X}(M)$$
$$(X, Y, Z) \mapsto (\nabla_X \nabla_Y - \nabla_Y \nabla_X - \nabla_{[X,Y]})Z =: R(X, Y)Z \tag{26}$$
$$\text{with } \mathcal{X}(M) \text{ the space of smooth vector fields on } M.$$

The Riemann curvature tensor has in the order of $\frac{1}{12} n^4$ independent entries for $\dim(M) = n$ and is thus a very general concept of curvature. More specific curvature quantities are e.g. the sectional curvature or the Ricci scalar. The latter indicates volume growth averaged over all directions, the former is defined below.
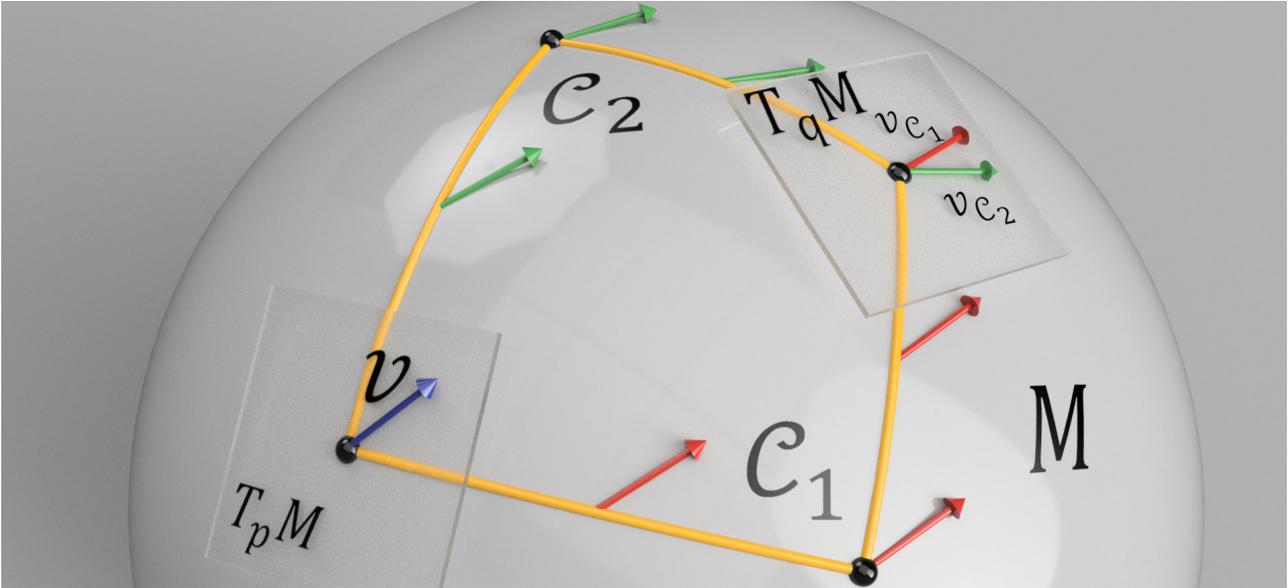
Figure 5: Illustration of infinitesimal parallel transports along a tiny rectangle. The horizontal and vertical pieces of the paths are induced by the flow of X and Y, respectively. The blue vector v at p is parallel transported along the curves $\mathcal{C}_1$ and $\mathcal{C}_2$. The difference between the resulting vectors $v_{\mathcal{C}_1}$ and $v_{\mathcal{C}_2}$ is proportional to the Riemann curvature tensor. Infinitesimally, it corresponds to the curvature tensor at p.

2.9 The sectional curvature

Instead of having a notion of the curvature in all directions at once, sectional curvature gives the scalar value of the curvature of a two-dimensional subspace spanned by two vectors emanating from the same point. Imagining a 2-dimensional surface in 3-dimensional space, the sectional curvature corresponds to the size of spheres or saddles which snuggle to the surface. It becomes apparent that this quantity is dependent on the evaluation point:
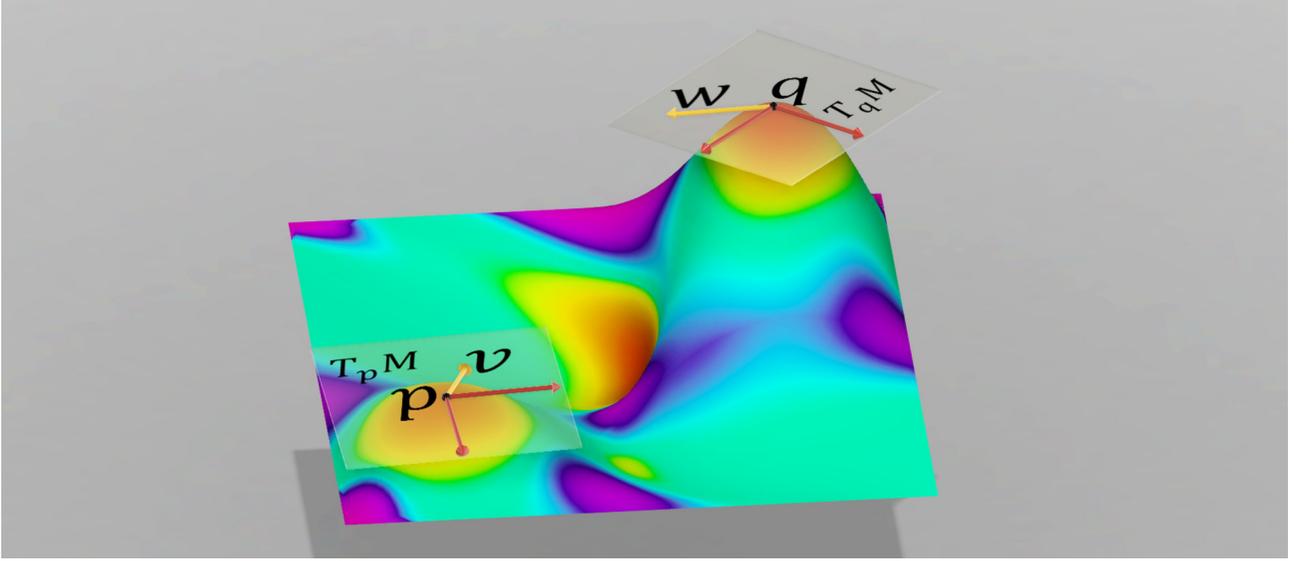
Figure 6: Illustration of the sectional curvature of the surface from fig. 3. Reddish color indicates positive curvature, blueish color indicates negative curvature. Positive curvature means that the surface in a small area looks like a sphere, negative curvature corresponds to saddle-shape of the neighbourhood. The absolute values of the curvature depend on the scale of the surface, e.g. the sectional curvature of a sphere of smaller radius is bigger than the one of big spheres.

The sectional curvature K can be computed using the Riemann curvature tensor as:

$$K(p, V, W) = \frac{g_p(R_p(V_p, W_p)W_p, V_p)}{g_p(V_p, V_p)g_p(W_p, W_p) - g_p(V_p, W_p)^2}, \quad V, W \in \mathcal{X}(M) \tag{27}$$

where g and R are dependent on p. At a point p with two tangent vectors at p, this boils down to:

$$K(v, w) = \frac{g(R(v, w)w, v)}{g(v, v)g(w, w) - g(v, w)^2}, \quad v, w \in T_pM \tag{28}$$

where g and R are implicitly evaluated at p. The quantity K is independent of the two vectors v and w due to the normalization, as long as they are linearly independent.
Euclidean space, spheres and hyperbolic space well-known spaces of constant sectional curvature. However, symmetric spaces in general do not have constant sectional curvature while maintaining a high degree of symmetry and hence conceptual simplicity.

What does the sectional curvature mean?

Sectional curvature of a space can be understood as the relation between the volume of a circle and its radius. This ratio is $\frac{\pi r^2}{r}$ for the flat space $\mathbb{R}^2$, $\frac{4\pi \sinh^2 r}{r}$ in negatively curved space H^2 and $\frac{2\pi(1-\cos r)}{r}$ for the positively curved sphere S^2. Notice that volume of a disk grows faster in hyperbolic space H^2 than in Euclidean space. Analogously, volume growth in spherical space S^2 is slower than in Euclidean space. Essentially, this is what the Ricci scalar signifies. The sectional curvature is in a way a restriction of the Ricci scalar to the two-dimensional

space which is spanned by the two vectors from eq. 28. The sectional curvature is hence more specific than the Ricci scalar curvature.
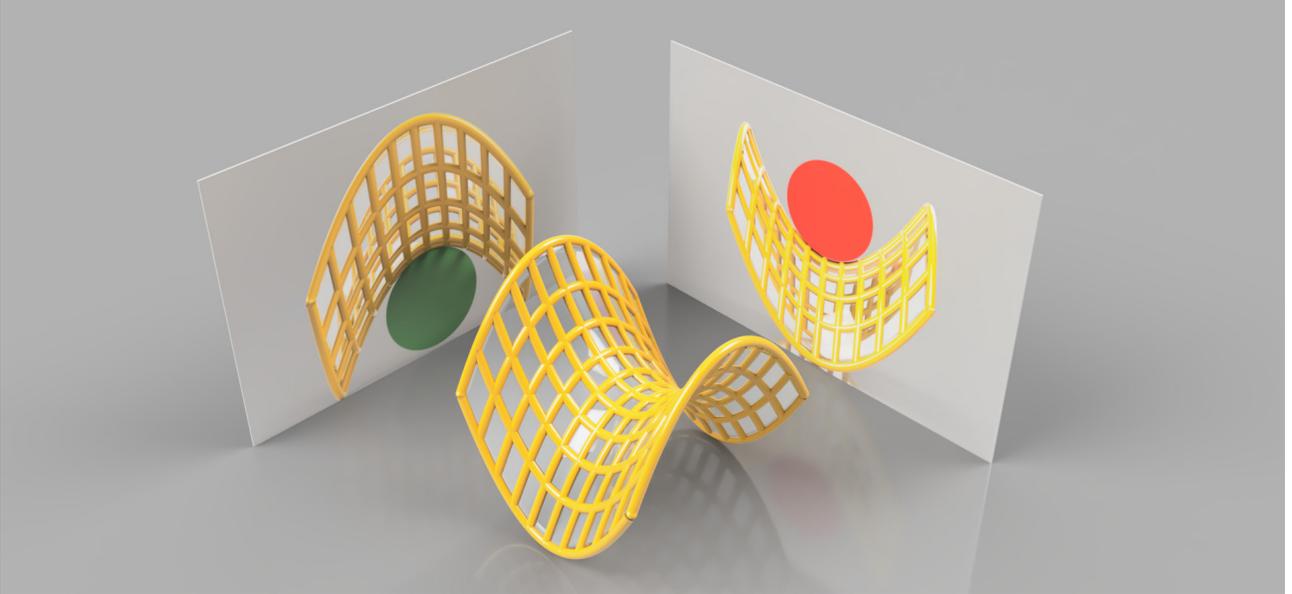
Figure 7: A curved surface and two projections with respective osculating circles at the touching point. These osculating circles have their midpoint on different sides of the surface. The curvature is hence defined to have negative sign; the surface looks like a saddle. The size of the osculating circles is an inverse measure of the magnitude of curvature. Small osculating circles result in a curvature of high magnitude. The depicted surface has overall negative sectional curvature, its magnitude is dependent on the evaluation point.

3 Euclidean, spherical and hyperbolic space

3.1 Euclidean space $\mathbb{R}^n$

Euclidean space is well-known to be a flat space. Due to its vector space structure, one can easily find the symmetries at any point p:

$$
\begin{aligned}
s_p : M &\to M \\
y = p + w &\mapsto p - w, \ \ p, v, w \in \mathbb{R}^n.
\end{aligned}
\tag{29}
$$

If one chooses p to be the origin, it becomes apparent that for every point p, s_p is a point reflection at p. Being slightly sloppy, from now on the notion of a point reflection and a geodesic symmetry should be considered identical.

The default metric on Euclidean space corresponds to the identity matrix as pointed out before. The Riemann curvature tensor of Euclidean space vanishes everywhere. It follows that the sectional curvature is also zero everywhere.

3.2 Spherical space S^n

The sphere is a space of positive curvature, it is hence bending inwards. The sphere of radius R can be *embedded* in $\mathbb{R}^{n+1}$ as $S_R^n = \{x \in \mathbb{R}^{n+1} | \, \|x\|_2 = R^2\}$ with $S_1^2 =: S^2$. This is the best-know embedding, but keep in mind that the sphere can also be embedded in $\mathbb{R}^n \cup \{\infty\}$ via the stereographic projection as in fig. 8. Although the former embedding is much more common, both embeddings are equally well suited and have their advantages. Both models will be denoted by S^n, as they represent the same space.

The common "sphere"-embedding however simplifies finding the symmetries because it enables to use the inner

product space structure of the ambient space $\mathbb{R}^{n+1}$ to construct the symmetries at a point p:

$$s_p(y) = p\langle y, p\rangle - (y - p\langle y, p\rangle) = -y + 2p\langle y, p\rangle. \tag{30}$$

Considering p and y as vectors, the first term is the component in direction of p, the second term is the remaining part perpendicular to p. This term is subtracted, i.e. reversed.
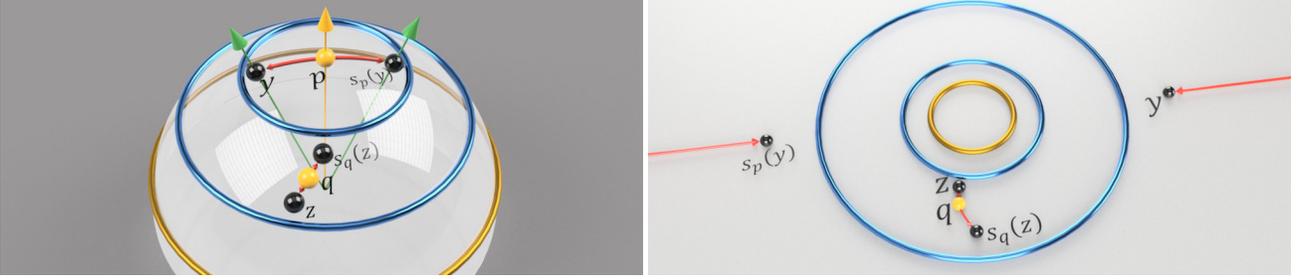
Figure 8: Two symmetries, namely s_p and s_q on the sphere S^2. On the left in the embedding in $\mathbb{R}^3$, on the right in the stereographic projection in $\mathbb{R}^2$. The blue circles help to orientate in the models. The green and yellow arrows emphasize the vector space structure of the embedding space $\mathbb{R}^3$ as used in eq. 30. The red geodesics in both models need not be straight in the Euclidean embedding spaces $\mathbb{R}^3$ and $\mathbb{R}^2$, respectively.
The point reflection is independent of a particular model; eq. 30 is hence just a realization of the symmetry with respect to the left 3-dimensional embedding. This equation does hence not hold for the right 2-dimensional embedding.

Distortion of the embedding

Note that the distance between two points in S^n is not equivalent to the distance between these points in the embedding manifold. Consider $(0,0,1)^T$ and $(0,0,-1)^T$ in $S^2 \subset \mathbb{R}^3$: The true distance is π, namely half the circumference. The embedding distance however is 2. Yet, the intrinsic length of an arc within S^2 in the embedding manifold $\mathbb{R}^3$ is identical to the length of that arc measured within $\mathbb{R}^3$. This embedding can be thought of as a non-distorting embedding. In fact, the sphere inherits its metric tensor from the Euclidean metric tensor in $\mathbb{R}^3$.
In contrast, the stereographic projection is a distorting embedding: The distance between two points measured with the metric of the embedding manifold is different from the intrinsic distance between two points of S^2. This can be taken care of by introducing a metric tensor (for $n \geq 2$):

$$g(x) = \frac{4}{(1+r^2)^2} \sum_i^n dx^i \otimes dx^i, \text{ with } r = \sqrt{\sum_i^n (x^i)^2}. \tag{31}$$

This metric tensor is obviously not inherited from the embedding space $\mathbb{R}^2$, hence this is a distorting model increasing with r corresponding to approaching the north pole of the sphere.

The sectional curvature

The sectional curvature K of a sphere of radius R can be computed via eq. 28:

$$K_R(p) = \frac{1}{R^2} \tag{32}$$

which is independent of the evaluation point p and supports the intuition of small spheres to be curved more strongly.
Scaling the metric tensor in the picture of the stereographic projection by a factor λ can be understood as multiplying all distances by λ. The stereographic projection hence represents a sphere of radius $R' = \lambda R$. Its sectional curvature thus behaves as:

$$g \mapsto g' = \lambda g \rightarrow K' = \frac{1}{\lambda} K. \tag{33}$$

3.3 Hyperbolic space H^n

Hyperbolic space is space of constant negative curvature. Because Euclidean space has curvature zero, and a sphere has positive curvature, hyperbolic space completes the trio due to its negative curvature. One can imagine it as being bent inwards and outwards at the same time in different directions like a saddle. Analogous to the sphere, hyperbolic space can be embedded in $\mathbb{R}^{n+1}$ as $H^n = \{x \in \mathbb{R}^{n+1} \mid \|x\|_M = -R^2\}$. Note that the norm is taken with respect to the Minkowski scalar product: $\langle x, y \rangle_M = (\sum_{i=1}^{n-1} x^i \cdot y^i) - x^n \cdot y^n$. Just as the sphere, this hyperboloid can also be embedded in $\mathbb{R}^n$, e.g. as the Poincaré ball (in 2 dimensions also called Poincaré disk). One can again make use of the embedding to find the symmetries at an arbitrary point p:

$$s_p(y) = p\langle y, p \rangle_M - (y - p\langle y, p \rangle_M) = -y + 2p\langle y, p \rangle_M. \tag{34}$$

Figure 9: Two symmetries, namely s_p and s_q in hyperbolic space H^2. On the left in the embedding in $\mathbb{R}^3$ as hyperboloid, on the right as Poincaré disk projection in $\mathbb{R}^2$. The geodesics within the manifold need not be straight in the embedding. Note that the symmetry s_p at point p in eq. 34 corresponds to the embedding in $\mathbb{R}^3$ using the vector space structure of the embedding space. Hyperbolic space is infinte, i.e. the hyperboloid extends to infinty in z-direction. The Poincaré disk is an open disk, the further upwards the points are in the hyperboloid model, the closer they are to the boundary of the disk.

Distortion of the embedding

The hyperboloid embedding is a non-distorting embedding in the sense that the metric on the hyperboloid is induced by the Minkowski metric of the embedding manifold $\mathbb{R}^{n+1}$. It is just the restriction of the Minkowski scalar product to the tangent space of the hyperboloid. This is analogous to the case of $S^n \subset \mathbb{R}^{n+1}$.

For the Poincaré disk model, one introduces a metric tensor g which carries the information about the distortion between the true hyperbolic space and the Euclidean disk it lives in. This tensor is then dependent on the coordinates of the embedding. The metric tensor for the Poincaré disk can be computed as pullback tensor via the transformation and reads:

$$g(x) = \frac{4}{(1 - r^2)^2} \sum_i dx^i \otimes dx^i, \;\; \text{with} \;\; r = \sqrt{\sum_i (x^i)^2} \tag{35}$$

Note the similarity between this metric tensor and eq. 31. Since the disk is bounded, the metric has to capture that hyperbolic space is infinite as it diverges to infinity as $r \to 1$.

The sectional curvature

The sectional curvature K of a hyperboloid of radius R can be computed via eq. 28:

$$K_R(p) = -\frac{1}{R^2} \tag{36}$$

which is again independent of the evaluation point p. The sharper and more cone-like the hyperboloid, the more negatively it is curved.

Models of hyperbolic space
Aside from the hyperboloid model and the Poincaré disk model which have been introduced in section
3.3, there are even more models of hyperbolic space such as the upper half plane embedding or the
Beltrami-Klein model. The upper half-space model will be useful for the Siegel space in section 14.
All models of hyperbolic space will be denoted by H^n, it is however clear from context, which model is
used.
Recall the models of hyperbolic space which we have encountered so far:

- The hyperboloid model:
  $H^n = \{x \in \mathbb{R}^{n+1} | \|x\|_M = 1\}$ with metric tensor $g(x) = \sum_i^n dx^i \otimes dx^i - dx^{n+1} \otimes dx^{n+1}$

- The Poincaré ball model:
  $H^n = \{x \in \mathbb{R}^n | \|x\|_2 < 1\}$ with metric tensor $g(x) = \frac{4}{(1-\|x^n\|_2^2)^2} \sum_i^n dx^i \otimes dx^i$

The model as upper half-space is $H^n = \{x \in \mathbb{R}^n | x^n > 0\}$ with metric tensor $g(x) = \frac{1}{(x^n)^2} \sum_i^n dx^i \otimes dx^i$.
The figure below shows the two-dimensional Poincaré ball and the model as upper half-space with two
geodesic symmetries.

Figure 10: Two symmetries, namely s_p and s_q in hyperbolic space H^2. The blue and gold circles help
orientating. On the left in the embedding as Poincaré disk, on the right as upper half-space. Note that
hyperbolic space is infinite, i.e. both the Poincaré disk and the upper half-space are open sets in $\mathbb{R}^2$.

Complex coordinates for H^2
It is convenient to consider 2-dimensional hyperbolic space in complex coordinates. This of course
holds for the Poincaré disk and the upper half-space, since they are embedded in 2-dimensional space,
whereas the hyperboloid is embedded in 3-dimensional space and hence does not naturally fit to complex
coordinates. These models become in complex coordinates:

- The Poincaré disk model:
  $H^2 = \{z \in \mathbb{C} | |z| < 1\}$ with metric tensor $g(z) = \frac{4}{(1-|z|^2)^2} dz \otimes d\bar{z}$

- The upper half-space model:
  $H^2 = \{z \in \mathbb{C} | \Im(z) > 0\}$ with metric tensor $g(z) = \frac{1}{\Im(z)^2} dz \otimes d\bar{z}$

Note that the natural analogy $\{z \in \mathbb{C}^n | |z| < 1\} \not\cong H^m$ does not hold for any m if $n > 1$. Hyperbolic
space in higher dimensions is represented by the models embedded in the real numbers as above.
As both models are equivalent, there is a map called "Cayley transform" sending the points of the
half-space to the disk and vice versa:

$$i : \text{Disk} \to \text{half-space} \quad z \mapsto \frac{zi+i}{-z+1} \tag{37}$$

$$i^{-1} : \text{half-space} \to \text{Disk} \quad z \mapsto \frac{z-i}{z+i}. \tag{38}$$

Note that these maps rotate by 90° as one can see in fig. 10 above.

4 Transvections as composition of symmetries

The defining property of a symmetric space is to posses a so-called symmetry at every point p. This symmetry is an isometry, i.e. it keeps distances and angles invariant. A priori, a symmetric space only possesses these geodesic symmetries. One can now ask if a symmetric space also possesses other isometries aside from geodesic symmetries like translations and rotations as in the example space $\mathbb{R}^2$. In fact, a combination of two geodesic symmetries at different points is a translation. This can be motivated easily for the example space $\mathbb{R}^2$, but holds for any symmetric space:
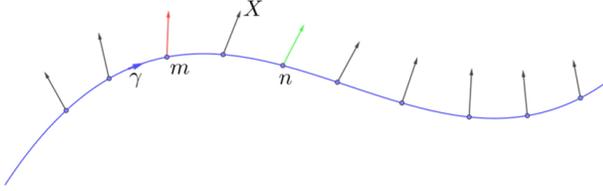
(a) Initial situation with geodesic γ and vector field X parallel to γ
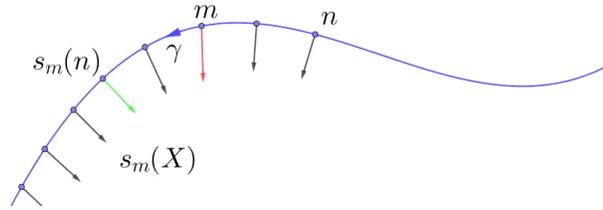
(b) Setup after the application of the first symmetry s_m
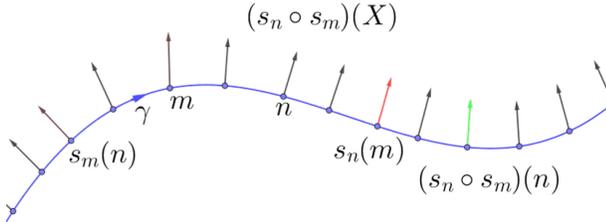
(c) Setup after the application of the transvection $s_n \circ s_m$

Figure 11: Illustration of a transvection $s_m \circ s_n$.

Consider a unit speed geodesic $\gamma(t)$, $t \in [-\infty, \infty]$ [a] and a vector field X parallel to γ [b] as depicted in fig. 11a. Apply a geodesic symmetry s_m WLOG at its midpoint $\gamma(0) = m$. By definition of the geodesic symmetry, the geodesic is mapped to itself. For X at m holds:

$$ds_m X(t)\Big|_{t=0} = -X(t)\Big|_{t=0} \tag{39}$$
$$\text{with } X(t) \coloneqq X(\gamma(t)).$$

Since s_m is an isometry, the reflected vector field is again parallel to the reflected geodesic, because angles are preserved. This means that eq. 39 holds for any t. The situation is depicted in fig. 11b.

If one applies another symmetry WLOG at point $n = \gamma(s)$ lying on the geodesic, γ and X are reflected again. X is hence pointing in the initial direction again. However, the combination of symmetries has moved points and vectors along the geodesic:

$$(s_n \circ s_m)(\gamma(t)) = \gamma(t+s),$$
$$d(s_n \circ s_m)_{\gamma(t)} X(t) = X(t+s). \tag{40}$$

This combination of symmetries is called a *transvection* and is the same as a parallel transport along the geodesic connecting m and n. The result can be seen in fig. 11c. A transvection is thus a generalization of the well-known translation in the space $\mathbb{R}^n$.

---

[a] Any geodesic can be extended to infinite length, as described in section 5

[b] This means that X is parallel transported along γ; the angle between the vector field X and the velocity vector $\dot{\gamma}$ is constant: $\nabla_{\dot{\gamma}} X = 0$.

The simple requirement to possess a geodesic symmetry yields another class of isometries, namely the transvections. Naturally, rotations around a point should also be isometries, as they leave angles and distances invariant. In fact, rotations can be generated via transvections in a little more complicated way via the Lie bracket of transvection Killing vector fields. This will be shown in detail in section 8.

Transvections, rotations, geodesic symmetries and their combinations are again isometries. They hence form a group, the isometry group.

Further detail on rotation groups

It will become important for the section about the Lie algebra of the isometry group that the group of isometries can have disjoint components. To motivate this, consider the example of the rotations on $\mathbb{R}^m$:

WLOG [10] choose the origin in $\mathbb{R}^m$. A point reflection with respect to the origin can be written as matrix $-Id_m$ with all entries on the diagonal equal to -1. This matrix has determinant $+1$ and -1 in even and odd dimensions and belongs to the group of orthogonal matrices $O(2n)$ and $O(2n+1)$, respectively. The group of orthogonal matrices is an unconnected manifold: Since the map $\det : O(m) \to \mathbb{R}$ is continuous, a closed image requires a closed pre-image. As $\{1, -1\}$ is the closed image, $O(m)$ splits into two closed disjoint components. These two components are $O^+(m) = SO(m)$ and $O^-(m)$ with positive and negative determinant, respectively. One can hence investigate if there is a path connecting the identity to the point reflection at the origin:

- In the case of $m = 2n$ even dimensions a point reflection can be obtained as n rotations about n axes. One can *continuously* rotate about these n axis to obtain the point reflection at the origin. Due to this continuity there is a path within the manifold $O(2n)$ connecting the identity Id_{2n} to $-Id_{2n}$. Hence both Id_{2n} and $-Id_{2m}$ live in the component $SO(2n)$.

Example: The rotation group SO(4)

Consider the case $m = 4$. Since there is a path within $SO(4)$ connecting Id_4 to $-Id_4$, there should be a vector $v \in T_e\big(SO(4)\big)$ such that the Riemann exponential map at $e = Id_4$ yields $-Id_4$. The vector v is itself a matrix. The tangent space $T_e\big(SO(4)\big)$ is denoted as $\mathfrak{so}(4)$[a]. This space can be represented as the set of all skew-symmetric (4×4)-matrices.[b]

The Riemann exponential map sends the infinitesimal transformation v to the point reflection:

$$T_e(SO(4)) = \mathfrak{so}(4) \ni v = \begin{pmatrix} 0 & \pi & 0 & 0 \\ -\pi & 0 & 0 & 0 \\ 0 & 0 & 0 & \pi \\ 0 & 0 & -\pi & 0 \end{pmatrix} \quad \text{with} \exp_e v = -Id. \tag{41}$$

In this case, the exponential map is the actual matrix exponential which explains the origin of the name. Note that a point reflection in $\mathbb{R}^{2n}$ is in fact the same as a rotation. In particular, the rotation in $\mathbb{R}^2$ is equivalent to a point reflection with respect to the origin:
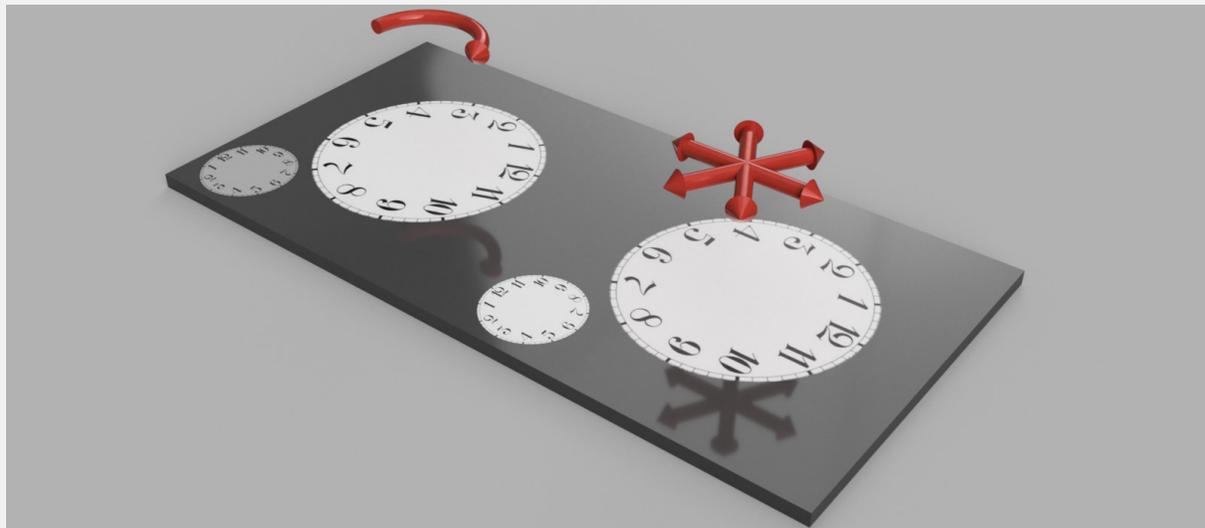
Figure 12: On the left, a rotation by almost π of the small clock is shown. On the right, the small clock is reflected with respect to the origin of the clock. One can see that a point reflection in $\mathbb{R}^{2n}$ can be obtained *continuously* by rotations.

---

[a]The Gothic letters are used for the tangent space at the identity if the manifold is a Lie group, see section 7
[b]Note that an element $v \in \mathfrak{so}(4)$ is not an element of $SO(4)$, since $\det v = 0$.

---

[10]A redefinition of coordinates can transform any point to the origin.

- Now consider the case of $2n + 1$ dimensions: There is no way to obtain *continuously* the point reflection. This is due to the fact that Id_{2n+1} has determinant $+1$, but the point reflection $-Id_{2n+1}$ has determinant -1. There is hence no path within $O(2n + 1)$ connecting these two matrices, as they lie in unconnected components of $O(2n + 1)$:

  In the language of manifolds it is impossible to find a vector $v \in T_e\big(O(2n + 1)\big)$ such that the emanating geodesic reaches $-Id_{2n1}$:

$$\nexists \ v \in SO(2n + 1) : \exp_e v = -Id. \tag{42}$$

Example: The rotation group $SO(3)$

The group $SO(3)$ is the well-known group of orientation-preserving rotations on $\mathbb{R}^3$. Apply a geodesic reflection with respect to the origin in $\mathbb{R}^3$. This point reflection can be represented by $-Id_3$. There is no orientation-preserving rotation (i.e. a rotation belonging to $SO(3)$) sending the basis vectors back to the original directions. There is hence no path within $O(3)$ connecting Id_3 with $-Id_3$ since they lie in different components.
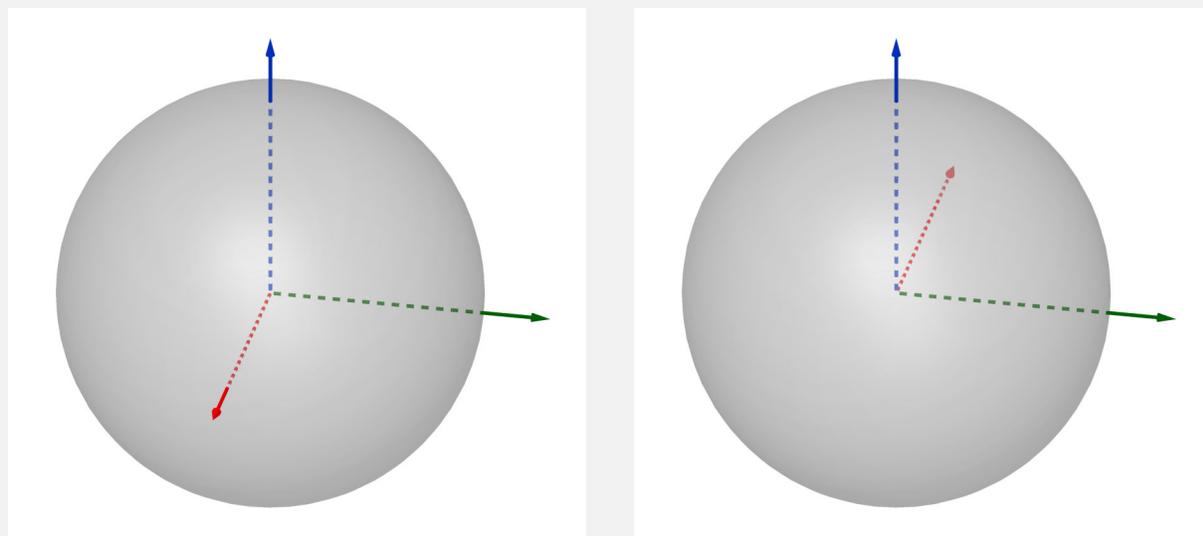
(a) The standard basis on $\mathbb{R}^3$. This is a right-handed system.

(b) After application of a point reflection with respect to the origin, one obtains a left-handed system.

Figure 13: In $\mathbb{R}^3$ one can see easily that a point reflection with respect to the origin cannot be obtained by a rotation from $SO(3)$, since the orientation has changed. This illustrates that $O(3)$ has two disconnected components, namely $O(3) = O^-(3) \cup SO(3)$. In particular, the point reflection matrix is contained in the component $O^-(3)$, the identity matrix is contained in $SO(3)$.

One denotes the component of the isometries on M which is connected to the identity by $I_0(M)$ or $Iso^+(M)$. For the case of $O(m)$ above, this component is written as $SO(m)$.

5 First results for symmetric spaces

The simple demand that the manifold M be a symmetric space leads to a number of interesting results (worked out in [Eschenburg, 1997]).

- A symmetric space is a complete metric space, since it is geodesically complete. Any geodesic can be extended to arbitrary length: WLOG let γ be a unit speed geodesic which is defined on $[0, s) \in \mathbb{R}$. The symmetry $s_{\gamma(s \cdot 3/4)}$ extends the geodesic to the interval $[0, s \cdot 3/2]$ but also reverses the direction of the geodesic. A second application of the above reflection reverses the direction to the geodesic's initial

direction. Using transvections, any geodesic can be extended to arbitrary length. This is motivated in fig. 11.

- A symmetric space is a homogeneous space. This means that one cannot distinguish between any two points of that manifold. An example of homogeneous space is the sphere S^2. We can always find an isometry which maps any arbitrary point to any other arbitrary point. If we regard the earth as S^2, there is no reason why e.g. the zeroth longitude needs to pass Greenwich. Any other choice for that longitude would describe the geometry of earth by no means worse.
  Mathematically speaking, the group of isometries on M called $\tilde{G} = I(M)$ acts transitively and smoothly on M: For any two points p and q one can find an isometry which maps p to q. An example of such an isometry is the transvection $s_q \circ s_{\gamma(r)}$, where $\gamma(r)$ is the midpoint of the geodesic connecting p and q.

  However, a homogeneous space is not equivalent to a space being isotropic. A space being isotropic means that at one special point p, one cannot find a distinct orientation. This becomes apparent when standing on top of a mountain. One may look around and see no differences whatsoever. This point is the only point of isotropy. Yet, this setting is not homogeneous, since every hiker can agree that there is a distinct point (namely the top of the mountain).
  Mathematically speaking, a space M is isotropic at p if for any two unit tangent vectors v, w at p there is an isometry φ with $\varphi(p) = p$ and $D\varphi(v) = w$. This is basically a rotation around p mapping the direction of v to the direction of w.
  The notions of homogeneity and isotropy are confused often, but their distinction is important for our purpose, as a symmetric space can be isotropic (such as S^2), but does not have to be isotropic (such as $S^2 \times \mathbb{R}^2$).
  According to the situation of isotropy at the top of the mountain, there is the reversed situation of homogeneity without istropy anywhere: A swimmer in the ocean feels the wind coming from one direction and can hence determine a distinct direction to which every other swimmer can agree. No swimmer is located at a special position, so this setting is homogeneous but not isotropic.

- A symmetric space has vanishing derivative of the curvature tensor ($\nabla R = 0$). One might be tempted to think that the curvature then has to be constant everywhere and the manifold then has to have either positive, negative or vanishing curvature everywhere, such as the sphere S^2, hyperbolic space H^2 or the flat Euclidean space $\mathbb{R}^n$. This is not the case, because the covariant derivative is a directional derivative. Since a symmetric space is homogeneous but need not be isotropic, there can be different subspaces of different curvature at the same time. When starting at point p and following one geodesic path γ_1, one will not measure a change of the curvature tensor, choosing a different geodesic path γ_2 however leads to a different, but also constant curvature tensor. This will become clearer with the examples in section 13 later, in particular $H^2 \times \mathbb{R}$ and $S^2 \times \mathbb{R}$.

6 The algebraic approach to symmetric spaces

A symmetric space is closely related to its isometry group. This can be motivated:
Consider $M = S^2 \subset \mathbb{R}^3$ as manifold and its isometry group $G = SO(3)$ of all distance-preserving rotations in $\mathbb{R}^3$. Since all $SO(3)$-rotations map the sphere to itself, it is promising to study the relation between a manifold M and its isometry group $G := I_0(M)$[11]. Because the isometry group $G = SO(3)$ is three-dimensional, but $M = S^2$ is only two-dimensional, the correspondence between M and G is a little more complex:

---

[11]Strictly speaking, the full isometry group is the group $O(3)$. The positive component $G := I_0(M)$ is the set of orientation-preserving isometries. In order to being able to reach every element of G via a path starting at the identity which is contained in G, one chooses only the component $I_0(M)$. The geometry of the underlying manifold is not changed by any means.

Definition: Group action on a manifold

Let G be a group and M be a manifold. G is said to *act* on M via the map θ if every $g \in G$ induces a map $M \to M$. That is $\theta : G \to \mathcal{M}aps(M; M)$. The map θ thus sends every element of G to a transformation on M in such a way that for $g_i \in G$ hold:

$$\theta(g_1 \cdot g_2) = \theta(g_1) \circ \theta(g_2). \tag{43}$$

The symbol '$\cdot$' stands for the group operation within G, the symbol '$\circ$' for the composition of two maps. G acting on M is written as $G \curvearrowright M$. Note that $\theta(g)(p),\ p \in M$ is often written as gp and the action θ is suppressed.

Example:

The group $SO(3)$ acts on S^2 as the usual matrix vector product:

$$\theta\left(\begin{pmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{pmatrix} \right) \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} = \begin{pmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{pmatrix} \cdot \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}. \tag{44}$$

In fact: $\theta(g) \in \mathcal{M}aps(S^2; S^2),\ g \in SO(3)$. **Example:**

The group $PSL(2, \mathbb{R}) = SL(2, \mathbb{R})/\{\pm Id\}$[a] acts on the upper half plane $H^2 = \{z \in \mathbb{C}|\ \Im(z) > 0\}$ [b] as Möbius transformations:

$$\theta\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} \right) \begin{pmatrix} \Re(z) \\ \Im(z) \end{pmatrix} = \frac{a(\Re(z) + \Im(z) \cdot i) + b}{c(\Re(z) + \Im(z) \cdot i) + d} \text{ as complex division.} \tag{45}$$

Explicitly, interpreting $g \in SL(2, \mathbb{R})$ as matrix vector product gives a different result, e.g.:

$$\begin{pmatrix} \cos(\frac{\pi}{2}) & -\sin(\frac{\pi}{2}) \\ \sin(\frac{\pi}{2}) & \cos(\frac{\pi}{2}) \end{pmatrix} \curvearrowright (0,1)^T = \frac{\cos(\frac{\pi}{2}) \cdot i - \sin(\frac{\pi}{2})}{i \cdot \sin(\frac{\pi}{2}) + \cos(\frac{\pi}{2})} = i \text{ interpreted as Möbius transformation} \tag{46}$$

$$\text{but: } \begin{pmatrix} \cos(\frac{\pi}{2}) & -\sin(\frac{\pi}{2}) \\ \sin(\frac{\pi}{2}) & \cos(\frac{\pi}{2}) \end{pmatrix} \cdot (0,1)^T = (-1,0)^T \to -1 + 0 \cdot i, \text{ interpreted as matrix vector product.} \tag{47}$$

---

[a]Note that $\theta(g) = \theta(-g)$ for $g \in SL(2, \mathbb{R})$. One hence only considers the corresponding equivalence class of M.
[b]More details about the action on the upper half-space model can be found in the box in section 14.

Remember that $G := I_0(M)$ is the path-connected component connected to the identity of the isometry group. Since a symmetric space is homogeneous, we can choose any point p as base point. Consider the isotropy group $G_p := \{g \in G | \theta(g)p = p\}$ which is a closed subgroup of G. We can now identify M with the coset space G/G_p because every point of the manifold can be obtained by letting a distinct $g \in G/G_p$ act on p.

G_p is often written as K, since the choice of the point p is suppressed because it is arbitrary.

The projection map π maps $g \in G$ to its equivalence class $[g] = gK$ where K stands for any arbitrary element of K. Any two elements g_1 and g_2 are considered equal in G/K if there is a $k \in K$ with $g_1 \cdot k = g_2$. It follows:

$$g_1(g_2 K) = (g_1 g_2)K \tag{48}$$

which the following commuting diagram illustrates:

$$
\begin{array}{ccc}
G & \xrightarrow{\ \pi\ } & G/K \\
\downarrow{\scriptstyle L_G} & & \downarrow{\scriptstyle L_G} \\
G & \xrightarrow{\ \pi\ } & G/K
\end{array}
\text{, with } L_G \text{ as left group multiplication in } G.
$$

M then can be identified with G/K, i.e. $M \cong G/K$. The diffeomorphism

$$
\begin{aligned}
G/K &\to M, \\
gK &\mapsto \theta(g)p
\end{aligned}
\tag{49}
$$

makes this correspondence explicit. It follows:

$$
\dim(M) = \dim(G) - \dim(K)
\tag{50}
$$

The concept of a symmetric space as coset space will become clear with the examples:

Example: $\mathbb{R}^2$ as coset space
The isometry group for Euclidean space $\mathbb{R}^2$ is the so-called Euclidean group $E(2) =: G$ which consists of both translations and rotations. WLOG one can choose the origin as base point p. Obviously, the isotropy group K is the group of all rotations around the origin - namely $O(2)$ - because only the origin is fixed under the action of K.
One can obtain any other point q in $\mathbb{R}^2$ from p since there exists a $g \in G$ such that $\theta(g)p = q$. However, identifying the point q with the transformation g is not unique, because there are several transformations $\{gk_1, gk_2, ...\}, k_i \in K$ which map the origin p to the point q: If $\theta(g)$ maps the origin p to q, so does $\theta(gk)$, since $\theta(gk) = \theta(g) \circ \theta(k)$ and $\theta(k)$ does not move the origin. One therefore introduces the projection π:

$$
\begin{aligned}
\pi : G &\to G/K \\
g &\mapsto gK.
\end{aligned}
\tag{51}
$$

The Euclidean plane can thus be identified with $E(2)/O(2)$ which corresponds to G/K.

Example: S^2 as coset space
The isometry group G for the sphere S^2 is $SO(3)$ - as mentioned before. The isotropy group K is $SO(2)$ which is the group of those rotations which leave one axis invariant. If we choose $p = (0,0,1)^T$, the isotropy group consists of rotations around the z-axis which can be described as $SO(2)$. Every point $q \in S^2$ can be obtained via a transformation from G. Again, if $\theta(g)(p) = q$, then $\theta(gk)(p) = q$ for $k \in K$. To make the correspondence unique, we use the projection π which maps every rotation g to the equivalence class gK which is the set of transformations which at first apply an arbitrary element of K and after that apply g.

Example: H^2 as coset space
The isometry group G of hyperbolic space H^2 in the hyperboloid model is $SO(2,1,\mathbb{R})$. These isometries correspond to translations and rotations in hyperbolic space leaving the hyperboloid shown in fig. 9 invariant. We can again choose an arbitrary but fixed point $p \in H^2$ and investigate its isotropy group. We may choose the bottom point $p = (0,0,1)^T$ for convenience. The subgroup $K \subset G$ which leaves p invariant is $SO(2)$: These are the rotations around the z-axis. Now one can consider the correspondence between a mapping $g \in G$ and a point q of manifold by $\theta(g)p = q$. Applying any rotation $k \in K$ before applying $\theta(g)$ on p yields the same point $\theta(gk)p = q$. To make the correspondence unique, we apply the projection π and obtain a one-to-one correspondence between M and G/K as in eq. 49.

7 Lie algebra of the isometry group as set of Killing vector fields

Due to a theorem by Myers and Steenrod [Myers and Steenrod, 1939] the isometry group G of a complete Riemannian manifold is a Lie group.[12] Since the isotropy group K is a closed subgroup, it is also a Lie group, proving that the space G/K is actually a homogeneous space. The Lie group G is in general a curved manifold. Instead of working with this curved object, many results also follow by just considering its Lie algebra which is just a flat vector space and hence often much easier to deal with.

Recap: Lie algebra

The Lie algebra $\mathfrak{g}$ of the Lie group G is a distinct vector space of the same dimension as G. Any Lie group is equipped with this vector space $\mathfrak{g}$ of dimension $\dim \mathfrak{g} = \dim G$. This vector space can be regarded as the tangent space at the identity element e, i.e. $\mathfrak{g} \cong T_e G \cong \mathbb{R}^{(\dim G)}$. The tangent space is a Euclidean vector space and hence easy to deal with. Although the curved Lie group G is in general difficult to study, its Lie algebra $\mathfrak{g}$ carries lots of information about G itself. This correspondence can be made explicit with the Riemann exponential map which was introduced in section 2.7. This exponential map is a map into the Lie group G (which is a manifold):

$$
\begin{aligned}
\exp : T_e G = \mathfrak{g} &\to G \\
v &\mapsto \exp_e(v) = \gamma_v(1) \\
\text{with } \gamma_v(0) = e \text{ and } \dot{\gamma}_v(0) &= v \in T_e G.
\end{aligned}
\tag{52}
$$

This map can be understood as a differential equation in G since both initial value $\gamma_v(0)$ and its derivative at 0 are given.

Further detail

There is another definition of the Lie algebra in terms of left-invariant[13] vector fields. For this definition one needs the notion of a push-forward on a manifold M:

Let $p, q \in M$ and $\phi : M \to M$. One can send a vector v at p to a vector $\phi_*(v)$ at q using the push-forward of the map ϕ with $\phi(p) = q$. The map ϕ_* is a linear map, sending a vector from the tangent space at p to a vector from the tangent space at q. This map ϕ_* can be understood as Jacobian matrix of ϕ.

Let us now consider the push-forward on a Lie group G, i.e. a manifold which also carries group structure. Introduce the left-multiplication:

$$
\begin{aligned}
L_g : G &\to G \\
h &\mapsto g \cdot h = L_g(h) \\
\Rightarrow L_g(\cdot) &\in \mathcal{M}aps(G; G).
\end{aligned}
\tag{53}
$$

Its push-forward is:

$$
\begin{aligned}
(L_g)_* : TG &\to TG \\
v &\mapsto (L_g)_* v \in T_{gh} G \text{ with } v \in T_h G.
\end{aligned}
\tag{54}
$$

The so-called left-invariant vector fields are those vector fields, which fulfill

$$
\left((L_h)_* X \right)(g) = X(hg).
\tag{55}
$$

These vector fields are special. In fact, on a manifold which is not a Lie group, there is no natural way to define a distinct class of vector fields.

---

[12]As a quick reminder, a Lie group is a smooth manifold with an additional group structure, i.e. points of the manifold can be multiplied using the group operation. An example is the Lie group $SO(2, \mathbb{R})$:

$$
SO(2, \mathbb{R}) = \begin{pmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{pmatrix}, \phi \in \mathbb{R}.
$$

It is a smooth 1-dimensional manifold (because it has only one free parameter ϕ) embedded in $\mathbb{R}^4$. In fact, it is equivalent to the circle $S^1 \subset \mathbb{R}^2$. Obviously, $SO(2, \mathbb{R})$ is also a group with matrix multiplication as group operation.

[13]There is also the equivalent method to choose the right-invariant vector fields, but it has become convention to consider the left-invariant vector fields.

The equivalence of the two definitions of the Lie algebra can be seen by noting that a vector at e induces a left-invariant vector field on all of G:

$$X(h) = (L_g)_* v \ \text{ with } \ g \cdot e = h. \tag{56}$$

Vice versa, every left-invariant vector field can be associated with its value at e which is a vector in $T_e G \cong \mathfrak{g}$. The isomorphism between the tangent space at e and the space of left-invariant vector fields is denoted by ξ and ξ^{-1}.

The Lie bracket on $\mathfrak{g}$

The Lie algebra possesses a Lie bracket, i.e. a bilinear, anti-symmetric product on $\mathfrak{g}$ fulfilling the Jacobi-identity. This Lie bracket is inherited from the Lie bracket of vector fields on the manifold G as:

$$[v_1, v_2] := \xi^{-1}\big([\xi(v_1), \xi(v_2)]\big), \ \ v_i \in T_e G, \mathcal{X}(G) \ni \xi(v_i) \text{ left-invariant.} \tag{57}$$

More on the Lie bracket of vector fields in section 8.

Recap: The flow of a vector field

A vector field is a smooth section of the tangent bundle which means that to every point p an element of the tangent space T_pM is assigned. This coincides with the notion of a vector field assigning a vector to every point of M.

This vector field induces a set of transformations on M as follows: A pair $(p, v) \in TM$ determines a differential equation with unique solution via the Riemann exponential map as in eq. 25. The solution to any pair (p, X_p) is a geodesic with starting point p and initial velocity vector $X_p := X(p)$. For a different pair (q, X_q) the geodesic will of course be different. This set of geodesics induced by X is called the flow $\{\phi_t\}$ of X. The free parameter t determines how far the starting point p will be transported along the geodesics $\gamma_{(p,X_p)}$ with initial values (p, X_p). The flow has following properties:

$$\phi_t(p) = \exp_p(tX) \tag{58}$$

$$\phi_0 = Id(M), \text{ the identity} \tag{59}$$

$$\phi_{t+s} = \phi_t \circ \phi_s. \tag{60}$$

The flow is often referred to as one-parameter-group, since it is a group in t:

$$\begin{aligned} \phi_t : \mathbb{R} &\to \mathcal{M}aps(M; M) \\ t &\mapsto \exp_p(tX_p) \ \forall p \in M. \end{aligned} \tag{61}$$
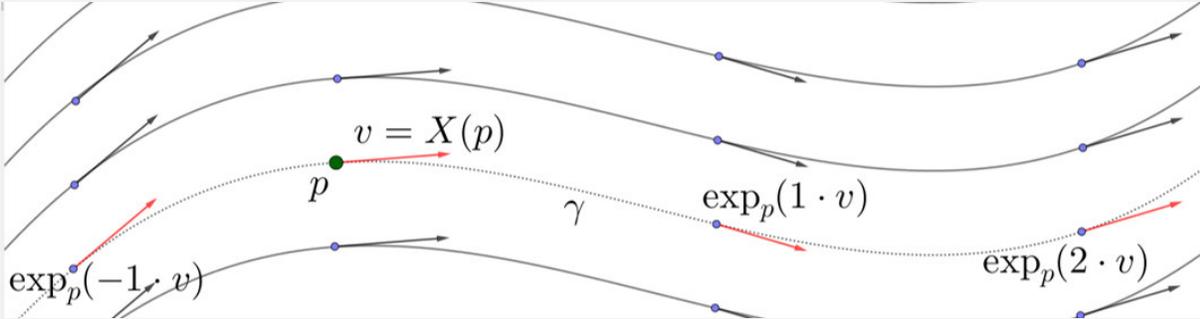
Figure 14: A vector field X generates the flow $\{\phi_t\}$: X assigns to every point p a velocity vector X_p which leads to a transformation along the geodesic $\gamma_{(p,X_p)}$. Note the group property in the variable t.

Killing vector fields

Because a vector field corresponds to a one-parameter-group of transformations on M, one can interpret a vector field as set of maps from M to M. One hence calls those vector fields whose induced transformations are isometries on M "Killing vector fields". In a coordinate frame, X is a Killing vector field, if:

$$(\nabla_\mu X)_\nu = -(\nabla_\nu X)_\mu. \tag{62}$$

Example

A trivial example is the zero vector field $X_0 = 0$ which assigns the zero vector to every point. The induced transformation is the identity on all of M which is trivially an isometry.

Example

The Killing vector field $X_{\text{trans}} = \partial_x$ on $M = \mathbb{R}^2$ obviously fulfills eq. 62. It corresponds to a shift in x-direction; its flow consists of transvections.

Example

The Killing vector field $X_{\text{rot}} = y\partial_x - x\partial_y$ on $M = \mathbb{R}^2$ also fulfills the Killing condition. It corresponds to a rotation around the origin and hence not to a transvection.

Since every Killing vector field generates a group of isometries on M, there is a clear correspondence between the space of Killing vector fields and the isometry group $G := I_0(M)$: Together with a fixed t_0, ϕ_{t_0} is an isometry. Due to the close relation between a Lie group and its Lie algebra, it is natural to investigate the connection between the space of Killing vector fields on M and the Lie algebra $\mathfrak{g}$ of the Lie group G:

As a Killing vector field X induces an isometric flow $\{\phi_t\}$ on M, one can associate X with the element $v \in \mathfrak{g}$ such that holds:

$$G \ni \exp_e(tv) = \phi_t \in G \ \ \forall t. \tag{63}$$

This map will be called Υ. Having found the one-to-one correspondence between the space of Killing vector fields and the Lie algebra $\mathfrak{g}$ of the Lie group G of isometries on M, one can consider them as identical and compute properties of Killing vector fields and hence of the manifold M just by investigating $\mathfrak{g}$:

$$\mathfrak{g} \times \mathbb{R} \xrightarrow{\exp_e(tv)} G \xrightarrow{\ \theta\ } M$$

X Killing vector field on M.

Further detail

In general, not all isometries of the full isometry group can be reached via a geodesic starting at $e \in G$. That is because the full isometry group is in general not connected[14]. Because of that the group $G := I_0(M)$ does in general not contain geodesic symmetries of $M \cong G/K$. This is remarkable, since the existence of a reflection at every point is the defining property of a symmetric space.

An illustration of this gives the 3-dimensional sphere $S^3 \cong SO(4,\mathbb{R})/SO(3,\mathbb{R})$ with north pole $p = (0,0,0,1)^T$. The geodesic symmetry is

$$s_{(0,0,0,1)^T}(x^1, x^2, x^3, x^4)^T = (-x^1, -x^2, -x^3, x^4)^T, \tag{64}$$

but for the corresponding matrix holds: $\mathrm{diag}(-1,-1,-1,1) \notin SO(4,\mathbb{R})$, since it is not orientation preserving. One can understand the coset form G/K of a symmetric space M as a characterization in terms of transvections starting at one base point p. Although a symmetric space necessarily possesses a geodesic symmetry at every point, G/K does not give any information how the geodesic symmetry on M looks like. In a pictoral sense, G/K is the space of all orientation preserving isometries without rotations. These then boil down to transvections.

8 Cartan Decomposition

When investigating the space of Killing vector fields on M further, one notices that they and hence $\mathfrak{g}$ can be classified:[15]

$$\mathfrak{k} = \{x \in \mathfrak{g} \mid \Upsilon(x)_p = 0\} \ \text{and} \ \mathfrak{p} = \{x \in \mathfrak{g} \mid (\nabla\Upsilon(x))_p = 0\} \ \text{with} \ \mathfrak{k} \cap \mathfrak{p} = 0 \in \mathfrak{g}. \tag{65}$$

This categorization of subspaces of $\mathfrak{g}$ is called *Cartan Decomposition*. Examining the vector space $\mathfrak{k}$, one notices that the corresponding flow needs to be the identity at p (for all t), as the velocity vector X_p of a geodesic through p is zero. The space $\mathfrak{k}$ reminds us of example Killing vector field X_{rot} in the box above: Let us choose $p = (0,0)^T$ in its example setting $\mathbb{R}^2$. Since X does not alter p, it belongs to $\mathfrak{k}$.

Examining the vector space $\mathfrak{p}$, one notices that the corresponding flow must be uniform in magnitude but pointing in a direction away from p because X_p is non-zero. The space $\mathfrak{p}$ however reminds us of example X_{trans} above. One may choose $p = (0,0)^T$ in its example setting $\mathbb{R}^2$. Since the derivative of X vanishes at p, but X does not vanish, it is associated with a constant shift in x-direction.

The example Killing vector field X_0 above is part of both subspaces of $\mathfrak{g}$. In fact, $\mathfrak{g}$ decomposes as direct sum $\mathfrak{g} = \mathfrak{p} \oplus \mathfrak{k}$. The zero Killing vector field thus corresponds to the zero element in $\mathfrak{g}$.

---

[14]Recall the discussion in section 4.
[15]Once again, the base point p is arbitrarily chosen.

To examine the Lie bracket on $\mathfrak{g}$, recall from above that $\mathfrak{g}$ carries its Lie bracket inherited from the Lie bracket of vector fields on its Lie group G: The following box deals with the Lie bracket on vector fields.

Recap: Lie bracket as non-commutative flows

The Lie bracket $[X, Y]$ of two vector fields has a geometric meaning as composition of the flows σ and τ which are generated by the vector fields X and Y, respectively. Consider a point x and move it a little along the flow σ signified with a red arrow in the figure below generated by the vector field X. This point is then moved a little along the flow τ generated by Y and vice versa. The parallelogram is not closed, since the flows are not constant on the manifold. In particular, the difference between the two paths is proportional to the Lie bracket $[X, Y]$.[a] Note that the green arrow is pointing in a maybe unintuitive direction.

Algebraically, the Lie bracket has the form:

$$[\cdot, \cdot] : \mathcal{X}(M) \times \mathcal{X}(M) \to \mathcal{X}(M)$$
$$X, Y \mapsto [X, Y] := (X^\alpha \partial_\alpha Y^\beta - Y^\alpha \partial_\alpha X^\beta)\partial_\beta \overset{\star}{=} \nabla_X Y - \nabla_Y X \tag{66}$$

where the equality "$\star$" is due to our geometry setup being torsion-free. An explicit example of the Lie bracket is given for the sphere S^2 in fig. 16.
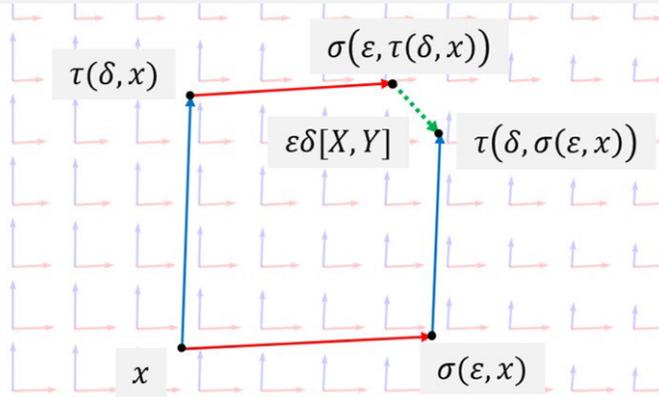
Figure 15: Lie bracket as non-commutative flows. The arrows in the background symbolize the non-constant vector fields X (red) and Y (blue). The parallelogram does not close due to the fact that X and Y are not constant. Note that the green arrow would be pointing in the opposite direction for the Lie bracket $[Y, X]$.

---

[a]Note that the difference between two points is in general not defined on a manifold. Since the parameters ϵ and δ are taken to be very small, the whole area falls into one tiny neighbourhood of x which can be equipped with Euclidean coordinates.

The space $\mathfrak{g}$ and the space of Killing vector fields on M are related via the map Υ. However, it is not obvious in which way Υ relates the Lie bracket on $\mathfrak{g}$ to the Lie bracket on the Killing vector fields. It can be shown [Sakai, 1996, chaper I, §2, E.8] that the map Υ is an anti-isomorphism in the sense that holds:

$$\Upsilon[x, y] = -[\Upsilon x, \Upsilon y], \quad x, y \in \mathfrak{g}. \tag{67}$$

The minus sign has to be considered, when transferring results from investigating $\mathfrak{g}$ to the space of Killing vector fields and vice versa.

It is now interesting to consider elements of $\mathfrak{k}$ and $\mathfrak{p}$ as inputs for the Lie bracket on $\mathfrak{g}$. It is shown below that:

$$[\mathfrak{k}, \mathfrak{k}] \subseteq \mathfrak{k}$$
$$[\mathfrak{p}, \mathfrak{p}] \subseteq \mathfrak{k} \tag{68}$$
$$[\mathfrak{p}, \mathfrak{k}] \subseteq \mathfrak{p}$$

which is to be understood element-wise, i.e. they hold for all elements from $\mathfrak{p}$ and $\mathfrak{k}$, respectively. The first two statements are straight-forward to show:

Let v, w be in the subspace $\mathfrak{k} \subset \mathfrak{g}$ with corresponding Killing vector fields V, W, i.e. $V_p = W_p = 0$. Then

$$[V,W]_p = (\nabla_V W)_p - (\nabla_W V)_p = 0 \tag{69}$$

since the derivative along the zero vectors V_p and W_p is trivially zero. From eq. 67 follows that $[\mathfrak{k}, \mathfrak{k}] \subseteq \mathfrak{k}$.

Let V, W now be Killing vector fields from $\mathfrak{p}$, i.e. $(\nabla V)_p = (\nabla W)_p = 0$. Then eq. 69 also holds by assumption, i.e. $[V,W]_p = 0$. Using the anti-isomorphism eq. 67 one arrives at $[\mathfrak{p}, \mathfrak{p}] \subseteq \mathfrak{k}$.

The third statement can be proven with aid of the following property of a Killing field [Eschenburg, 1997, 3. Lemma 1]:

$$\nabla_A \nabla_V W = \nabla_{\nabla_A W} V - R(A, V)W. \tag{70}$$

Let now $\Upsilon^{-1}(V) \in \mathfrak{k}$, $\Upsilon^{-1}(W) \in \mathfrak{k}$. It needs to be shown that $(\nabla_A[V,W])_p = 0$ for all vector fields $A \in \mathcal{X}(M)$, dropping the subscript p for convenience:

$$\begin{aligned}
\nabla_A[V,W] &= \nabla_A \nabla_V W - \nabla_A \nabla_W V \\
&= -R(V,A)W + \nabla_{\nabla_A W} V + R(W,A)V - \nabla_{\nabla_A V} W \\
&= R(A,V)W + R(W,A)V \\
&= R(V,W)A \\
&= 0
\end{aligned} \tag{71}$$

where the second equality is due to the identity eq. 70 above. The third equality follows because $(\nabla W)_p = 0$, the forth follows from the Bianchi identity. Since the equation needs to hold for any A, in particular for the case $A_p = 0$, it needs to be zero. With eq. 67 one obtains:

$$(\nabla_A[V,W])_p = 0 \Rightarrow [\Upsilon(v), \upsilon(w)] \subseteq \Upsilon(\mathfrak{p}) \Rightarrow -\Upsilon[v,w] \subseteq \Upsilon(\mathfrak{p}) \Rightarrow [\mathfrak{p}, \mathfrak{k}] \subseteq \mathfrak{p}. \tag{72}$$

There is a handy way to calculate the Riemann curvature tensor at the base point p by using generators of transvections:

$$R_p(V,W)U = [U, [V,W]]_p, \quad U, V, W \in \Upsilon(\mathfrak{p}). \tag{73}$$

By comparing that to the formula to compute the sectional curvature as in eq. 28 one notices that the sectional curvature of a plane spanned by two normed vectors can be easily calculated this way. The set $\mathfrak{p}$ together with the triple product $[\cdot, [\cdot, \cdot]]$ forms a so-called Lie triple system, because it leaves $\mathfrak{p}$ invariant, i.e. $[[\mathfrak{p}, \mathfrak{p}]\mathfrak{p}] \subseteq \mathfrak{p}$. This can be seen easily with eq. 68.

The equation eq. 73 above can be computed directly, for simplicity drop the base point p:

$$\begin{aligned}
[U, [V,W]] &= \nabla_U[V,W] - \nabla_{[V,W]}U \\
&= \nabla_U[V,W] \\
&= \nabla_U \nabla_V W - \nabla_U \nabla_W V \\
&\stackrel{\text{eq. 70}}{=} \nabla_{\nabla_U V} W - R(W,U)V - \nabla_{\nabla_U V} W + R(V,U)W \\
&= -R(W,U)V + R(V,U)W \\
&= R(V,W)U
\end{aligned} \tag{74}$$

where $[V,W] \in \Upsilon(\mathfrak{k})$, hence $(\nabla_{[V,W]}A)_p = 0 \ \forall A \in \mathcal{X}(M)$ and the last equality is due to the Bianchi identity.

$\square$

Recall that the Lie algebra of the isometry group G is in $1-1$ correspondence to the space of Killing vector fields on $M \cong G/K$. This has been pointed out extensively in section 7. To get an intuition for eq. 68, it makes sense to consider the Lie bracket on $\mathfrak{g}$ in the picture of Killing vector fields on M as in eq. 67:

Cartan decomposition on $\mathfrak{so}(3)$

Let us take a look at the symmetric space $S^2 \cong SO(3)/SO(2)$. The Lie algebra of the isometry group $SO(3)$ is $\mathfrak{so}(3)$ and can be decomposed as $\mathfrak{g} = \mathfrak{p} \oplus \mathfrak{k}$. Note that $\mathfrak{so}(3)$ can be represented of skew-symmetric 3×3 matrices. An element of $\mathfrak{g}$ is hence no rotation matrix, since its determinant is zero.

With base point $p = (0,0,1)^T$, the space $\mathfrak{k}$ corresponds to all rotations around the z-axis. The space $\mathfrak{p}$ corresponds to all rotations around axis different from the z-axis which shift p, since p does not lie on the rotation axis and is hence not invariant. Note that this fits to the defining property of $\mathfrak{p}$ in eq. 65.

A basis of $\mathfrak{so}(3)$ are the well-known angular momentum operators L_i:

$$L_x = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad L_y = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad L_z = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{75}$$

Because the base point is $p = (0,0,1)^T$, L_z obviously generates $\mathfrak{k}$, whereas L_x and L_y generate $\mathfrak{p}$.

The commutation relations thus nicely illustrate the Lie bracket operation on a general Cartan decomposition:

$$\begin{aligned} [L_z, L_z] &\subseteq L_z & [\mathfrak{k}, \mathfrak{k}] &\subseteq \mathfrak{k} \\ [L_x, L_y] &= L_z & [\mathfrak{p}, \mathfrak{p}] &\subseteq \mathfrak{k} \\ [L_y, L_z] &= L_x & [\mathfrak{p}, \mathfrak{k}] &\subseteq \mathfrak{p}. \end{aligned} \tag{76}$$

To get a feel for this formula, we use the Lie bracket as measure of commutativity of flows. The figure below illustrates $[\mathfrak{p}, \mathfrak{p}] \subseteq \mathfrak{k}$ on the sphere using the generated vector fields of $\mathfrak{p}$ and $\mathfrak{k}$ via Υ.
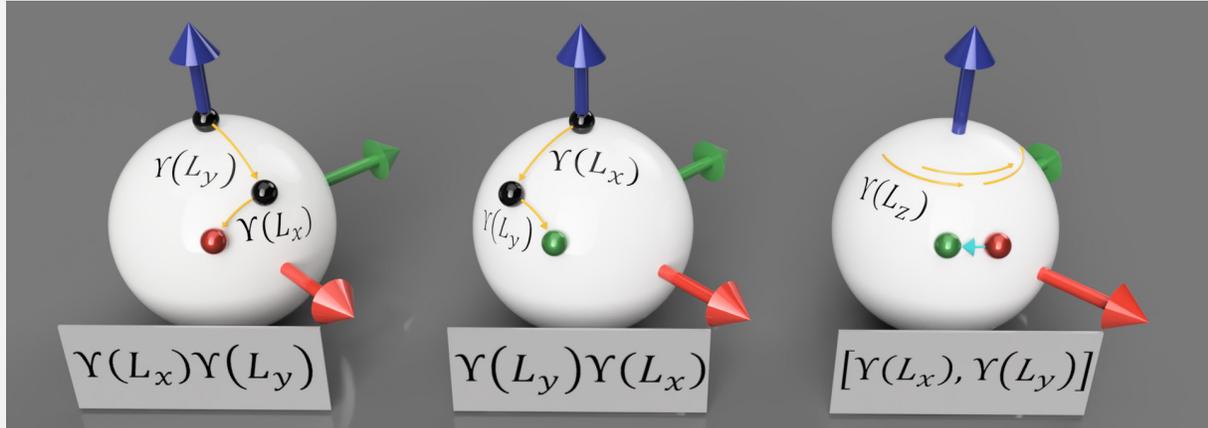
Figure 16: Visualization of $[\Upsilon(L_x), \Upsilon(L_y)] = -\Upsilon(L_z)$ on the sphere S^2 in three steps. The flows of the vector fields are indicated by thin yellow arrows.

On the left one can see the result of firstly transporting along the flow generated by $\Upsilon(L_y)$ and secondly transporting along the flow generated by $\Upsilon(L_x)$.

In the middle, one can see the result of the transports in opposite order.

On the right, the difference between the two results is indicated by a turquoise arrow going from right to left. It corresponds to $[\Upsilon(L_x), \Upsilon(L_y)]$. However, the flow generated by $\Upsilon(L_z)$ points from left to right. This reflects the anti-isomorphism of $\mathfrak{g}$ and the Lie algebra of Killing vector fields as in eq. 67.

Remember that the dimension of a Lie algebra is the same as of the Lie group, simply because the tangent space is a vector space of the same dimension as the manifold. For S^2 one can check that

$$2 = \dim S^2 = \dim\left(SO(3)/SO(2)\right) = \dim SO(3) - \dim SO(2) = 3 - 1 = \dim \mathfrak{g} - \dim \mathfrak{k} = \dim \mathfrak{p}. \tag{77}$$

9 Cartan Involution

A different way to obtain the Cartan decomposition on $\mathfrak{g}$ is by examining the eigenspaces of the derivative of a special map, which is called *Cartan involution*.

Define the Cartan involution on G as:

$$\begin{aligned} \sigma_p : \ & G \to G \\ & g \mapsto g_{s_p} \cdot g \cdot g_{s_p}, \ \text{with } \theta(g_{s_p}) = s_p, \ g_{s_p} \in I(M). \end{aligned} \tag{78}$$

It has been pointed out extensively that in general s_p need not be an element of $G = I_0(M)$ but an element of $I(M)$. The combination of two symmetries however is again in G:

The map $\phi : \ (M, M) \to I(M), \ (m, n) \mapsto s_m \circ s_n$ is continuous in M. [16] A smooth variation in (M, M) can hence not reach a different unconnected component of $I(M)$ via ϕ. The element (q, q) however is mapped to the identity for all $q \in M$, since s_q is self-inverse. This means that $\mathrm{im}(\phi)$ never leaves the identity component of $I(M)$, namely $G = I_0(M)$.

An involution is a map which is of order two, i.e. twofold application yields the identity:

$$(\sigma)^2(g) = g_{s_p} \cdot g_{s_p} \cdot g \cdot g_{s_p} \cdot g_{s_p} = Id(g), \tag{79}$$

since s_p is self-inverse. The derivative of σ maps elements of the tangent space of G to itself:

$$\sigma_* : \ T_h G \to T_{\sigma(h)} G, \ \text{in particular:} \tag{80}$$
$$\sigma_*(v) : \ \mathfrak{g} \cong T_e G \to \mathfrak{g}, \ v \in T_e G. \tag{81}$$

Because $(\sigma)^2 = Id$, σ_* can only have eigenvalues $\{-1, 1\}$. One defines the spaces

$$\mathfrak{k} = \{x \in \mathfrak{g} \mid \sigma_*(x) = -x\} \ \text{and} \ \mathfrak{p} = \{x \in \mathfrak{g} \mid \sigma_*(x) = x\} \ \text{with} \ \mathfrak{k} \cap \mathfrak{p} = 0 \in \mathfrak{g}. \tag{82}$$

Note that although the spaces are called the same way as in the transvection and rotation Killing vector fields in eq. 65, their equivalence needs to be proven. If the different definitions on $\mathfrak{k}$ coincide, it follows that they also coincide on $\mathfrak{p}$ just by reasons of dimension of $\mathfrak{g}$.

The illustration of fig. 17 helps to understand the nature of the Cartan involution and hence the following proof. Consider the commuting [17] diagram:

$$\begin{array}{ccc} \mathfrak{g} & \xrightarrow{\ \exp\ } & G \\ \ \downarrow{\scriptstyle \sigma_*} & & \ \downarrow{\scriptstyle \sigma} \ , \ \to \exp_e(t\sigma_* x) = \sigma(\exp_e(tx)) = \ g_{s_p} \cdot \exp_e(tx) \cdot g_{s_p} \\ \mathfrak{g} & \xrightarrow{\ \exp\ } & G \end{array} \tag{83}$$

Let $\sigma_*(x) = x$, it needs to be shown that $\Upsilon(x)_p = 0$. Using eq. 83 one obtains:

$$\theta\big(\exp(t\sigma_* x)\big)(p) = \theta\big(g_{s_p} \cdot \exp(tx) \cdot g_{s_p}\big)(p)$$
$$\Rightarrow \theta\big(\exp(tx)\big)(p) = \theta\big(g_{s_p} \cdot \exp(tx)\big)(p)$$
$$\text{if } \Upsilon(x)_p \neq 0, \text{ then the flow induced by } x \text{ yields: } \theta\big(\exp(tx)\big)(p) = q \text{ for small } t$$
$$\Rightarrow q = \theta\big(g_{s_p}\big)(q) = s_p(q).$$

which cannot hold, since s_p has as only fixed point $p \neq q$. It hence follows that $\Upsilon(x)_p = 0$.

Vice versa, if $\Upsilon(x)_p = 0$, one needs to show that $\sigma_* x = x$:

Let $q \neq p$ be a point close to p. There is hence a unit speed geodesic γ connecting p and q with $\gamma(0) = p$ and $\gamma(s) = q$. For the right hand side of eq. 83 follows:

$$\theta\big(g_{s_p} \cdot \exp(tx) \cdot g_{s_p}\big)(\gamma(s)) = \theta\big(g_{s_p} \cdot \exp(tx)\big)(s_p(\gamma(s)) = \theta\big(g_{s_p} \cdot \exp(tx)\big)\gamma(-s) = s_p(\tilde{\gamma}(-s)) = \tilde{\gamma}(s)$$

---

[16]because $\varphi : \ M \to I(M), \ m \mapsto s_m$ is continuous [Iozzi, 2018, Lemma 2.15].

[17]The commutativity of these isometries (labelling them f_i, $f_i : \mathfrak{g} \to G$) can be seen by noting that $f_1(0) = f_2(0)$, $Df_1\big|_0 = Df_2\big|_0 \Rightarrow f_1 \equiv f_2$.

since from $\Upsilon(x)_p = 0$ it follows that p is invariant under $\exp(tx)$, but it maps a geodesic γ isometrically to a different (WLOG unit speed) geodesic $\tilde{\gamma}$ with $\tilde{\gamma}(0) = p$.

For the left hand side assume that $\sigma_* x = -x$:

$$\theta\big(\exp(t\sigma_* x)\big)(\gamma(s)) = \theta\big(\exp(-tx)\big)(\gamma(s)) = \tilde{\gamma}(-s).$$

The two sides together yield: $\tilde{\gamma}(s) = \tilde{\gamma}(-s)$ which is contradictory as $\tilde{\gamma}$. It follows that the eigenvalue must be $+1$. Due to reasons of dimension of $\mathfrak{g}$, this shows that the two definitions in eq. 82 and eq. 65 do in fact coincide.

The statements from eq. 68 follow easily using $\phi_*[x,y] = [\phi_* x, \phi_* y]$, ϕ arbitrary smooth map. Choosing the Cartan involution as map, one arrives at: $\sigma_*[x,y] = [\sigma_* x, \sigma_* y] = (\lambda_1 \cdot \lambda_2)[x,y]$ with λ_i as corresponding eigenvalues.

We arrive at the following diagram which illustrates the correspondence between a symmetric manifold M and its representation as coset G/K where G and K are subspaces of the isometry group acting on M. This correspondence also lifts to a correspondence of their tangent spaces:

$$
\begin{array}{c}
G/K \cong M \\[6pt]
\exp \Big\uparrow \quad \exp \Big\uparrow \qquad \text{with } \mathfrak{g}/\mathfrak{k} = \{x + \mathfrak{k} \,|\, x \in \mathfrak{p}\}. \\[6pt]
\mathfrak{g}/\mathfrak{k} \cong T_p M
\end{array}
\tag{84}
$$

(a) Setup on the model space S^2 with clock to show orientation and green arrow indicating the base point $p = (0,0,1)^T$.
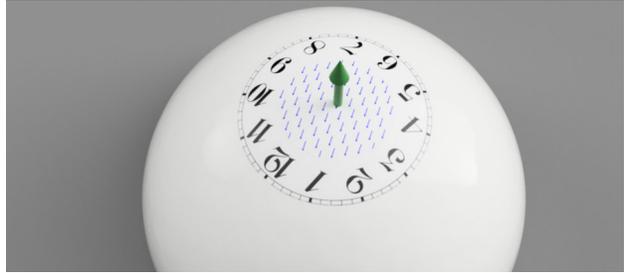
(b) After the application of s_p, the orientation of the clock has changed. The blue arrows represent the transvection Killing field.
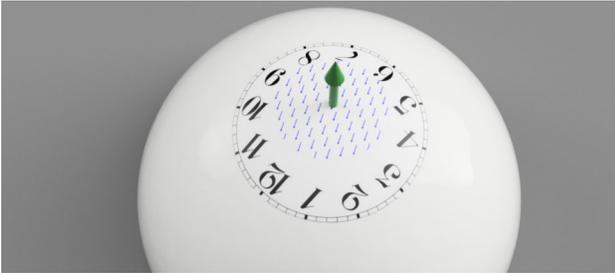
(c) After the application of $\theta(\exp_e(x))$, the clock has shifted in direction of the vector field.
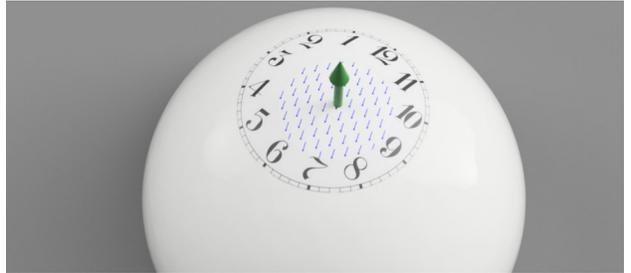
(d) A second application of s_p reflects the clock again. It has now shifted to the top.

Figure 17: Illustration of the Cartan involution of eq. 78. The element $g \in G$ is a transvection, its corresponding Killing vector field $\Upsilon(x)$, $x \in \mathfrak{p}$ points downwards. As one can see comparing 17a and 17d, the resulting transformation is a shift in the opposite direction of the small vector arrows. The eigenvalue of x under σ_* is hence -1.

10 Totally geodesic submanifolds

It is interesting to examine the structure of a subspace N of a symmetric space M. A very crucial property of a submanifold is to be *totally geodesic*: Consider a complete submanifold $N \subseteq M$. This submanifold is called totally geodesic, if for every pair (n, v), $n \in N$, $v \in T_n N$ the induced geodesic is contained in N. The shortest path between two points from a totally geodesic submanifold thus lies within this submanifold. Illustrating examples are the sphere and a plane within Euclidean space $\mathbb{R}^3$:
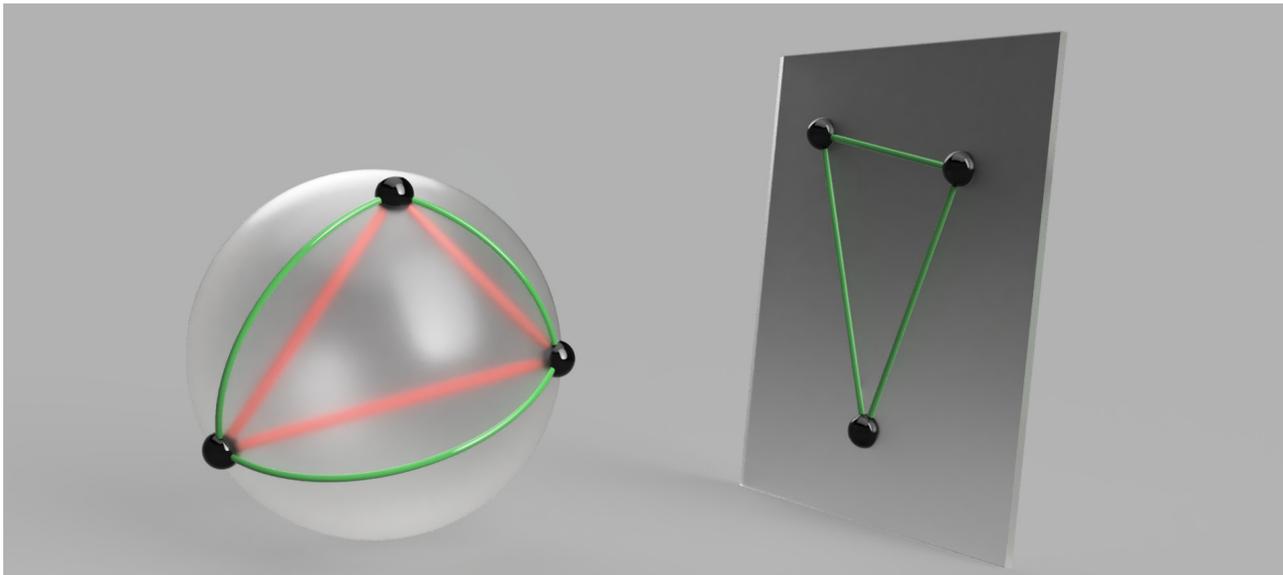
Figure 18: Visualization of the embedding of a sphere and a plane into $\mathbb{R}^3$.
On the left: Three points on the sphere are connected by green paths. These paths are geodesics within the sphere, but they are not geodesics with respect to the whole manifold $\mathbb{R}^3$, as these geodesics are shown by red straight paths. The sphere is hence not a totally geodesic submanifold of $\mathbb{R}^3$. Note however that the sphere is nevertheless positively curved, with the Euclidean metric tensor it inherits from the embedding.
On the right: An embedding of a plane within $\mathbb{R}^3$. The geodesics within this submanifold are also geodesics within the embedding space. This submanifold is thus a totally geodesic submanifold of $\mathbb{R}^3$.

A totally geodesic submanifold $N \subset M$ is generated by a subset $\mathfrak{p}' \subset \mathfrak{p}$ and is of the form $N = \theta\big(\exp_e(\mathfrak{p}')\big)(q)$. The point q is any arbitrary point contained in the manifold M which is often written as: $N = \exp(\mathfrak{p}')$.
Pictorially speaking, travelling in N is the same as travelling in M. This can be seen easily in fig. 18 above. Note that the Euclidean plane would not be a totally geodesic submanifold if it was embedded as plane $(x, y, 1)$ into H^3 (in the upper half-space model) with the inherited hyperbolic metric. This is due to the fact that geodesics in H^3 are half-circles standing perpendicular on $z = 0$. The shortest path between two points in the embedded Euclidean plane would hence leave the submanifold.

There is now a close correspondence between a Lie triple system and a totally geodesic submanifold. Firstly, recall the definition of a Lie triple product as in eq. 73:

$$[\cdot, [\cdot, \cdot]] : \mathfrak{p}' \times \mathfrak{p}' \times \mathfrak{p}' \to \mathfrak{g}, \ \ \mathfrak{p}' \subseteq \mathfrak{g}$$
$$(p_1, p_2, p_3) \mapsto [p_1, [p_2, p_3]] \tag{85}$$

If the image of this map is contained in $\mathfrak{p}'$, we call $\mathfrak{p}'$ a *Lie triple system*. We had found out that this is the case for both $\mathfrak{p}$ and trivially for $\mathfrak{k}$ in the normal Cartan decomposition [18] as in eq. 68. Now, one is interested in finding a non-trivial subset $\mathfrak{p}' \subset \mathfrak{p}$ which forms a Lie triple system to find interesting totally geodesic submanifolds which is justified by following theorem:

---

[18]Recall that $\mathfrak{p}$ is not a subalgebra of $\mathfrak{g}$, since $\mathfrak{p}$ is not invariant under the Lie bracket. In contrast, $\mathfrak{k}$ is a subalgebra. Both $\mathfrak{p}$ and $\mathfrak{k}$ are however invariant under the triple product.

> **Correspondence between Lie triple systems and totally geodesic submanifolds [Helgason, 1979, Theorem 7.2]**
>
> Let $\mathfrak{p}' \subset \mathfrak{p}$ be a Lie triple system, i.e. $[\mathfrak{p}', [\mathfrak{p}', \mathfrak{p}']] \subseteq \mathfrak{p}'$, then $N = \theta(\exp(t \cdot \mathfrak{p}'))$ is a totally geodesic submanifold of M.

Sketch of the proof:

Using that $\mathfrak{p}'$ is a Lie triple system, one obtains that $\mathfrak{g}' = \mathfrak{p}' + [\mathfrak{p}', \mathfrak{p}'] \subset \mathfrak{g}$ is a subalgebra of $\mathfrak{g}$ and determines a symmetric submanifold $N \subset M$ with transvection space $\mathfrak{p}'$. It follows that a geodesic $(p, \theta(\exp(t \cdot \mathfrak{p}')))$ is contained in N.

11 Rank of a symmetric space

The rank of a manifold is defined as the dimension of the space of parallel Jacobi vector fields and is hence a natural number which is globally defined. It is the highest dimension of a flat totally geodesic submanifold in M. The rank of a manifold is an important quantity of a symmetric space, because it gives a notion of the flatness of the manifold. In particular, a manifold is flat ($K = 0$) if and only if $\dim(M) = \text{rank}(M)$.

Since M is a homogeneous space, every point thus lies inside a flat submanifold of dimension equal to the rank of M.

> **Recap: Jacobi vector fields**
>
> A vector field X which is defined along an arbitrary geodesic γ is a Jacobi vector field, if it fulfills:
>
> $$\nabla^2_{\dot{\gamma}(t)} X - R(\dot{\gamma}(t), X)\dot{\gamma}(t) = 0 \tag{86}$$
>
> with $R(\cdot, \cdot)(\cdot)$ as Riemann curvature tensor. If one restricts the vector fields to parallel vector fields along γ (i.e. $\nabla_{\dot{\gamma}(t)} X = 0$), then the equation simplifies to:
>
> $$R(\dot{\gamma}(t), X)\dot{\gamma}(t) = 0. \tag{87}$$
>
> The dimension of parallel Jacobi vector fields is at least one, since the one-dimensional space $\lambda \cdot V = \lambda \cdot \dot{\gamma}(t)$ fulfills the equation. Trivially, the sectional curvature of the plane spanned by $(X, \dot{\gamma})$ is zero along γ. This can be seen from eq. 28:
>
> $$K(X, \dot{\gamma}) = \frac{g(-R(\dot{\gamma}, X)\dot{\gamma}, X)}{g(X, X)g(\dot{\gamma}, \dot{\gamma}) - g(X, \dot{\gamma})^2} = 0,$$
>
> with X Jacobi vector field, since the denominator is zero.
>
> Pictorially, a vector field being parallel to a curve can be understood as having a constant angle and magnitude with respect to that curve. Those vector fields which fulfill eq. 87 can be decomposed into parts parallel and orthogonal to the curve. The orthogonal Jacobi fields point into those directions orthogonal to $\dot{\gamma}$ in which the curvature does not change. Since $\nabla_{\gamma} R$ holds anyway for a symmetric space, the dimension spanned by all Jacobi vector fields represents the dimension of flat subspaces along the geodesic γ. Since the orthogonal Jacobi vector fields run along geodesics perpendicular to γ, this flat subspace is a totally geodesic submanifold.

It can be very cumbersome to determine the rank of M via parallel Jacobi fields. It is more elegant, to work in the picture of the Lie algebra $\mathfrak{g}$:

Recall eq. 73

$$R(V_1, V_2)V_3 = [V_3, [V_1, V_2]], \quad V_i \in \Upsilon(\mathfrak{p}), q.$$

Using that Υ is an anti-isomorphism, one obtains:

$$R(V_1, V_2)V_3 = [\Upsilon(p_3), [\Upsilon(p_1), \Upsilon(p_2)]]$$
$$= [\Upsilon(p_3), -\Upsilon([p_1, p_2])] = \Upsilon([p_3, [p_1, p_2]]), \quad V_i = \Upsilon(p_i), \quad p_i \in \mathfrak{p}. \tag{88}$$

Choosing a subspace $\mathfrak{a} \subseteq \mathfrak{p}$ such that the Lie bracket vanishes means that algebra elements commute, i.e. the Lie algebra $\mathfrak{a}$ needs to be abelian. It follows that the Riemann tensor vanishes on that generated subspace, it is hence flat. This flat submanifold is also totally geodesic, since $\mathfrak{a}$ is trivially closed under the Lie triple product as $[\mathfrak{a}, [\mathfrak{a}, \mathfrak{a}]] = 0 \in \mathfrak{a}$.

The dimension of the maximal abelian subalgebra $\mathfrak{a} \subseteq \mathfrak{p}$ thus determines the rank of the manifold.

12   A product of symmetric spaces

In the following section properties of a Cartesian product of a symmetric space are examined. These properties easily follow from the fact that all of the constituents are symmetric spaces. This is the reason why literature focuses on indecomposable[19] symmetric spaces.

Let $(M_i, s_{M_i}, G_i, K_i, \mathfrak{g}_i, \mathfrak{p}_i, \mathfrak{k}_i, \sigma_i, (\sigma_i)_*, \mathfrak{a}_i), \ \ i \in \{1, 2\}$ be symmetric spaces of dimension n_i as defined before. The meaning of those symbols is summarized:

Symbol	Name	Meaning
M	symmetric space M	Riemannian manifold with isometric geodesic symmetry
s_p	symmetry at base point p	isometric geodesic reflection at $p \in M$
G	isometry group on M	identity containing group of all isometric maps acting on M
K	isotropy group on M	group of all isometries, which leave a base point p invariant
$\mathfrak{g}$	Lie algebra of G	vector space of all infinitesimal isometries on M
$\mathfrak{p}$	subspace of transvections	space of all infinitesimal shifts at base point p
$\mathfrak{k}$	subspace of rotations	space of all infinitesimal rotations at p
σ	Cartan involution on G at p	map sending isometry to isometry by reflecting before and afterwards
σ_*	lifted Cartan involution on $\mathfrak{g}$	Cartan involution on the space of infinitesimal isometries
$\mathfrak{a}$	abelian subalgebra of $\mathfrak{p}$	subspace of $\mathfrak{g}$ which generates a flat subspace via $\theta(\exp_e(\mathfrak{a}))$

Table 1: Table showing the index of notation with a short explanation

One makes following observations:

- $M_1 \times M_2$ is a symmetric space: The geodesic symmetry on $M_1 \times M_2$ is just (s_{M_1}, s_{M_2}) where the symmetries act on the first n_1 and last n_2 components, respectively.

- For the coset formulation holds: $M_1 \times M_2 \cong G_1/K_1 \times G_2/K_2 \cong (G_1 \times G_2)/(K_1 \times K_2)$.

- The isometry group G and the isotropy group K act component-wise. The dimension of G is hence $\dim G_1 + \dim G_2$ as the coset formulation suggests. Analogously: $\dim K = \dim K_1 + \dim K_2$.

- The Lie algebra $\mathfrak{g}$ of G consists of all infinitesimal transformations of both spaces. It decomposes as:

$$\mathfrak{g} = (g_1, g_2), \ \ g_i \in \mathfrak{g}_i, \ \ \text{because } \mathfrak{g} = \mathfrak{g}_1 \oplus \mathfrak{g}_2. \tag{89}$$

The Cartan decomposition of $\mathfrak{g} = \mathfrak{p} \oplus \mathfrak{k}$ holds with $\mathfrak{p} = (p_1, p_2), \ p_i \in \mathfrak{p}_i$ and $\mathfrak{k} = (k_1, k_2), \ k_i \in \mathfrak{k}_i$. The algebra operations on $\mathfrak{g}$ lead to eq. 68:

$$\left[\begin{pmatrix} p_1 \\ p_2 \end{pmatrix}, \begin{pmatrix} \tilde{p}_1 \\ \tilde{p}_2 \end{pmatrix} \right] = \begin{pmatrix} [p_1, \tilde{p}_1] \\ [p_2, \tilde{p}_2] \end{pmatrix} \in \mathfrak{k} \ \ \text{and analogously} \ \ [\mathfrak{k}, \mathfrak{k}] \subseteq \mathfrak{k} \ \ [\mathfrak{p}, \mathfrak{k}] \subseteq \mathfrak{p}. \tag{90}$$

- The elements (p_1, k_2) or (k_1, p_2) neither belong to $\mathfrak{p}$ nor to $\mathfrak{k}$. This does not contradict to $\mathfrak{g} = \mathfrak{p} \oplus \mathfrak{k}$.

- The Cartan involution σ is lifted component-wise. Its derivative σ_* does again possess eigenvalues $\{-1, 1\}$, for $\mathfrak{p}$ and $\mathfrak{k}$, respectively. An element $j = (p_1, k_2)$ however is not an eigenvector for σ_* since:

$$\sigma_* \begin{pmatrix} p_1 \\ k_2 \end{pmatrix} = \begin{pmatrix} (\sigma_1)_* p_1 \\ (\sigma_2)_* k_2 \end{pmatrix} = \begin{pmatrix} -p_1 \\ k_2 \end{pmatrix}. \tag{91}$$

---

[19]or to be less restrictive: isotropy irreducible

This is obvious, since j is not contained in an eigenspace of any eigenvalue. The same holds of course for an element of $(\mathfrak{k}_1, \mathfrak{p}_2)$.

- The abelian subalgebras have the form:

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \quad a_i \in \mathfrak{a}_i \text{ because } \left[\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \begin{pmatrix} \tilde{a}_1 \\ \tilde{a}_2 \end{pmatrix} \right] = \begin{pmatrix} [a_1, \tilde{a}_1] \\ [a_2, \tilde{a}_2] \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{92}$$

In particular, this means that (p_1, p_2) is an abelian subalgebra, if both p_1 and p_2 are one-dimensional since they then commute with themselves. The rank of a product manifold is $\text{rank}(M_1) + \text{rank}(M_2)$ and hence at least 2, since the rank of any symmetric space is at least 1.

- The metric tensor decomposes into the metric tensors of the factors. Therefore, the distance between two points in $M_1 \times M_2$ is easily computed as:

$$d(p, q) = \sqrt{d_1^2(p_1, q_1) + d_2^2(p_2, q_2)} \text{ with respective distances } d_i. \tag{93}$$

Recall the geodesic equation:

$$\nabla_{\dot{\gamma}} \dot{\gamma} \stackrel{!}{=} 0. \tag{94}$$

The decomposition of the metric tensor ensures that geodesics can be decomposed into geodesics within the product spaces.

$$\nabla_{\dot{\gamma}} \dot{\gamma} = \nabla_{\dot{\gamma}_1 + \dot{\gamma}_2}(\dot{\gamma}_1 + \dot{\gamma}_2) = \nabla_{\dot{\gamma}_1} \dot{\gamma}_2 + \nabla_{\dot{\gamma}_2} \dot{\gamma}_1 = 0, \tag{95}$$

where the second equality is due to the fact that γ_1 and γ_2 are geodesics, the last equality holds since the values of γ_i (first summand below) and the entries of the metric tensor which belong to space M_i (second summand below) are not dependent on the coordinates of M_j and vice versa:

$$\nabla_{\dot{\gamma}_i} \dot{\gamma}_j = \dot{\gamma}_i{}^\mu \partial_\mu \dot{\gamma}_j{}^\nu \partial_\nu + \dot{\gamma}_i{}^\mu \dot{\gamma}_j{}^\nu \Gamma^\kappa_{\mu\nu} \partial_\kappa = 0, \quad i \neq j. \tag{96}$$

- The Riemann tensor decomposes into its respective Riemann tensors (R_i):

$$R(X, Y)Z = (R_1)(X_1, Y_1)Z_1 + (R_2)(X_2, Y_2)Z_2 \tag{97}$$

with decompositions of the vector fields into their components in M_1 and M_2.

13 Examples of "simple" symmetric spaces

13.1 The sphere S^2

In section 3.2 we have already investigated the symmetries on S^n. For the 2-sphere embedded in $\mathbb{R}^3$ the symmetry at $p = (0, 0, 1)^T$ is:

$$\begin{aligned} s_p : S^2 &\to S^2 \\ (x, y, z)^T &\mapsto (-x, -y, z)^T. \end{aligned} \tag{98}$$

As the sphere often served as an example space before, we know that a n-sphere has a coset structure of $SO(n)/SO(n-1)$ which leads to $S^2 \cong SO(3)/SO(2)$ and $T_p(SO(3)/SO(2)) = \mathfrak{so}(3)/\mathfrak{so}(2)$.

The Lie algebra $\mathfrak{so}(3)$ corresponds to the set of all skew-symmetric 3×3 matrices and hence $\mathfrak{so}(2)$ to the set of skew-symmetric 2×2 matrices. Choosing $p = (0, 0, 1)^T$ the Lie algebra decomposes as:

$$\mathfrak{p} = \left\{ \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & \alpha \\ 0 & -\alpha & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & \beta \\ 0 & 0 & 0 \\ -\beta & 0 & 0 \end{pmatrix} \,\middle|\, \alpha, \beta \in \mathbb{R} \right\} \text{ and } \mathfrak{k} = \left\{ \begin{pmatrix} 0 & \gamma & 0 \\ -\gamma & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \,\middle|\, \gamma \in \mathbb{R} \right\} \tag{99}$$

where the elements in $\exp_e(\mathfrak{p})$ rotate about angles α, β around x-, y- axes, respectively. Accordingly, an element $\exp_e(k), k \in K$ is a rotation around the z-axis. Note that $\mathfrak{k}$ really consists of 2×2 matrices within a 3×3 system. The Cartan involution follows as:

$$\sigma_p : SO(3) \to SO(3)$$

$$g \mapsto \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot g \cdot \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{100}$$

Its derivative $(\sigma_p)_*$ acts on $\mathfrak{p}$ and $\mathfrak{k}$ with eigenvalues $-1, 1$, respectively. This has been motivated in fig. 17. The abelian subspaces of $\mathfrak{p}$ are one-dimensional, hence trivial. The rank of S^2 is hence 1. This fits to intuition. The totally geodesic submanifolds of rank 1 are precisely the great circles S^1.

13.2 The symmetric space H^2

Hyperbolic 2-space has a coset structure of $H^2 \cong SO(2,1)/SO(2)$ in the hyperboloid model. [20] We may choose the point $p = (0,0,1)^T$ in the hyperboloid embedding. This leads to $T_p\big(SO(2,1)\big)/SO(2)\big) = \mathfrak{so}(2,1)/\mathfrak{so}(2)$. The Lie algebra $\mathfrak{so}(2,1)$ corresponds to the set of matrices of the form:

$$\mathfrak{so}(2,1) = \left\{ \begin{pmatrix} A & x \\ x^T & 0 \end{pmatrix} \,\middle|\, A^T = -A, \ A \in M_{2\times 2}(\mathbb{R}) \right\}. \tag{101}$$

Again, $\mathfrak{so}(2)$ is the set of skew-symmetric 2×2 matrices. Choosing $p = (0,0,1)^T$ the Lie algebra decomposes as:

$$\mathfrak{p} = \left\{ \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & \alpha \\ 0 & \alpha & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & \beta \\ 0 & 0 & 0 \\ \beta & 0 & 0 \end{pmatrix} \,\middle|\, \alpha, \ \beta \in \mathbb{R} \right\} \text{ and } \mathfrak{k} = \left\{ \begin{pmatrix} 0 & \gamma & 0 \\ -\gamma & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \,\middle|\, \gamma \in \mathbb{R} \right\} \tag{102}$$

where the elements in $\exp_e(\mathfrak{p})$ rotate about angles α, β around x-, y- axes, respectively. Note the difference between this Cartan decomposition and the Cartan decomposition of S^2 in eq. 99: The former emerges from rotations in Minkowski embedding space, the latter from rotations in Euclidean embedding space.

Since the geodesic symmetry at p sends (x,y,z) to $(-x,-y,z)$, the Cartan involution follows as:

$$\sigma_p : SO(2,1) \to SO(2,1)$$

$$g \mapsto \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot g \cdot \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{103}$$

The derivative $(\sigma_p)_*$ has eigenvalues $\{-1,1\}$ for $\mathfrak{p}$ and $\mathfrak{k}$, respectively.

The abelian subspaces of $\mathfrak{p}$ are one-dimensional, hence trivial. Thus, the rank of H^2 is 1. This again fits to intuition. The totally geodesic submanifolds of rank 1 are precisely the geodesics in H^2.

13.3 The symmetric space $H^2 \times \mathbb{R}$

As pointed out before, hyperbolic space is a symmetric space. A product $M = M_1 \times M_2$ of symmetric spaces $M_i = G_i/K_i$ is also a symmetric space. Algebraically, M is equivalent to $G_1/K_1 \times G_2/K_2$. The coset formulation of $\mathbb{R}$ is just $\mathbb{R} \cong \mathcal{E}(1)/SO(1)$. Since $SO(1)$ consists only of $\{1\}$, $\mathcal{E}(1)$ is just the one-dimensional group of translations, as there are no rotations in 1 dimension. It follows that $H^2 \times \mathbb{R} \cong SO(2,1)/SO(2) \times \mathcal{E}(1)/SO(1)$. At a point $p = (p_1, p_2)$, $p_i \in M_i$ the symmetry s_p acts on q component-wise as $s_p(q) = (s_{1p_1}(q_1), s_{2p_2}(q_2))$. The metric tensor does not posess any cross terms, since H^2 and $\mathbb{R}$ stand perpendicular on each other. Choosing the Poincaré disk model for the H^2, the metric tensor becomes:

$$g_{\mu\nu} = \begin{pmatrix} \frac{4}{(1-r^2)^2} & 0 & 0 \\ 0 & \frac{4}{(1-r^2)^2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ in Cartesian coordinates } (x,y,z) \text{ with } r^2 = x^2 + y^2, \ x,y \in H^2. \tag{104}$$

[20]We choose it for simplicity for this section. However, the discussion would of course also be possible in the other models.

13.3.1 Where are the flat subspaces?

Since a straight path within a hyperbolic space is flat in the sense that one does not observe bending of the geodesic path when travelling along it, it thus corresponds to $\mathbb{R}$. Because of the Cartesian product in $H^2 \times \mathbb{R}$, the Cartesian product of two one-dimensional geodesics is a 2-dimensional flat Euclidean space $\mathbb{R}^2$. A combination of any two geodesics from both of the hyperbolic spaces thus forms a Euclidean subspace.

This can be formulated more rigorously in terms of subspaces of the Lie algebra of the group of isometries as in eq. 92. Any abelian subalgebra of H^2 commutes with the abelian subalgebra of $\mathbb{R}$. The abelian subalgebra $\mathfrak{a}_{H^2} \subset \mathfrak{p}_{H^2}$ is only one dimensional and consists of infinitesimal transvections ($\rightarrow$ geodesics). The abelian subalgebra $\mathfrak{a}_{\mathbb{R}} \subseteq \mathfrak{p}_{\mathbb{R}}$ is identical to $\mathfrak{p}$, since $\mathfrak{p}$ itself is only one-dimensional. The flat totally geodesic subspaces are thus:

$$\exp_p \begin{pmatrix} \mathfrak{a}_{H^2} \\ \mathfrak{p}_{\mathbb{R}} \end{pmatrix}, \quad \mathfrak{a}_{H^2} \text{ denotes abelian, hence one-dimensional subalgebra of } \mathfrak{p}_{H^2}. \tag{105}$$

13.3.2 Where are the $K = -1$ negatively curved subspaces?

As seen above, a plane which is spanned by two geodesics which live in different H^2 spaces is equivalent to $\mathbb{R}^2$. However, a plane spanned by two geodesics which both live in H^2 obviously do not form a flat space, but a negatively curved space of fixed curvature $K = -1$.

These spaces are totally geodesic submanifolds, as their generating subspace $\mathfrak{p} \subseteq \mathfrak{g}$ forms a Lie triple system. This makes perfect sense, because it is just the transvection space of H^2. The $K = -1$ totally geodesic subspaces are thus:

$$\exp_p \begin{pmatrix} \mathfrak{p}_{H^2} \\ 0 \end{pmatrix}. \tag{106}$$

13.3.3 Visualization of $H^2 \times \mathbb{R}$

2-dimensional hyperbolic space can be visualized with aid of the Poincaré disk model. If one puts the Euclidean space $\mathbb{R}$ perpendicular to this disk, $H^2 \times \mathbb{R}$ has the shape of a cylinder. The geodesics in H^2 are the usual geodesics in the x-y-plane with arbitrary constant z. Accordingly, the geodesics in $\mathbb{R}$ are straight lines in z-direction with arbitrary x and y values within the disk.
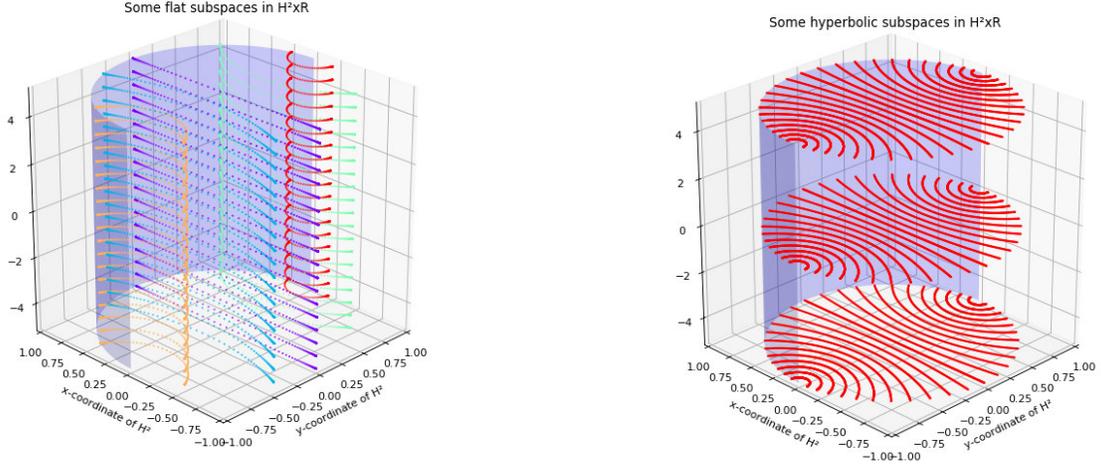
(a) These planes are spanned by vectors which live in H^2 and $\mathbb{R}$. Due to the symmetry of the disk, any plane which can be obtained by rotation along the z-axis is also a flat subspace.

(b) Hyperbolic subspaces with geodesics within the subspaces. Of course, any other x-y-plane is also a hyperbolic subspace of sectional curvature $K = -1$.

Figure 19: Special subspaces in $H^2 \times \mathbb{R}$. Note that in the left picture the points which form the geodesics are equidistant, i.e. along the vector ∂_x the distance between two points is constant. Of course, two planes are not parallel, i.e. the distance between points of different coloring increases and approaches infinity as $r \to 1$.

13.3.4 Planes of non-constant sectional curvature

In the section above, we have investigated the flats and the subspaces with $K = -1$. It is well known that the sectional curvature is confined to $[-1, 0]$ for $H^2 \times \mathbb{R}$ [Gu et al., 2019, Lemma 1]. To get a feel for that, one can choose two arbitrary vectors at an arbitrary point and calculate the sectional curvature. It is easiest for the calculation to choose the origin as evaluation point. Since the space is homogeneous, the result holds for all other points.

Choosing the vectors $v, w \in T_{p=(0,0,0)}(H^2 \times \mathbb{R})$ with mixed terms such as

$$v = a \cdot \partial_x + b \cdot \partial_z \text{ and } w = \partial_y, \tag{107}$$

the vector v does not only live in the Euclidean part of the tangent space, but also in the hyperbolic part. With aid of the scalar values a and b one can manipulate the tilting of the plane. Via eq. 28 one can calculate the sectional curvature of the plane spanned by v, w within $T_{p=(0,0,0)}(H^2 \times \mathbb{R})$. Using the correspondence between the Lie algebra $\mathfrak{p}$ and Killing vector fields, one can simplify the calculation with eq. 73. The result is:

$$K = -\frac{4a^2}{4a^2 + b^2}. \tag{108}$$

Naturally, the sectional curvature K is fixed to the interval $[-1, 0]$ for $b = 0$ and $a = 0$, respectively.

To construct such a smooth submanifold such that at $p = (0, 0, 0)$ the vectors v and w from eq. 107 belong to the tangent space, one may use the map

$$F : H^2 \to H^2 \times \mathbb{R} \tag{109}$$

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x \\ y \\ f(x, y) \end{pmatrix} \tag{110}$$

37

and the induced vectors $v = \partial_x F(x, y)$ and $w = \partial_y F(x, y)$. The following figure is obtained by setting $f(x, y) = f_m(x) = 2 \cdot m \cdot \operatorname{arctanh}(x)$ and hence $v = \partial_x + 2m\partial_z$ and $w = \partial_y$ with $v, w \in T_{(0,0,0)} H^2 \times \mathbb{R}$.
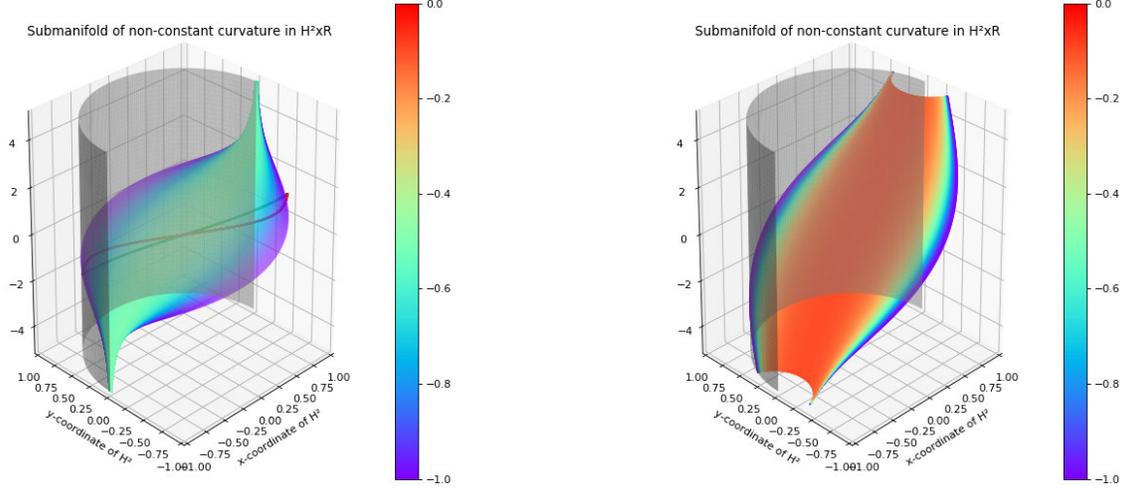
(a) Submanifold of non-constant negative sectional curvature. This submanifold is not totally geodesic: The green path is a geodesic within the submanifold, the red path is a geodesic in the full embedding space. The red path is shorter with respect to the metric.

(b) As the metric tensor tends to infinity at the disk boundary, the hyperbolic part of the spanning vectors dominates the Euclidean part. The sectional curvature thus tends to $K = -1$ at the boundary. Note that the submanifold in this figure extends to infinity in z-direction.

Figure 20: Sectional curvature of submanifolds within $H^2 \times \mathbb{R}$ indicated by coloring. On the left: Submanifold generated by $F(x, y) = (x, y, 2 \cdot \operatorname{arctanh}(x))$ from eq. 109 with $m = 1$ and vectors $v = \partial_x + 2(1 - x^2)^{-1}\partial_z$ and $w = \partial_y$. At the origin $p = (0, 0, 0)$ these vectors become $v = \partial_x + 2 \cdot \partial_z$ and $w = \partial_y$ with $K_p = -0.5$ from eq. 108.
On the right: Submanifold induced by $F(x, y) = (x, y, 2 \cdot 3 \cdot \operatorname{arctanh}(x))$ with $m = 3$ and vectors $v = (1, 0, 2 \cdot 3(1 - x^2)^{-1})^T$ and $w = (0, 1, 0)^T$. At the origin these vectors become $v = \partial_x + 6 \cdot \partial_z$ and $w = \partial_y$ with $K_p = -0.1$. One may have a look at fig. 9 to recall distortion of the Poincaré disk model.

Are these subspaces totally geodesic?

The spaces which are spanned by these vectors are no totally geodesic submanifolds. A geodesic connecting two points of this submanifold is thus not necessarily contained within the submanifold which is shown explicitly: Consider the transvection space of $H^2 \times \mathbb{R}$ written in components as:

$$\begin{pmatrix} \mathfrak{p}_{H^2} \\ \mathfrak{p}_{\mathbb{R}} \end{pmatrix}. \tag{111}$$

Define

$$\left\{ \begin{pmatrix} \alpha \cdot a \cdot p_x + \beta \cdot p_y \\ \alpha \cdot b \cdot p_z \end{pmatrix} \right\}, \quad p_x, p_y, p_z \text{ forming a basis of } \begin{pmatrix} \mathfrak{p}_{H^2} \\ \mathfrak{p}_{\mathbb{R}} \end{pmatrix} \tag{112}$$

as two-dimensional linear span of the respective transvection spaces. The Greek letters are parameters which span the linear subspace of $\mathfrak{g}$, the Latin letters determine the ratio of the respective directions within this space. It turns out that this is not a Lie triple system:
If it were a Lie triple system, then for all $a, b, \alpha_i, \beta_i, i \in \{1, 2, 3\}$ there would exist $\alpha_4, \beta_4 \in \mathbb{R}$ such that

$$\left[\begin{pmatrix} \alpha_1 \cdot a \cdot p_x + \beta_1 \cdot p_y \\ \alpha_1 \cdot b \cdot p_z \end{pmatrix}, \left[\begin{pmatrix} \alpha_2 \cdot a \cdot p_x + \beta_2 \cdot p_y \\ \alpha_2 \cdot b \cdot p_z \end{pmatrix}, \begin{pmatrix} \alpha_3 \cdot a \cdot p_x + \beta_3 \cdot p_y \\ \alpha_3 \cdot b \cdot p_z \end{pmatrix} \right] \right] = \begin{pmatrix} \alpha_4 \cdot a \cdot p_x + \beta_4 \cdot p_y \\ \alpha_4 \cdot b \cdot p_z \end{pmatrix} \tag{113}$$

is fulfilled. However, explicit calculation yields:

$$
\begin{aligned}
&\left[\begin{pmatrix} \alpha_1 \cdot a \cdot p_x + \beta_1 \cdot p_y \\ \alpha_1 \cdot b \cdot p_z \end{pmatrix}, \left[\begin{pmatrix} \alpha_2 \cdot a \cdot p_x + \beta_2 \cdot p_y \\ \alpha_2 \cdot b \cdot p_z \end{pmatrix}, \begin{pmatrix} \alpha_3 \cdot a \cdot p_x + \beta_3 \cdot p_y \\ \alpha_3 \cdot b \cdot p_z \end{pmatrix} \right] \right] \\
&= \left[\begin{pmatrix} \alpha_1 \cdot a \cdot p_x + \beta_1 \cdot p_y \\ \alpha_1 \cdot b \cdot p_z \end{pmatrix}, \begin{pmatrix} [\alpha_2 a p_x, \beta_3 p_y] + [\beta_2 p_y, \alpha_3 a p_x] \\ 0 \end{pmatrix} \right] \\
&= \left[\begin{pmatrix} \alpha_1 \cdot a \cdot p_x + \beta_1 \cdot p_y \\ \alpha_1 \cdot b \cdot p_z \end{pmatrix}, \begin{pmatrix} (a\alpha_2\beta_3 - a\alpha_3\beta_2)[p_x, p_y] \\ 0 \end{pmatrix} \right] \\
&= \begin{pmatrix} \alpha_1 a(a\alpha_2\beta_3 - a\alpha_3\beta_2)\big[p_x, [p_x, p_y]\big] + \beta_1(a\alpha_2\beta_3 - a\alpha_3\beta_2)\big[p_y, [p_x, p_y]\big] \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} \beta_1(\alpha_2\beta_3 - \alpha_3\beta_2) \cdot a \cdot p_x + \alpha_1 a(a\alpha_2\beta_3 - a\alpha_3\beta_2) \cdot p_y \\ 0 \end{pmatrix} \notin \begin{pmatrix} \alpha_4 \cdot a \cdot p_x + \beta_4 \cdot p_y \\ \alpha_4 \cdot b \cdot p_z \end{pmatrix}
\end{aligned}
\tag{114}
$$

which shows that the Lie triple product is not contained within itself for all α_i and hence not a totally geodesic submanifold.

In fact, the above computation shows that there are no more 2-dimensional totally geodesic submanifolds than the flats and the space H^2. This is due to the fact that the space $\mathfrak{p}_\mathbb{R}$ is abelian. Any mixed terms in the transvection space hence vanish in the Lie triple product.

13.3.5   Planes of constant negative sectional curvature

In sec. 13.3.4 submanifolds of non-constant curvature have been shown. It is interesting to know if and how submanifolds of constant negative curvature K can be embedded in $H^2 \times \mathbb{R}$. This subspace is locally isomorphic to $H^2_{\sqrt{-K^{-1}}}$, but the embedding is not totally geodesic.[21]

Using the construction as above in eq. 109, one needs to solve for $f(x,y)$ such that the sectional curvature is constant:

For this one chooses cylindrical coordinates such that the metric tensor becomes:

$$
g_{\mu\nu} = \begin{pmatrix} \frac{4}{(1-r^2)^2} & 0 & 0 \\ 0 & \frac{4}{(1-r^2)^2} \cdot r^2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{in coordinates } (r, \phi, z).
\tag{115}
$$

The map F from eq. 109 becomes:

$$
F : \begin{pmatrix} r \\ \phi \end{pmatrix} \mapsto \begin{pmatrix} r \\ \phi \\ f(r, \phi) \end{pmatrix} \quad \text{with resulting vectors } v = (1, 0, \partial_r f)^T, w = (0, \frac{1}{r}, \partial_\phi f)^T.
\tag{116}
$$

The solution to the equation below is in particular easy, if one chooses to set $f(r, \phi) = f(r)$. The sectional curvature on every point $q = (r, \phi, z)$ can be calculated via eq. 28:

$$
\begin{aligned}
K(r, \phi, z) &= \frac{g(R(v,w)w,v)}{g(v,v)g(w,w) - g(v,w)^2} = \frac{g_{\alpha\beta}R^\alpha{}_{\gamma\delta\epsilon}v^\delta w^\epsilon w^\gamma v^\beta}{g(v,v)g(w,w) - g(v,w)^2} = \frac{R_{\alpha\gamma\delta\epsilon}v^\delta w^\epsilon w^\gamma v^\alpha}{g(v,v)g(w,w) - g(v,w)^2} \\
&= \frac{R_{1212} \cdot 1 \cdot \frac{1}{r} \cdot \frac{1}{r} \cdot 1}{\left(\frac{4}{(1-r^2)^2} + (\partial_r f(r))^2\right)\frac{4}{(1-r^2)^2} - 0} = \frac{-\left(\frac{4}{(1-r^2)^2}\right)^2 \cdot r^2 \cdot \frac{1}{r^2}}{\left(\frac{4}{(1-r^2)^2}\right)^2 \cdot \left(1 + \frac{(1-r^2)^2}{4}(\partial_r f(r))^2\right)} \\
&= -\frac{1}{1 + \frac{(1-r^2)^2}{4}\left(\partial_r f(r)\right)^2}.
\end{aligned}
\tag{117}
$$

In order to keep K constant, one finds the solutions:

$$
f(r) = \pm 2 \cdot m \cdot \operatorname{arctanh}(r) \quad \text{with } m \text{ as free parameter to determine } K \text{ via } K = -\frac{1}{1+m^2}.
\tag{118}
$$

---

[21]In fact, it is not even a smooth manifold at the cone as one can see in fig. 21.

Note that the vectors $v = (1, 0, \partial_r f)^T$, $w = (0, \frac{1}{r}, \partial_\phi f)^T$ are not well-defined at $r = 0$. The sectional curvature at the origin is hence also not defined. The space is hence in a strict sense not diffeomorphic to hyperbolic space of corresponding negative curvature.[22] However, except for this point, the space looks like $H_{\sqrt{-K^{-1}}}$.

The submanifold of constant sectional curvature is not a totally geodesic submanifold[23] of $H^2 \times \mathbb{R}$. After transforming the map F to Cartesian coordinates again, the vectors v, w become:

$$v = (1, 0, \frac{2x}{(1-r^2)^2 r})^T \text{ and } w = (0, 1\frac{2x}{(1-r^2)^2 r})^T. \tag{119}$$

The corresponding elements of the transvection space $\mathfrak{p}$ do not induce a Lie triple product, since again $[p_z, p_z]$ vanishes as in eq. 114 before. One can also see the submanifold not being totally geodesic by looking at fig. 21a: The geodesics follow the curvature of the whole space $H^2 \times \mathbb{R}$ and hence deviate from the shortest paths within the submanifold.

---

[22]Every complete simply connected space of constant negative curvature K is diffeomorphic to $H^2_{\sqrt{-K^{-1}}}$. Since the submanifold's sectional curvature is not defined at $r = 0$, this is hence not the case.

[23]Disregarding the fact that it is not even a smooth manifold, hence the geodesics are not well-defined at the cone point.

Visualization of the embedding of constant sectional curvature in $H^2 \times \mathbb{R}$
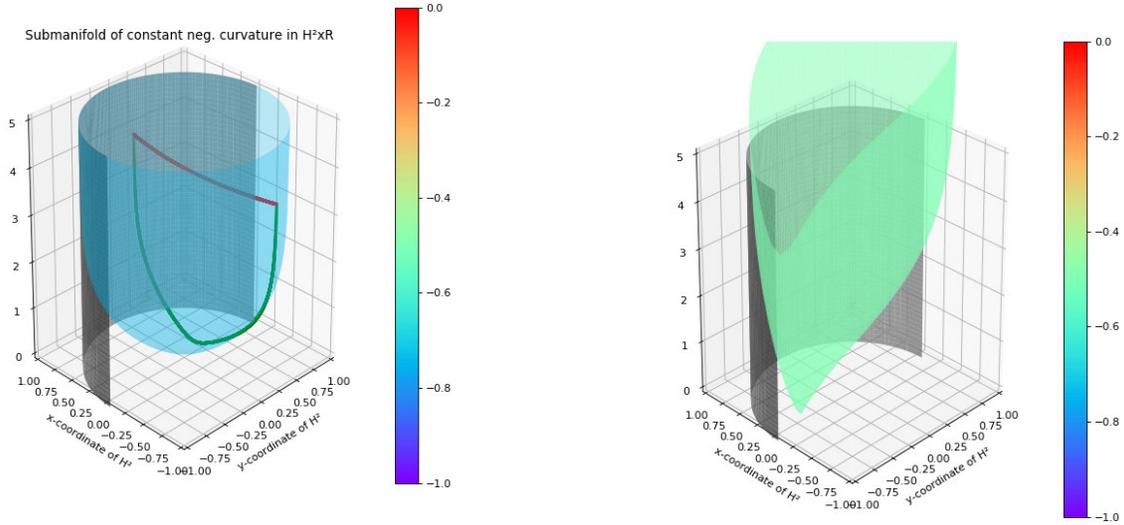
(a) Embedding locally isomorphic to $H^2_{\sqrt{4/3}}$ within $H^2 \times \mathbb{R}$. This submanifold is not totally geodesic: The green path is a geodesic within the submanifold, but not a geodesic within the embedding manifold as the red path. The red path is shorter.

(b) Subspace of sectional curvature $K = -0.5$ after a change of coordinates. This illustrates that the subspace is perfectly regular everywhere except for the cone point.
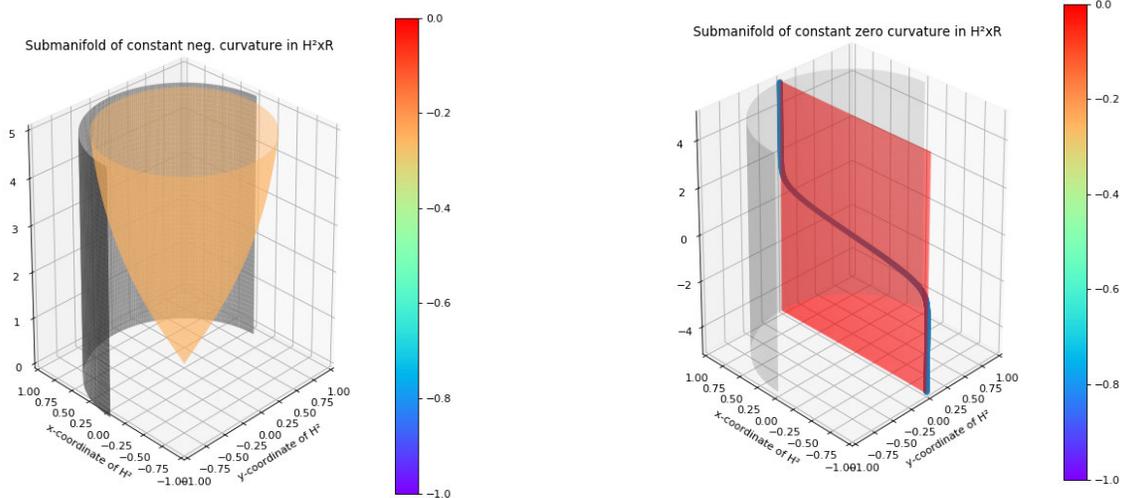
(c) Sectional curvature of a submanifold with $K = -0.25$. Note that the funnel surface extends to infinity in z-direction approaching the boundary $r = 1$.

(d) Sectional curvature along a plane which possesses curvature zero with geodesic. Note that the geodesic is a straight line with respect to the curved metric. This submanifold is totally geodesic.

Figure 21: Sectional curvature of submanifolds of constant curvature within $H^2 \times \mathbb{R}$ with hyperbolic space shown in the Poincaré model. The coloring of the surface indicates its sectional curvature. The cylindrical shape in the background indicates the full volume of $H^2 \times \mathbb{R}$.

13.4 The symmetric space $S^2 \times \mathbb{R}$

The sphere S^2 is naturally embedded in $\mathbb{R}^3$, but recall that it is a two-dimensional surface. The space $S^2 \times \mathbb{R}$ is hence three-dimensional and can be embedded in $\mathbb{R}^3$. For visualization purposes we consider S^2 in the stereographic projection model as introduced in section 3.2.

As seen in the example above, symmetries and metric tensor decompose:

$$s_{p=(0,0,0)}(x, y, z)^T = (-x, -y, -z)^T \tag{120}$$

$$g_{\mu\nu} = \begin{pmatrix} \frac{4}{(1+r^2)^2} & 0 & 0 \\ 0 & \frac{4}{(1+r^2)^2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ in coordinates } x, y, z, \ r^2 = x^2 + y^2, \ x, y \in S^2 \text{ in stereographic proj.} \tag{121}$$

13.4.1 Where are the flat subspaces?

To examine the curvature of two-dimensional subspaces, one needs to investigate the geodesics on the sphere. In the natural embedding of $S^2 \subset \mathbb{R}^3$, geodesics on the sphere are great-circles, i.e. circles whose midpoint is the origin of $\mathbb{R}^3$. Every geodesic resembles the circle S^1 but locally it looks like $\mathbb{R}$. A surface which is spanned by a geodesic on S^2 and a perpendicular geodesic in $\mathbb{R}$ thus locally looks like $\mathbb{R}^2$, but globally the surface is a cylinder: In fact, it is just $S^1 \times \mathbb{R}$: Travelling along the sphere, one eventually passes the starting point. This corresponds to travelling around the cylinder. Travelling in $\mathbb{R}$-direction leads into positive or negative infinity. This corresponds to travelling upwards or downwards the cylinder. Note that the sectional curvature of the cylinder is flat and that the cylinder locally looks like $\mathbb{R}^2$.

In the stereographic projection, geodesics are ellipses (or in the degenerate case straight lines through the origin). Because $\mathbb{R}$ stands perpendicular to these ellipses the flat subspaces are squeezed cylinders. The squeezing arises from the distortion of the stereographic projection.

The flat totally geodesic subspaces correspond to a vanishing Lie triple product on the generating subalgebra $\mathfrak{a}$. Analogous to H^2, the sphere S^2 only posseses *one-dimensional* abelian subspaces of $\mathfrak{p}_{S^2}$ which are trivial. The geodesic submanifolds have the form:

$$\exp_p \begin{pmatrix} \mathfrak{a}_{S^2} \\ \mathfrak{p}_{\mathbb{R}} \end{pmatrix}, \quad \mathfrak{a}_{S^2} \text{ denotes abelian, hence one-dimensional subalgebra of } \mathfrak{p}_{S^2}. \tag{122}$$

13.4.2 Where are the $K = 1$ positively curved subspaces?

A plane, which is spanned by two geodesics which both live in S^2 is positively curved with $K = 1$. This is especially easy to see for the case of the Euclidean coordinate being zero, as the spanned space is just S^2. It might hence be misleading to call this a plane since it is a closed surface due to its positive curvature.

These subspaces are totally geodesic submanifolds: They are generated by the Lie triple system $\mathfrak{p}_{S^2}$ from eq. 99, i.e. they have the form:

$$\exp_p \begin{pmatrix} \mathfrak{p}_{S^2} \\ 0 \end{pmatrix}. \tag{123}$$

13.4.3 Visualization of $S^2 \times \mathbb{R}$

Recall that the sphere can be visualized via the stereographic projection in $\mathbb{R}^2$, see fig.8. The space $S^2 \times \mathbb{R}$ can hence be visualized in $\mathbb{R}^3$, because the one-dimensional Euclidean space $\mathbb{R}$ is perpendicular to the plane of the stereographic projection of S^2. The geodesics in S^2 are the geodesics within the stereographic projection, i.e. they are ellipses or in the degenerated cases lines through the origin. The geodesics in z-direction within $\mathbb{R}$ are trivial.

Spheres of the same curvature in S²xR
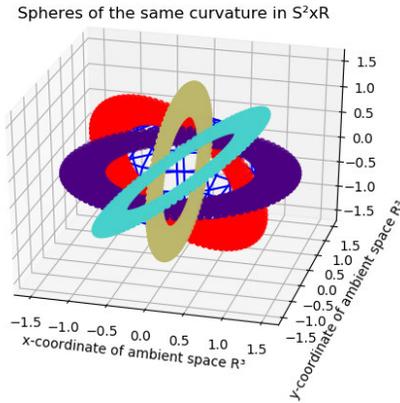
(a) These planes are spanned by vectors which live in S^2 and $\mathbb{R}$. Locally, it looks like $\mathbb{R}^2$. Note that the plane is not only infinite in positive, but also in negative r-direction, i.e. "inwards" and "outwards".

Some flat submanifolds of $K = 0$ within S²xR
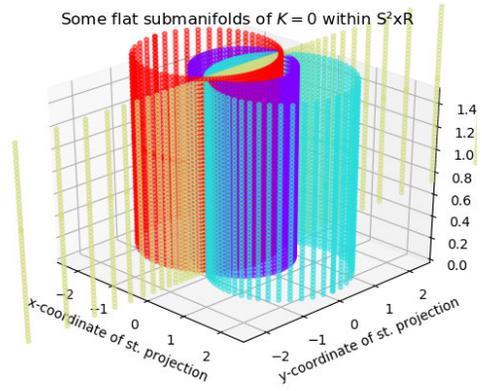
(b) Using the stereographic projection, it is apparent that the flat subspaces are cylinders. Note that the cylinders seem to be squeezed, as geodesics in the stereographic projection are ellipses.

Spheres of the same curvature in S²xR
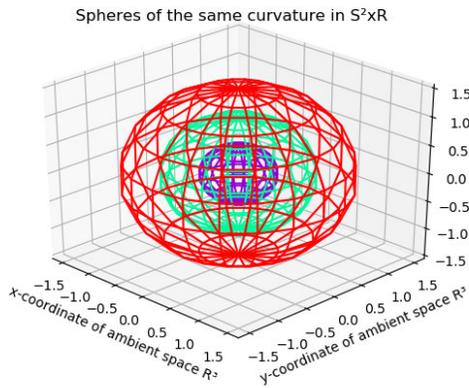
(c) Spheres of the same radius and curvature. The spherical flats are spheres of radius=1 with different value in the coordinate associated to $\mathbb{R}$. Although these spheres seem to have different size and curvature, this is only an artefact of the model. Note that the Euclidean direction is perpendicular to the surface.

Some spherical submanifolds of $K = 1$ within S²xR
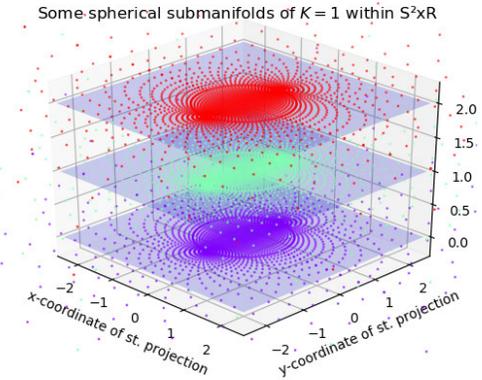
(d) The x-y-planes are planes of constant sectional curvature $K = 1$. Note that they are all of the same size and that the distance in z-direction is Euclidean.

Figure 22: Flat and spherical subspaces in $S^2 \times \mathbb{R}$. On the left as ordinary representation of the sphere within $\mathbb{R}^3$, on the right as stereographic projection.

13.4.4   Planes of non-constant sectional curvature

Analogous to the space $H^2 \times \mathbb{R}$, we now want to examine spaces spanned by vectors with mixed terms, i.e. vectors which do not lie only within one factor of the Cartesian product. One may choose the following vectors living in $T_{p=(0,0,0)}(S^2 \times \mathbb{R})$:

$$v = a \cdot \partial_x + b \cdot \partial_z \ \text{ and } \ w = \partial_y.^{24} \tag{124}$$

---

[24]Note that the vector ∂_z stands perpendicular to the sphere. The vectors ∂_x and ∂_y are of course chosen with respect to the stereographic projection and its coordinate system (x, y).

The ratio of the scalar values a and b determines the tilting of the space with respect to the $\mathbb{R}$-axis. Calculating the sectional curvature using eq. 28 and plugging eq. 99 in eq. 73 results in

$$K = \frac{4a^2}{4a^2 + b^2},\tag{125}$$

which is completely analogous to the results for $H^2 \times \mathbb{R}$ in eq. 108 and is a nice sanity check. To construct such a smooth submanifold such that at $p = (0,0,0)$ the vectors v and w from eq. 124 belong the the tangent space, one may use the map

$$G : S^2 \to S^2 \times \mathbb{R}\tag{126}$$

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x \\ y \\ g(x,y) \end{pmatrix}\tag{127}$$

and the induced vectors $v = \partial_x G(x,y)$ and $w = \partial_y G(x,y)$. The following figure is obtained by setting $g(x,y) = g_m(x) = 2 \cdot m \cdot \arctan(x)$ and hence $v = \partial_x + 2m\partial_z$ and $w = \partial_y$ with $v, w \in T_{(0,0,0)}(S^2 \times \mathbb{R})$.
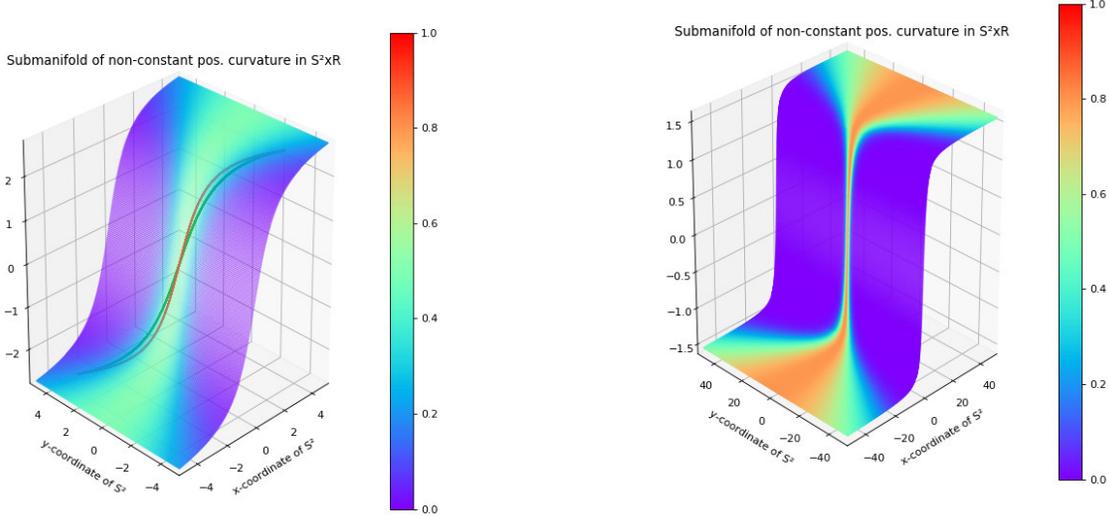
(a) Submanifold of non-constant positive sectional curvature. This submanifold is not totally geodesic: The green path is a geodesic within the submanifold, the red path is a geodesic in the full embedding space. The red path is shorter with respect to the metric.

(b) As the metric tensor tends to zero for $r \to \infty$, the Euclidean part of the spanning vectors dominates the Euclidean part for $y \to \infty$. The sectional curvature thus tends to $K = 0$ in this direction. Note that the submanifold in this figure extends to infinity in $x - y$-direction but is limited in z to $\big(2m\arctan(-\infty), 2m\arctan(\infty)\big) = \big(-m\pi, m\pi\big)$.

Figure 23: Sectional curvature of submanifolds within $S^2 \times \mathbb{R}$ indicated by coloring. On the left: Submanifold generated by $G(x,y) = (x,y,2 \cdot \arctan(x))$ from eq. 109 with $m = 1$ and vectors $v = \partial_x + 2(1+x^2)^{-1}\partial_z$ and $w = \partial_y$. At the origin $p = (0,0,0)$ these vectors become $v = \partial_x + 2 \cdot \partial_z$ and $w = \partial_y$ with $K_p = 0.5$ from eq. 125. On the right: Submanifold induced by $G(x,y) = (x,y,2 \cdot 3 \cdot \arctan(x))$ with $m = 0.5$ and vectors $v = (1,0,2 \cdot 0.5(1+x^2)^{-1})^T$ and $w = (0,1,0)^T$. At the origin these vectors become $v = \partial_x + \partial_z$ and $w = \partial_y$ with $K_p = 0.8$. One may have a look at fig. 8 to recall distortion of the stereographic projection model.

Are these subspaces totally geodesic?

The submanifolds which are spanned by these vectors are not totally geodesic. Define

$$\mathfrak{a} = \left\{ \begin{pmatrix} \alpha \cdot a \cdot p_1 + \beta \cdot p_2 \\ \alpha \cdot b \cdot p_3 \end{pmatrix} \right\}, \quad p_1, p_2, p_3 \text{ forming a basis of } \begin{pmatrix} \mathfrak{p}_{S^2} \\ \mathfrak{p}_{\mathbb{R}} \end{pmatrix} \tag{128}$$

as two-dimensional linear span (with Greek parameters) of the respective transvection spaces with ratios determined by Latin letters. Analogous to the computation for $H^2 \times \mathbb{R}$ in eq. 114, the space $\mathfrak{p}_{\mathbb{R}}$ is abelian and hence the commutator in this component vanishes. Again, the only totally geodesic submanifolds are flat cylinder and the space S^2 itself.

13.4.5 Planes of constant positive sectional curvature

One can find planes of constant sectional curvature analogous to the procedure in section 13.3.5 above: In cylindrical coordinates the metric tensor becomes:

$$g_{\mu\nu} = \begin{pmatrix} \frac{4}{(1+r^2)^2} & 0 & 0 \\ 0 & \frac{4}{(1+r^2)^2} \cdot r^2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{in coordinates } (r, \phi, z). \tag{129}$$

The map G from eq. 126 becomes:

$$\begin{pmatrix} r \\ \phi \end{pmatrix} \mapsto \begin{pmatrix} r \\ \phi \\ g(r, \phi) \end{pmatrix} \quad \text{with resulting vectors } v = (1, 0, \partial_r g)^T, w = (0, \frac{1}{r}, \partial_\phi g)^T. \tag{130}$$

The solution is again in particular easy, if one chooses to set $g(r, \phi) = g(r)$. The sectional curvature on every point $q = (r, \phi, z)$ is:

$$K(r, \phi, z) = \frac{1}{1 + \frac{(1+r^2)^2}{4} \left(\partial_r g(r)\right)^2}. \tag{131}$$

In order to keep K constant, one finds the solutions:

$$g(r) = \pm 2 \cdot m \cdot \arctan(r) \text{ with } m \text{ as free parameter to determine } K \text{ via } K = -\frac{1}{1+m^2}. \tag{132}$$

The discussion of the space being isomorphic to $S^2_{\sqrt{K}}$ is analogue to the discussion for $H^2 \times \mathbb{R}$ above.

Visualization of the embedding of constant sectional curvature in $S^2 \times \mathbb{R}$
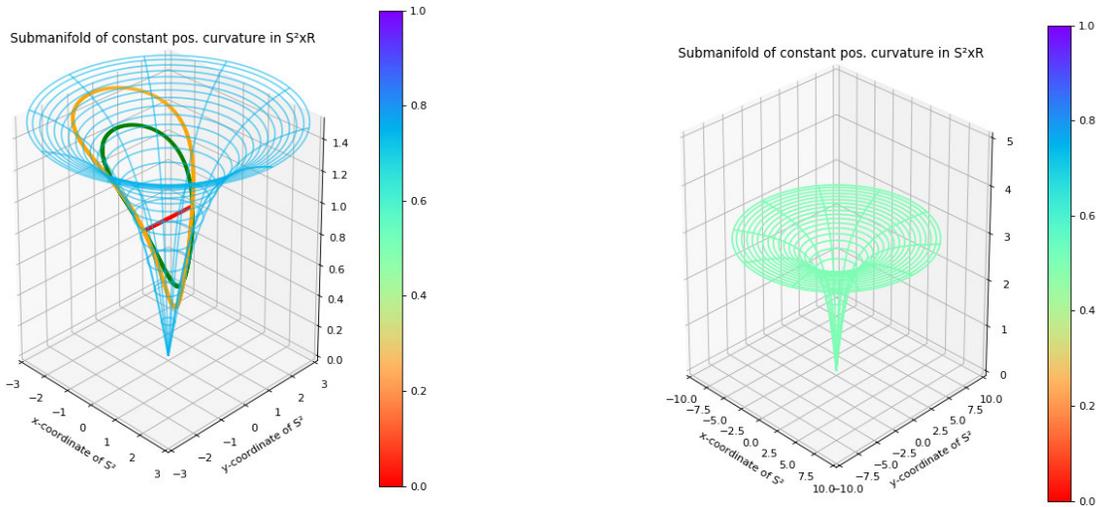
(a) Subspace of constant curvature $K = 0.75$ with green and orange geodesics on that subspace. The red geodesic within the embedding space shows that these subspaces are note totally geodesic. Note that geodesics intersect each other twice.

(b) Note the different range of the axes: This funnel is "steeper" than the funnel on the left. The tangent vectors have a greater part in the Euclidean direction, i.e. this funnel corresponds to the sphere of *bigger* radius.
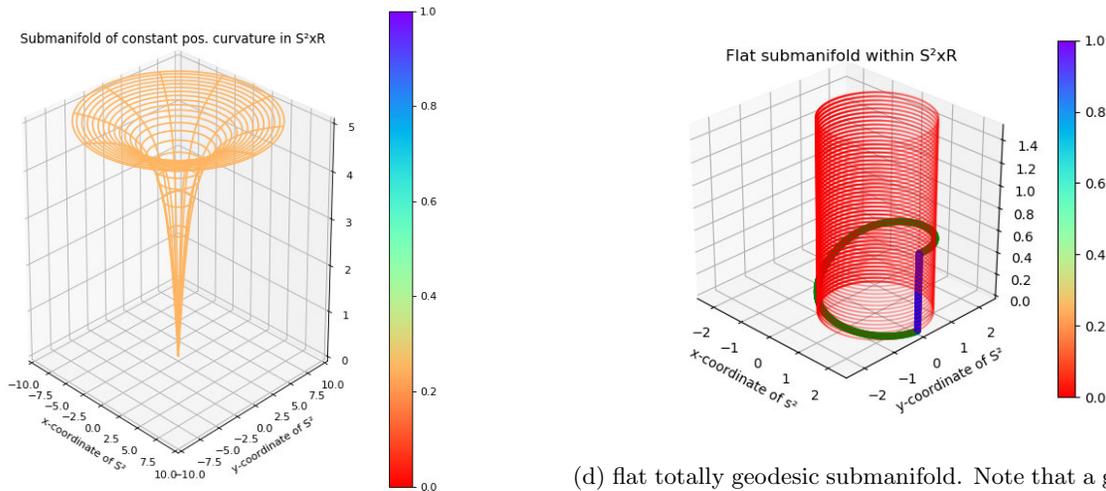
(c) The sectional curvature of this funnel is close to zero. Note that the slim part of the funnel already looks like a cylinder.

(d) flat totally geodesic submanifold. Note that a geodesic is only locally the shortest route. Globally, there can be several geodesics as indicated by the green and blue paths. Note that for two points on a sphere there are at least two geodesic paths connecting them.

Figure 24: Subspaces of constant positive sectional curvature. These can be embedded in $\mathbb{R}^3$ as $\mathbb{R}^2 \times \mathbb{R}$ because the stereographic projection of the sphere spans $\mathbb{R}^2$ and the Euclidean factor spans $\mathbb{R}$ perpendicular to it.
The funnel surfaces possess constant positive curvature, and are hence isomorphic to a sphere of radius ≥ 1 except for the cone point: These funnel-shaped subspaces are not smooth manifolds due to the cone point $(0, 0, 0)$.

13.5   The symmetric space $H^2 \times H^2$

$H^2 \times H^2$ shares some of its properties with $H^2 \times \mathbb{R}$, because $H^2 \times \mathbb{R} \subset H^2 \times H^2$. It hence also possesses 2-dimensional flat subspaces and hyperbolic subspaces at the same time. This again shows that $H^2 \times H^2$ is not the same as H^4, since H^4 does not possess 2-dimensional flat subspaces, more precisely it is a manifold of rank 1.

The same argument holds for the torus $T = S^1 \times S^1$ which is of rank 2 and thus not equivalent to the sphere S^2 which is of rank 1.

Let us examine the properties of $H^2 \times H^2$:

The first two and last two components behave as pairs of two-dimensional hyperbolic space. The Cartesian product however is the reason, why this space also has infinitely many two-dimensional flat Euclidean subspaces as does $H^2 \times \mathbb{R}$. Additionally, $H^2 \times H^2$ space has another property which arises from the fact that it is the product of two *distinct* hyperbolic spaces: These two hyperbolic spaces are independent of each other which permits an embedding of two different hierarchies at the same time. The two factors will be enumerated with index i as $(_iH^2)$.

13.5.1   Where are the flat subspaces?

As pointed out before, the spaces $(_iH^2)$ are independent of each other due to the Cartesian product. Let us remind ourselves that the Euclidean space $\mathbb{R}^2$ is the Cartesian product $\mathbb{R} \times \mathbb{R}$. Therefore we need a Cartesian product of two one-dimensional flat spaces $\mathbb{R}$. These one-dimensional flat spaces are the geodesics within H^2. The plane which is spanned by two vectors from different hyperbolic spaces is hence a flat subspace.

More formally, the one-dimensional subspaces of the respective transvection spaces $_i\mathfrak{p}$ are trivially abelian:

$$[\alpha \cdot p, \beta \cdot p] = (\alpha \cdot \beta)[p, p] = 0, \;\; p \in \,_i\mathfrak{p}. \tag{133}$$

The vectors

$$v = (a_1 \cdot \partial_{x^1} + a_2 \cdot \partial_{y^1}) \;\; \text{and} \;\; w = (b_1 \cdot \partial_{x^2} + b_2 \cdot \partial_{y^2}), \;\; a_i, b_i, c_i \in \mathbb{R} \tag{134}$$

at the base point $p = (0, 0, 0, 0)$ in the setting of a product of Poincaré disks span a 2-dimensional submanifold. The coefficients determine the specific geodesics in each hyperbolic factor. The generated submanifold is flat since the corresponding transvections in the Lie algebra picture are abelian. Writing the space spanned by these vectors in rows where each row corresponds to a H^2-factor of the Cartesian product, one obtains in Lie algebra formulation

$$\mathfrak{a} := \left\{ \begin{pmatrix} \alpha \cdot (a_1 \cdot \,_1p_x + a_2 \cdot \,_1p_y) \\ \beta \cdot (b_1 \cdot \,_2p_x + b_2 \cdot \,_2p_y) \end{pmatrix} \right\}, \tag{135}$$

with $_jp_{x,y}$ corresponding to transvection in x^j, y^j direction within the factor $(_jH^2)$.

The Lie bracket of this space vanishes, because both components are abelian subalgebras of their respective $_i\mathfrak{p}$, since they are one-dimensional.

Pictorially, the entries in the components in eq. 135 each correspond to a transvection in direction of the sum of the constituent transvections. These transvections generate geodesics in both spaces $(_iH^2)$ independently. Together, they span $\mathbb{R}^2$.

There are no three-dimensional flat submanifolds since rank$(H^2 \times H^2) = 2$. A combination of two one-dimensional subspaces from the same $\mathfrak{p}_{(_iH^2)}$ is not an abelian Lie triple system and hence not flat. The span of two geodesics from the same hyperbolic space does not yield a 2 dimensional flat subspace, simply because $H^2 \not\cong \mathbb{R} \times \mathbb{R}$.

13.5.2   Where are the $K = -1$ negatively curved subspaces?

Subspaces which lie in one hyperbolic product only are trivially two-dimensional subspaces of sectional curvature $K = -1$. They are totally geodesic subspaces generated by:

$$\exp \begin{pmatrix} \mathfrak{p}_{H^2} \\ 0 \end{pmatrix}_p \;\; \text{and} \;\; \exp \begin{pmatrix} 0 \\ \mathfrak{p}_{H^2} \end{pmatrix}_p . \tag{136}$$

13.5.3 Planes of more interesting curvature

Analogy to $H^2 \times \mathbb{R}$

Since $H^2 \times \mathbb{R} \subset H^2 \times H^2$, the two-dimensional negatively curved subspaces of $H^2 \times \mathbb{R}$ can be found within $H^2 \times H^2$. The discussion of the spaces of the form

$$\left\{ \begin{pmatrix} \alpha \cdot a \cdot {}_1 p_x + \beta \cdot b \cdot {}_1 p_y \\ \alpha \cdot {}_1 p_x \end{pmatrix} \right\}, {}_j p_{x,y} \text{ corresponding to transvection in } x^j, y^j \text{ direction in the space } ({}_j H^2) \quad (137)$$

is identical to that of paragraph 13.3.5, since the one-dimensional transvection in ${}_i\mathfrak{p}$ of $({}_2 H^2)$ corresponds to a transvection in $\mathbb{R}$.

Other planes within $H^2 \times H^2$

Looking at eq. 135 and eq. 137, one notices that not all two-dimensional subspaces are exhausted. Consider the vectors:

$$v = a_1 \partial_{x^1} + a_2 \partial_{x^2} \text{ and } w = b_1 \partial_{x^1} + c_1 \partial_{y^1} + b_2 \partial_{x^2} + c_2 \partial_{y^2}, \ \ a_i, b_i, c_i \in \mathbb{R} \quad (138)$$

which is the most general form how two vectors can lie in $H^2 \times H^2$, because on can choose the coordinates of both factors in such a way that v is pointing in x^1 and x^2-direction.
The space generated by these vectors is a totally geodesic submanifold. To prove that, consider that space in the picture of transvection spaces:

$$\left\{ \begin{pmatrix} \alpha \cdot a_1 \cdot {}_1 p_x + \beta\big(b_1 \cdot {}_1 p_x + c_1 \cdot {}_1 p_y\big) \\ \alpha \cdot a_2 \cdot {}_2 p_x + \beta\big(b_2 \cdot {}_2 p_x + c_2 \cdot {}_2 p_y\big) \end{pmatrix} \right\},$$

$${}_j p_{x,y} \text{ corresponding to transvection in } x^j, y^j \text{ direction in the space } ({}_j H^2). \quad (139)$$

To determine whether the space induced by these vectors is totally geodesic, we investigate if the Lie triple product is closed under the subspace of $\mathfrak{p}$ induced by the vectors. For all parameters $\alpha_i, \beta_i, a_i, b_i, c_i$ there must exist real parameters α_4, β_4 such that following holds:

$$\left[\left[\begin{pmatrix} \alpha_1 a_1 \, {}_1 p_x + \beta_1\big(b_1 \, {}_1 p_x + c_1 \, {}_1 p_y\big) \\ \alpha_1 a_2 \, {}_2 p_x + \beta_1\big(b_2 \, {}_2 p_x + c_2 \, {}_2 p_y\big) \end{pmatrix}, \begin{pmatrix} \alpha_2 a_1 \, {}_1 p_x + \beta_2\big(b_1 \, {}_1 p_x + c_1 \, {}_1 p_y\big) \\ \alpha_2 a_2 \, {}_2 p_x + \beta_2\big(b_2 \, {}_2 p_x + c_2 \, {}_2 p_y\big) \end{pmatrix} \right], \begin{pmatrix} \alpha_3 a_1 \, {}_1 p_x + \beta_3\big(b_1 \, {}_1 p_x + c_1 \, {}_1 p_y\big) \\ \alpha_3 a_2 \, {}_2 p_x + \beta_3\big(b_2 \, {}_2 p_x + c_2 \, {}_2 p_y\big) \end{pmatrix} \right]$$

$$\overset{!}{=} \begin{pmatrix} \alpha_4 a_1 \, {}_1 p_x + \beta_4\big(b_1 \, {}_1 p_x + c_1 \, {}_1 p_y\big) \\ \alpha_4 a_2 \, {}_2 p_x + \beta_4\big(b_2 \, {}_2 p_x + c_2 \, {}_2 p_y\big) \end{pmatrix}.$$

$$(140)$$

Direct calculation yields for the left side:

$$\begin{pmatrix} c_1 a_1 (\alpha_1 \beta_2 - \beta_1 \alpha_2)\Big((\alpha_3 a_1 + \beta_3 b_1)\big[[{}_1 p_x, {}_1 p_y], {}_1 p_x\big] + \beta_3 c_1 \big[[{}_1 p_x, {}_1 p_y], {}_1 p_y\big]\Big) \\ c_2 a_1 (\alpha_1 \beta_2 - \beta_1 \alpha_2)\Big((\alpha_3 a_2 + \beta_3 b_2)\big[[{}_2 p_x, {}_2 p_y], {}_2 p_x\big] + \beta_3 c_2 \big[[{}_2 p_x, {}_2 p_y], {}_2 p_y\big]\Big) \end{pmatrix}. \quad (141)$$

Putting both sides together one arrives at:

$$\beta_4 = a_1(\alpha_3 a_1 + \beta_3 b_1) \overset{!}{=} a_2(\alpha_3 a_2 + \beta_3 b_2) \quad (142)$$

$$\alpha_4 = \big(\beta_3 c_1^2 - \alpha_3 a_1 b_1 + \beta_3 b_1^2\big) \overset{!}{=} \big(\beta_3 c_2^2 - \alpha_3 a_2 b_2 + \beta_3 b_2^2\big) \quad (143)$$

which ultimately leads to $a_1 = a_2, b_1 = b_2, c_1 = \pm c_2$, [25] since the coordinate system has been chosen to align the two x-axes and since can eq. 140 must hold for all values α_i, β_i.
This space is hence only a totally geodesic submanifold, if it lies diagonally within $H^2 \times H^2$. The sectional

---

[25]except for trivial cases without coupling between the two factors H^2 which reduce to the spaces H^2 or $\mathbb{R}^2$.

curvature of this 2-dimensional space will be constant due to $\nabla R = 0$. With the decomposition with respect to the two underlying hyperbolic spaces follows:

$$K = \frac{g_1(R_1(v^1, w^1)w^1, v^1) + g_2(R_2(v^2, w^2)w^2, v^2)}{(g_1(v^1, v^1) + g_2(v^2, v^2))(g_1(w^1, w^1) + g_2(w^2, w^2)) - \left(g_1(v^1, w^1) + g_2(v^2, w^2)\right)^2}$$
$$= \frac{-a_1^2 c_1^2 - a_2^2 c_2^2}{a_1^2 c_1^2 + a_1^2 c_2^2 + a_2^2 c_1^2 + a_2^2 c_2^2} = \frac{-2a^2 c^2}{4a^2 c^2} = -\frac{1}{2}. \tag{144}$$

In those cases in which the vectors each only lie in one hyperbolic factor only (such as $a_1 = b_2 = c_2 = 0$) the subspace is flat as in section 13.3.1. For those cases in which the vectors both lie in the same hyperbolic factor (such as $a_2 = b_2 = c_2 = 0$), the sectional curvature is -1 as in section 13.3.2. As one has seen in section 13.3, the sectional curvature of the whole space confined to the interval $[-1, 0]$. In comparison to $H^2 \times \mathbb{R}$, $H^2_{\sqrt{2}}$ is a totally geodesic submanifold.

13.6 More general Cartesian products of "simple" symmetric spaces

One can form a symmetric space as Cartesian product from arbitrarily many symmetric spaces. So far, we have only investigated "simple" symmetric spaces of at most two constituents. However, Cartesian products of many "standard" symmetric spaces as $\mathbb{R}^n, H^2, S^n$ are interesting, because computations decompose and the space is hence easy to deal with. Helpful information about such a product space are its rank and its totally geodesic submanifolds.

Rank of Cartesian products

For these kinds of "simple" product spaces, the results of this chapter allow to read off rank and the maximal dimension of negatively and positively curved subspaces: The maximal dimension of negatively curved subspaces is the maximal dimension of a negatively curved subspace in any factor of the product, e.g. for $\mathbb{R}^2 \times H^3 \times H^4$ it is four.
Completely analogous to this, the maximal dimension of positively curved subspaces can be determined. For $\mathbb{R}^2 \times S^3 \times S^4$ it is also four.
The rank of a symmetric space is a little bit more complicated to determine, since every factor adds at least one dimension since every factor has at rank 1. This is because a one-dimensional subalgebra $\mathfrak{p}_i \subset \mathfrak{g}_i$ of the respective Lie algebras of every constituent is trivially abelian. As the rank of a Cartesian product is the sum of the ranks of the constituents, the rank of $H^2 \times S^3 \times \mathbb{R}^4$ is $1 + 1 + 4 = 6$.

Totally geodesic submanifolds of Cartesian products

It was shown in the sections about $H^2 \times \mathbb{R}$ and $S^2 \times \mathbb{R}$ that there are no totally geodesic submanifolds except for the trivial ones. However, the example of $H^2 \times H^2$ showed that a Cartesian product can have interesting totally geodesic submanifolds. The following will investigate that a bit further:

13.6.1 $S^2 \times S^2$

The space $S^2 \times S^2$ is very similar in nature to $H^2 \times H^2$ which one can read off from their similar metric tensors eq. 31 and eq. 35.[26] Carrying out the computations from above, one finds: The totally geodesic submanifolds are the spaces S^1, $S^1 \times S^1$, $S^2 \times S^1$ and S^2. Note that the torus $S^1 \times S^1$ is a flat 2-dimensional manifold. There are additionally infinitely many spaces $S^2_{R=\sqrt{2}}$ with sectional curvature $K = 0.5$ lying diagonally within $S^2 \times S^2$.

13.6.2 $H^2_{r_1} \times H^2_{r_2}$

This is the Cartesian product of hyperbolic spaces of curvatures different from $K = -1$. In fact, the sectional curvature of H^2_r is $-1/r^2$. The coset formulation of H^2_r is $SO(2,1)/SO(2)$. Nevertheless, the base point in the embedding space of the hyperboloid is not $(0, 0, 1)$ but $(0, 0, r)$. This means that the Lie algebra of the isometry

---

[26]In fact, S^2 is *dual* to H^2 with respect to the commutation relations on their respective Lie algebras of the isometry group.

group is the same as for $H^2_{r=1}$. The metric tensor of H^2_r in the Poincaré model is just scaled with r^2.
The totally geodesic submanifolds of $H^2_{r_1} \times H^2_{r_2}$ are hence analogous to the case $H^2 \times H^2$ with different curvatures:
The interesting 2-dimensional submanifolds from eq. 139 have the curvature:

$$K = \frac{r_1^2 g_1(R_1(v^1, w^1)w^1, v^1) + r_2^2 g_2(R_2(v^2, w^2)w^2, v^2)}{\left(r_1^2 g_1(v^1, v^1) + r_2^2 g_2(v^2, v^2)\right)\left(r_1^2 g_1(w^1, w^1) + r_2^2 g_2(w^2, w^2)\right) - \left(r_1^2 g_1(v^1, w^1) + r_2^2 g_2(v^2, w^2)\right)^2}$$
$$= -\frac{r_1^4 + r_2^4}{\left(r_1^2 + r_1^2\right)^2}. \tag{145}$$

This subspace is still a totally geodesic submanifold, as the Lie algebra structures of $H^2_{r_1}$ and $H^2_{r_2}$ fit together, as they are both hyperbolic spaces. This does not hold for the following case:

13.6.3 $\quad H^2 \times S^2$

This product space has of course the following trivial totally geodesic submanifolds:

$$H^r \times S^s, \ \ r, s \leq 2. \tag{146}$$

This also includes the flat cylinder $\mathbb{R} \times S^1$ (because $\mathbb{R} \subset H^2$ and $S^1 \subset S^2$). However, one might think that in analogy to the case $H^2 \times H^2$ examined in detail above, there is another 2-dimensional totally geodesic subspace which lies diagonally within $H^2 \times S^2$. This is not the case since the Lie algebras $_{H^2}\mathfrak{p}$ and $_{S^2}\mathfrak{p}$ are not compatible with each other: Solving the Lie triple product explicitly in eq. 141 is only possible if the Lie algebras of both spaces $(_iH^2)$ are compatible in the sense that their commutation relations match.
Since the Lie algebra of S^2 generates infinitesimal rotations in 3-dimensional Euclidean space, but the Lie algebra of H^2 generates infinitesimal rotations in 3-dimensional Minkowski space, their commutation relations are different. This ultimately leads to equations analogous to eq. 142 which cannot be fulfilled:

$$a_1(\alpha_3 a_1 + \beta_3 b_1) \overset{!}{=} -a_2(\alpha_3 a_2 + \beta_3 b_2) \text{ or} \tag{147}$$

$$\left(\beta_3 c_1^2 - \alpha_3 a_1 b_1 + \beta_3 b_1^2\right) \overset{!}{=} -\left(\beta_3 c_2^2 - \alpha_3 a_2 b_2 + \beta_3 b_2^2\right). \tag{148}$$

13.6.4 $\quad (H^2)^n$

The flat submanifolds are $\mathbb{R}^n$ and hence $\mathbb{R}^m$, $m < n$ which are trivially totally geodesic. Other trivially totally geodesic submanifolds are: $(\bigtimes_{i=1}^r H^2) \times \mathbb{R}^s$, $r + s \leq n$. Analogous to eq. 140 for $H^2 \times H^2$ there are 2-dimensional subspaces of constant curvature different from -1:
Consider the vectors

$$v = (a_1 \cdot \partial_{x_1} + ... + a_n \cdot \partial_{x_n}) \text{ and } w = (b_1 \cdot \partial_{x_1} + c_1 \cdot \partial_{y_1} + ... + b_n \cdot \partial_{x_n} + c_n \cdot \partial_{y_n}), \ \ a_i, b_i, c_i \in \mathbb{R} \tag{149}$$

which is the most general setting of vectors in $(H^2)^n$, as one can orient each space H^2 in such a way that the corresponding v-components each point in respective x-direction. This can also be written in a more comprehensive form where each line corresponds to a H^2-factor in the Cartesian product:

$$v = \begin{pmatrix} a_1 \partial_{x_1} \\ \vdots \\ a_n \partial_{x_n} \end{pmatrix}, w = \begin{pmatrix} b_1 \partial_{x_1} + c_1 \partial_{y_1} \\ \vdots \\ b_n \partial_{x_n} + c_1 \partial_{y_n} \end{pmatrix}. \tag{150}$$

To be a totally geodesic submanifold, the vectors need to form a Lie triple system. Analogous to the computation eq. 140, one arrives at:

$$a_i = a_j, b_i = b_j, c_i = c_j \ \forall i, j \ ^{27}. \tag{151}$$

 As the submanifold is totally geodesic, its sectional curvature must be constant since the sectional curvature is a scalar. This allows to compute the sectional curvature at any point:

$$K = \frac{-\sum_{i=1}^n a_i^2 c_i^2}{\left(\sum_{i=1}^n a_i^2\right)\left(\sum_{i=1}^n b_i^2 + c_i^2\right) - \left(\sum_{i=1}^n a_i^2 b_i^2\right)^2} = -\frac{n}{n^2} = -\frac{1}{n}. \tag{152}$$

---

[27] The cases with some coefficients being zero are mentioned below as $(H^2)^r$ is contained within $(H^2)^n$, $r < n$.

As the spaces $(H^2)^r$ $r \leq n$ are also contained within $(H^2)^n$, there are each infinitely many 2-dimensional totally geodesic submanifolds of constant sectional curvature $-1, -\frac{1}{2}, -\frac{1}{3}, -\frac{1}{4}, ..., -\frac{1}{n}$. Hence the spaces

$$H^2_{R=\sqrt{1}}, H^2_{R=\sqrt{2}}, H^2_{R=\sqrt{3}}, ..., H^2_{R=\sqrt{n}} \tag{153}$$

are totally geodesic submanifolds. Because the spaces $H^2_{R=\sqrt{2}}$ and $H^2_{R=\sqrt{3}}$ etc. do not stand perpendicular on each other, these submanifolds do not generate further interesting totally geodesic submanifolds as in section 13.6.2.

Since $H^3, H^4, ...$ etc. are not contained in $(H^2)^n$, the mentioned spaces exhaust all totally geodesic submanifolds.

13.6.5 $(S^2)^n$

For $(S^2)^n$ the results are analogous to the case of $(H^2)^n$: Aside from the trivial totally geodesic submanifolds, there are each infinitely many 2-dimensional totally geodesic submanifolds of constant sectional curvature $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, ..., \frac{1}{n}$ which corresponds to spheres of radius $1, \sqrt{2}, \sqrt{3}, ..., \sqrt{n}$.

13.6.6 $S^2_{r_1} \times S^2_{r_2}$

The discussion can be taken almost verbatim from the case $H^2_{r_1} \times H^2_{r_2}$ from section 13.6.2. The interesting 2-dimensional totally geodesic submanifolds have the sectional curvature:

$$K = \frac{r_1^4 + r_2^4}{\left(r_1^2 + r_1^2\right)^2}. \tag{154}$$

13.6.7 $H^m \times \mathbb{R}^n$ and $S^m \times \mathbb{R}^n$

These products do not possess any totally geodesic submanifolds except for the trivial ones which are respectively: $H^r \times \mathbb{R}^s$, $r \leq m$, $s \leq n$ and $S^r \times \mathbb{R}^s$, $r \leq m$, $s \leq n$ and of course $\mathbb{R}^{n+1}$, as these product spaces possess rank $n+1$.

14 The Siegel space

Introduction

The spaces which have been investigated before, are Cartesian products of spaces of constant sectional curvature. Investigating their properties thus reduces to investigating the properties of every single factor of the Cartesian product. The Siegel space is non-decomposable, i.e. it is not a Cartesian product of simpler spaces. For the Siegel space, the introduction to the theory of symmetric spaces becomes powerful.

The Siegel space (also called Siegel upper half plane or Siegel upper half-space) has been introduced and thoroughly investigated by CARL LUDWIG SIEGEL [Siegel, 1943]. It is a generalization of hyperbolic space in the sense that the models and metric are similar in structure to those of hyperbolic space. Whilst points in hyperbolic space are represented by vectors, points in the Siegel upper half-space are represented by matrices. The implications of the matrix product and its non-commutativity change a lot of the properties of the Siegel space compared to "regular" hyperbolic space. To facilitate getting a feel for the Siegel space, the box below provides a more detailed insight into the coset formulation of hyperbolic space than the introductory section 3.3 before.

The coset formulation of hyperbolic space

The following will elaborate on the Lie group quotient picture of 2-dimensional hyperbolic space in the introduced models. It later becomes apparent that 2-dimensional hyperbolic space is a natural subspace of the Siegel space.

- Hyperbolic space in the hyperboloid model can be written as $SO(2,1,\mathbb{R})/SO(2,\mathbb{R})$. The action isometry and isotropy groups is a regular matrix-vector multiplication in $\mathbb{R}^3$.

- The coset formulation of hyperbolic space as Poincaré disk model in complex coordinates is $SU(1,1,\mathbb{C})/SU(1,\mathbb{C})$ with

$$SU(1,1,\mathbb{C}) = \left\{ \begin{pmatrix} u & v \\ v^* & u^* \end{pmatrix} \middle| uu^* - vv^* = 1, \ u,v \in \mathbb{C} \right\}, \ SU(1,\mathbb{C}) = \left\{ \begin{pmatrix} e^{i\frac{\phi}{2}} & 0 \\ 0 & e^{-i\frac{\phi}{2}} \end{pmatrix} \middle| \phi \in \mathbb{R} \right\}.$$
(155)

The group $SU(1,1,\mathbb{C})$ and hence $SU(1,\mathbb{C})$ act as Möbius transformations on the disk:

$$\theta\left(\begin{pmatrix} u & v \\ v^* & u^* \end{pmatrix} \right) z = \frac{u \cdot z + v}{v^* \cdot z + u^*} \ \text{ as complex division.}$$
(156)

- The coset formulation in the upper half-space model in complex coordinates is $SL(2,\mathbb{R})/SO(2,\mathbb{R})$ with

$$SL(2,\mathbb{R}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \middle| ab - cd = 1 \right\} \ \text{and} \ SO(2,\mathbb{R}) = \left\{ \begin{pmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{pmatrix} \middle| \phi \in \mathbb{R} \right\}.$$
(157)

The group $SL(2,\mathbb{R})$ and hence $SO(2,\mathbb{R})$ also act as Möbius transformations on the complex upper half-space:

$$\theta\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} \right) z = \frac{a \cdot z + b}{c \cdot z + d} \ \text{ as complex division.}$$
(158)

Note that the actions on the Poincaré disk and the action on the upper half-space are conjugate via the Cayley transform i from eq. 37. Explicitly:

$$(\mathsf{i}^{-1} \circ \theta(g) \circ \mathsf{i}) = \theta(\tilde{g}) \ \ g \in SU(1,1,\mathbb{C}), \ \ \tilde{g} \in SL(2,\mathbb{R}).$$
(159)

In fact, $SU(1,1,\mathbb{C})$ and $SL(2,\mathbb{R})$ are isomorphic as well as $SU(1,\mathbb{C})$ and $SO(2,\mathbb{R})$. This again justifies why we focus on the model which fits best to convenience. For visualization purposes one often uses the disk, but for computational purposes we will use the upper half-space model.

14.1 The upper half-space model of the Siegel space

The Siegel space is a generalization of hyperbolic space. In particular, the space SH^n and H^n coincide for $n = 1$. For higher dimensions this -of course- does not hold. Analogue to hyperbolic space, there are several equivalent models of the Siegel space. The following model of the Siegel space is a generalization of the upper half-space. It can be defined as:

$$SH^n = \{Z \in Sym(n,\mathbb{C})|\Im(Z) > 0\}, \ \text{with} \ Sym(n,\mathbb{C}) \ \text{symmetric} \ n \times n\text{-matrices with entries in } \mathbb{C}.^{28} \quad (160)$$

The similarity to the upper half-space model of hyperbolic space is evident. The restriction to matrices with positive definite imaginary part can be understood when writing an element $P \in SH^n$ as $P = X + iY$. The matrix Y is symmetric and has real entries. It is hence diagonalizable with positive eigenvalues by definition of SH^n.

---

[28] The capital letter Z stands for a matrix, whereas a small letter z stands for a complex number.

Finding a coset formulation of SH^n in terms of a Lie group of isometries and a corresponding subgroup induces a Riemannian structure on SH^n. In fact, the coset formulation of the Siegel space reads:

$$SH^n = Sp(2n, \mathbb{R})/SpO(2n, \mathbb{R}). \tag{161}$$

The Symplectic group $Sp(2n, \mathbb{R})$

The group $Sp(2n, \mathbb{R})$ is the group of matrices which leaves the symplectic scalar product invariant. To get an intuition for that, recall that the group $O(n, \mathbb{R})$ leaves the Euclidean scalar product $\langle \cdot, \cdot \rangle_E$ invariant:

$$\langle Mv, Mw \rangle_E = (Mv)^T \cdot Id \cdot Mw = v^T M^T \cdot Id \cdot Mw = v^T \cdot Id \cdot w = \langle v, w \rangle_E, \ \ v, w \in \mathbb{R}^n, \ \ M \in O(n, \mathbb{R}). \tag{162}$$

The defining property of a matrix M to be orthogonal is hence $M^T \cdot Id \cdot M = Id$.

We now define the symplectic matrices as those matrices leaving the symplectic scalar product $\langle \cdot, \cdot \rangle_{Sp}$ invariant:

$$\langle Mv, Mw \rangle_{Sp} = (Mv)^T \cdot J \cdot Mw = v^T M^T \cdot J \cdot Mw = v^T \cdot J \cdot w = \langle v, w \rangle_{Sp}, \ \ v, w \in \mathbb{R}^{2n}, \ \ M \in Sp(2n, \mathbb{R}). \tag{163}$$

The defining property of a matrix M to be symplectic is hence $M^T \cdot J \cdot M = J$.
The symplectic form J is a $(2n \times 2n)$-matrix:

$$J = \begin{pmatrix} 0 & Id_n \\ -Id_n & 0 \end{pmatrix} \tag{164}$$

in block matrix notation. From this definition it becomes apparent, why the symplectic group only exists in even dimensions. By investigating the defining property of a matrix to be symplectic, we arrive at the block form

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \ \ A^T D - C^T B = Id_n, \ \ A^T C \ \text{and} \ B^T D \ \text{symmetric matrices.} \tag{165}$$

It can be inferred that a symplectic matrix has determinant $+1$ [Freitas, 1999, Prop. 2.2.3].

Analogue to the action of $SL(2n, \mathbb{R}) \curvearrowright H^2$, the symplectic group $Sp(2n, \mathbb{R})$ acts on SH^n as linear fractional transformations via the map $\theta : Sp(2n, \mathbb{R}) \to \mathcal{M}aps(SH^n; SH^n)$:

$$\theta \Big(\begin{pmatrix} A & B \\ C & D \end{pmatrix} \Big) Z = (AZ + B)(CZ + D)^{-1} \in SH^n. \tag{166}$$

Compare this with the action $SL(2, \mathbb{R}) \curvearrowright H^2$ of eq. 158. Note that $(CZ + D)^{-1}$ is the matrix analogue to the complex division $\frac{1}{cz+d}$. Also, note that the matrix acting on SH^n is a $(2n \times 2n)$-matrix with blocks A, B, C, D of size $n \times n$. The matrix Z is a complex $(n \times n)$-matrix. It can be shown that the resulting matrix is an element of SH^n. The action $Sp(2n, \mathbb{R}) \curvearrowright SH^n$ is hence well defined.

14.2 The ball model of the Siegel space

The ball model of the Siegel space is again similar to the Poincaré ball model of hyperbolic space. Explicitly:

$$SB^n = \{ Z \in Sym(n, \mathbb{C}) | Id_n - Z\bar{Z} > 0 \}. \tag{167}$$

This model is equivalent to the upper half-space model, [Siegel, 1943, II, 4] since the generalization of the Cayley transform as in eq. 37 maps one model to the other:

$$\mathfrak{i} : \text{Ball model} \to \text{half-space model}, \ Z \mapsto i(Z + Id_n)(-Z + Id_n)^{-1} \tag{168}$$

$$\mathfrak{i}^{-1} : \text{half-space model} \to \text{Ball model}, \ Z \mapsto (Z - iId_n)(Z + iId_n)^{-1}. \tag{169}$$

These maps can be expressed as generalized Möbius transformations associated with the matrices

$$\begin{pmatrix} Id_n & -Id_n \\ Id_n & iId_n \end{pmatrix} \text{ and } \frac{1}{2i} \begin{pmatrix} iId_n & iId_n \\ -1 & 1 \end{pmatrix}.^{29} \tag{170}$$

Again, the isometry group of the ball model is conjugated to the isometry group of the half-space model and can be written symbolically as $i^{-1} \circ Sp(2n, \mathbb{R}) \circ i$ which is the set of matrices:

$$i^{-1} \circ Sp(2n, \mathbb{R}) \circ i = \begin{pmatrix} U & B \\ \overline{B} & \overline{U} \end{pmatrix}, \ UU^\dagger - BB^\dagger = Id_n, \ UB^T = BU^T, \ U, B \text{ complex.} \tag{171}$$

This justifies working with the model which is most convenient for the purpose.

14.3 The symmetry on SH^n

The symmetry at $P = 0$ in the ball model is easy to see: It just maps Z to $-Z$. In matrix form acting as a linear fractional it reads

$$I_0(SB^n) \ni g_{s_{P=0}} = \begin{pmatrix} -iId_n & 0 \\ 0 & iId_n \end{pmatrix} \text{ in the ball model.} \tag{172}$$

This map can be lifted to the half-space model via eq. 168. This yields at $P = iId_n$ the symmetry given by $s_P : SH^n \to SH^n, \ Z \mapsto -Z^{-1}$ [30]. It also acts as linear fractional transformation and can be written in matrix form as

$$I_0(SH^n) \ni g_{s_{P=iId}} = \begin{pmatrix} 0 & -Id_n \\ Id_n & 0 \end{pmatrix} \text{ in the half-space model.} \tag{173}$$

We choose as base point $P = iId_n$, analogue to having chosen $p = i$ for H^2.[31] The Cartan involution is hence:

$$\sigma : Sp(2n, \mathbb{R}) \to Sp(2n, \mathbb{R}) \tag{174}$$

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \mapsto \begin{pmatrix} -D & C \\ B & -A \end{pmatrix} \cong \begin{pmatrix} D & -C \\ -B & A \end{pmatrix} \text{ in the half-space model.} \tag{175}$$

Since the elements of G_P commute with g_{s_P}, i.e. $\sigma(k) = k$ for all $k \in G_P$, we find the isotropy group $G_{P=Id}$ as

$$SpO(2n, \mathbb{R}) = \begin{pmatrix} A & B \\ -B & A \end{pmatrix}, \ A^T A + B^T B = Id_n, \ A^T B \text{ symmetric matrices.} \tag{176}$$

This group is the intersection $SpO(2n, \mathbb{R}) = Sp(2n, \mathbb{R}) \cap O(2n, \mathbb{R})$ which is trivially a subgroup of $Sp(2n, \mathbb{R})$, since the product of orthogonal matrices is again orthogonal.

14.4 The dimension of the Siegel space

The Siegel space SH^n can be embedded in $\mathbb{C}^{n^2}$, as it consists of complex $(n \times n)$-matrices. This corresponds to an embedding space of real dimension $2n^2$. However, the intrinsic dimension of the Siegel space is only $n^2 + n$, since both real and imaginary part of the matrix representation as in eq. 160 are symmetric.
One can also infer this by looking at the coset formulation $SH^n \cong Sp(2n, \mathbb{R})/SpO(2n, \mathbb{R})$. The dimension of $\mathfrak{p}$ of the corresponding Cartan decomposition is equal to the dimension of SH^n. The Lie algebra $\mathfrak{sp}(2n, \mathbb{R})$ of $Sp(2n, \mathbb{R})$ reads:

$$\mathfrak{sp}(2n, \mathbb{R}) = \begin{pmatrix} A & B \\ C & -A^T \end{pmatrix}, \ B \text{ and } C \text{ symmetric matrices, } A \text{ arbitrary.} \tag{177}$$

---

[29]The factor $\frac{1}{2i}$ is necessary to make these matrices inverse to each other. The action has not changed since the factor cancels in $(AZ + B)(CZ + D)^{-1}$.

[30]Again, the symmetry at $Q \in SH^n$ is given by $s_Q = \theta(g) \circ s_P \circ \theta(g^{-1})$ for $\theta(g)P = Q, \ g \in Sp(2n, \mathbb{R})$.

[31]Recall that the choice of P is arbitrary since the Siegel space is a homogeneous space. If one chose Q instead, the group of isometries $Sp(2n, \mathbb{R})$ would stay the same. Instead the formulation of the isotropy group would change. In particular, if G_P is the isotropy group at P, the isotropy group at Q is $G_Q = g \cdot G_P \cdot g^{-1}$ with $\theta(g)P = Q$ and $g \in Sp(2n, \mathbb{R})$ since the isometry group acts transitively on SH^n. Both isotropy groups are isomorphic and one chooses the base point $P = iId$ because the isotropy group has a form easy to deal with.

Its Cartan decomposition is:

$$\mathfrak{sp}(2n, \mathbb{R}) = \mathfrak{p} \oplus \mathfrak{k} = \begin{pmatrix} A & B \\ B & -A \end{pmatrix} \oplus \begin{pmatrix} C & D \\ -D & C \end{pmatrix}, \quad A, B, D \text{ symmetric}, \ C \text{ skew-symmetric.}^{32} \qquad (178)$$

From the explicit form of $\mathfrak{p}$ one can read off the dimension of $\mathfrak{p}$ and hence SH^n as $2 \cdot \frac{n^2+n}{2}$ as both A and B have $\frac{n^2+n}{2}$ degrees of freedom.

14.5 The metric tensor

The Riemannian structure is (up to a constant) determined by the isometry group, i.e. the metric tensor g needs to have such a form that

$$g_Q(V, W) = \theta(h^{-1})^* g_P(V, W) = g_P\big(\theta(h^{-1})_* V, \theta(h^{-1})_* W\big)$$
$$\text{with } V, W \in T_Q SH^n, \ h^{-1} \in Sp(2n, \mathbb{R}), \ \theta(h^{-1})Q = P \qquad (179)$$

which means that the metric tensor is determined by its value at one single point P due to the transitivity of the action of the isometry group on the manifold.

For $P = iId$ one may choose the Euclidean metric. This does not mean that SH^n is Euclidean, since the metric only coincides with the Euclidean metric at one single point $P = iId$. For all other points it is given by eq. 179 which makes sure that the metric carries the non-Euclidean structure of SH^n.

This is also the case for the metric on the upper half-space H^2: The metric tensor $1/\Im(z)\mathrm{d}z \otimes \mathrm{d}\bar{z}$ coincides with the Euclidean metric at $z = i$. This motivates the choice of the Euclidean metric at $P = iId$, as the Siegel space SH^n should coincide with hyperbolic space H^2 for $n = 1$. ^{33}

The Euclidean metric on $\mathbb{C}^{n^2}$

The space SH^n has a representation as $Sym(n, \mathbb{C})$-matrices and is obviously a subset of $\mathbb{C}^{n^2}$ where the matrices are "vectorized", i.e. written as complex n^2-vectors. On $\mathbb{C}^{n^2}$, the familiar Euclidean scalar product has the form

$$\langle v, w \rangle = \sum_{i=1}^{n^2} v^i \bar{w}^i, \quad v, w \text{ complex } n^2\text{-vectors.} \qquad (180)$$

The Euclidean scalar product on $Sym(n, \mathbb{C})$ is the Frobenius inner product, which is the Euclidean scalar product above for vectorized matrices. The Frobenius inner product in matrix notation reads:

$$\langle V, W \rangle = \mathrm{tr}(V\overline{W}), \quad V, W \text{ complex } (n \times n)\text{-matrices.} \qquad (181)$$

It now only remains to determine $\theta(h^{-1})_*$ in eq. 179. The computation is carried out explicitly in section 14.10.1 resulting in:

$$\theta(h^{-1})_*(V) = \sqrt{Y_Q^{-1}} \cdot V \cdot \sqrt{Y_Q^{-1}}, \quad V \in T_Q SH^n, \ Q = X_Q + iY_Q \in SH^n. \qquad (182)$$

The metric at an arbitrary point Q is hence:

$$g_Q(V, W) = \mathrm{tr}\big(\sqrt{Y_Q^{-1}} \cdot V \cdot \sqrt{Y_Q^{-1}} \cdot \overline{\sqrt{Y_Q^{-1}} \cdot W \cdot \sqrt{Y_Q^{-1}}}\big) = \mathrm{tr}(Y_Q^{-1} V Y_Q^{-1} \overline{W}) \qquad (183)$$

using that the trace is cyclic and that Y_Q is a real matrix.

---

^{32}Note that $SpO(2n, \mathbb{R})$ in eq. 176 and $\mathfrak{k} = \mathfrak{spo}(2n, \mathbb{R}) = \mathfrak{sp}(2n, \mathbb{R}) \cap \mathfrak{so}(2n, \mathbb{R})$ in eq. 178 are not the same groups because of the additional requirements on the blocks of matrices. In particular $0 \in M_{2n \times 2n}$ is an element of $\mathfrak{spo}(2n, \mathbb{R})$ but not of $SpO(2n, \mathbb{R})$.

^{33}One could of course scale the metric tensors of both SH^n and H^2 as $g' = \lambda g$ to preserve both the compatibility with the isometry groups and the generalization property. This would result in the sectional curvature to become $K' = \lambda^{-1}K$. It is hence sufficient to consider the "standard" hyperbolic and Siegel spaces because all properties are just scaled by terms in λ.

14.6 The Riemannian gradient on SH^n

The gradient of a real valued function f is a vector-valued function on the manifold SH^n. The Siegel manifold can be represented as set of matrices with coordinates as in eq. 160. With respect to these coordinates one can find the derivative of a real valued function f. However, one has to take into account that the magnitude of a tiny displacement δX^i at Q is dependent on the evaluation point Q. The metric tensor takes care of that by scaling the Euclidean gradient according to the evaluation point. As vectors in SH^n are represented by matrices, the Riemannian gradient $\mathrm{grad}_{Rie}f|_Q$ has matrix form and lives in the tangent space of SH^n at Q.

With aid of the Euclidean metric one can now find the Riemannian gradient with [Theis, 2005, Theorem 3.3] via:

$$g_P(V, \mathrm{grad}_{Rie}f\big|_Q) = \mathrm{d}f\big|_Q(V) = \langle V, \mathrm{grad}_{Euc}f\big|_Q \rangle \text{ with evaluation point } Q = X_Q + iY_Q \tag{184}$$

which shows the relation between the Euclidean gradient and the Riemannian gradient. Rearranging the left hand side yields:

$$g_Q(V, \mathrm{grad}_{Rie}f\big|_Q) = \mathrm{tr}(Y_Q^{-1}VY_Q^{-1}\overline{\mathrm{grad}_{Rie}f\big|_Q}) = \mathrm{tr}(V\overline{Y_Q^{-1}\mathrm{grad}_{Rie}f\big|_Q Y_Q^{-1}}) = \langle V, \overline{Y_Q^{-1}\mathrm{grad}_{Rie}f\big|_Q Y_Q^{-1}} \rangle. \tag{185}$$

Putting eq. 184 and eq. 185 together and using that the matrix Y_Q^{-1} is a matrix with real entries one arrives at:

$$\langle V, \mathrm{grad}_{Euc}f\big|_Q \rangle = \langle V, Y_Q^{-1}\mathrm{grad}_{Rie}f\big|_Q Y_Q^{-1} \rangle$$
$$\Rightarrow \mathrm{grad}_{Euc}f\big|_Q = Y_Q^{-1}\mathrm{grad}_{Rie}f\big|_Q Y_Q^{-1} \tag{186}$$
$$\Rightarrow \mathrm{grad}_{Rie}f\big|_Q = Y_Q\mathrm{grad}_{Euc}f\big|_Q Y_Q.$$

The Riemannian gradient can thus be calculated at Q from the Euclidean gradient via two simple matrix multiplications. Note that Riemannian and Euclidean gradient coincide at $P = iId$ since $Y_P = Id$.

14.7 The submanifold $(H^2)^n \subset SH^n$

The subspace $(H^2)^n$ can be found within SH^n. Note that this does not mean that SH^n is decomposable. Clearly, the matrix

$$\begin{pmatrix} z_1 & & & \\ & z_2 & & \\ & & \ddots & \\ & & & z_n \end{pmatrix} \text{ with } \Im(z_i) > 0 \tag{187}$$

corresponds to a point in SH^n, because it is trivially symmetric and its imaginary part is positive definite. To explicitly see why $(H^2)^n$ is a submanifold, one wants to find those isometries which only alter one H^2 subspace of $(H^2)^n$, i.e. $SL(2,\mathbb{R})^n \subset Sp(2n,\mathbb{R})$. The subspaces $SO(2,\mathbb{R})^n$ are then trivially contained in $Sp(2n,\mathbb{R})$, because $SO(2,\mathbb{R}) \subset SL(2,\mathbb{R})$.

We show explicitly that there are n $SL(2,\mathbb{R})$-subgroups by introducing matrices which correspond to each subgroup: The n matrices of the block matrix form

$$M_1 := \left(\begin{array}{ccccc|ccccc} a_1 & & & & & b_1 & & & & \\ & 1 & & & & & & & & \\ & & 1 & & & & & & & \\ & & & \ddots & & & & & & \\ & & & & 1 & & & & & \\ \hline c_1 & & & & & d_1 & & & & \\ & & & & & & 1 & & & \\ & & & & & & & 1 & & \\ & & & & & & & & \ddots & \\ & & & & & & & & & 1 \end{array}\right), \quad M_2 := \left(\begin{array}{ccccc|ccccc} 1 & & & & & & & & & \\ & a_2 & & & & & b_2 & & & \\ & & 1 & & & & & & & \\ & & & \ddots & & & & & & \\ & & & & 1 & & & & & \\ \hline & & & & & 1 & & & & \\ & c_2 & & & & & d_2 & & & \\ & & & & & & & 1 & & \\ & & & & & & & & \ddots & \\ & & & & & & & & & 1 \end{array}\right), \quad \ldots \tag{188}$$

56

with zeros everywhere else are symplectic $(2n \times 2n)$-matrices. The matrix M_i only acts on z_i in the H^2 form of eq. 187. Each action of M_i can be identified with a matrix in $SL(2, \mathbb{R})$, which shows that $SL(2, \mathbb{R})^n \subset Sp(2n, \mathbb{R})$. We now check that the Lie algebras which only correspond to one H^2 subspace of $(H^2)^n$ are contained within the Lie algebra of $Sp(2n, \mathbb{R})$, i.e. $\bigoplus_{i=1}^{n} \mathfrak{sl}(2, \mathbb{R}) \subset \mathfrak{sp}(2n, \mathbb{R})$.

Lie algebra of $SL(2, \mathbb{R})$

The Lie algebra of $SL(2, \mathbb{R})$ is $\mathfrak{sl}(2, \mathbb{R})$, the vector space of traceless (2×2)-matrices. The Lie algebra of the subgroup $SO(2, \mathbb{R}) \subset SL(2, \mathbb{R})$ is $\mathfrak{so}(2, \mathbb{R})$, the vector space of skew-symmetric (2×2)-matrices. One can find a basis of $\mathfrak{sl}(2, \mathbb{R})$, namely:

$$\mathfrak{sl}(2, \mathbb{R}) = \mathfrak{p} \oplus \mathfrak{k} = \left\langle \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right\rangle \oplus \left\langle \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \right\rangle \tag{189}$$

where the brackets stand for the linear span. The first summand correspond to $\mathfrak{sl}(2, \mathbb{R}) / \mathfrak{so}(2, \mathbb{R}) \cong \mathfrak{p}$ in the Cartan decomposition, the second summand corresponds to $\mathfrak{so}(2, \mathbb{R}) \cong \mathfrak{k}$.

The elements of $\mathfrak{p}$ generate transvections via the flow of their respective corresponding Killing vector fields. The first basis vector (which is a matrix) of $\mathfrak{p}$ from eq. 189 corresponds to the vector ∂_y. The second basis vector corresponds to the vector ∂_x at $p = i$.

This can be seen via exponentiation of the matrix in $\mathfrak{p}$: One obtains a matrix of $SL(2, \mathbb{R})$ which acts on H^2 as an isometry (a transvection) shifting the base point $p = i$ into the direction of the Killing vector fields, i.e. along geodesics.
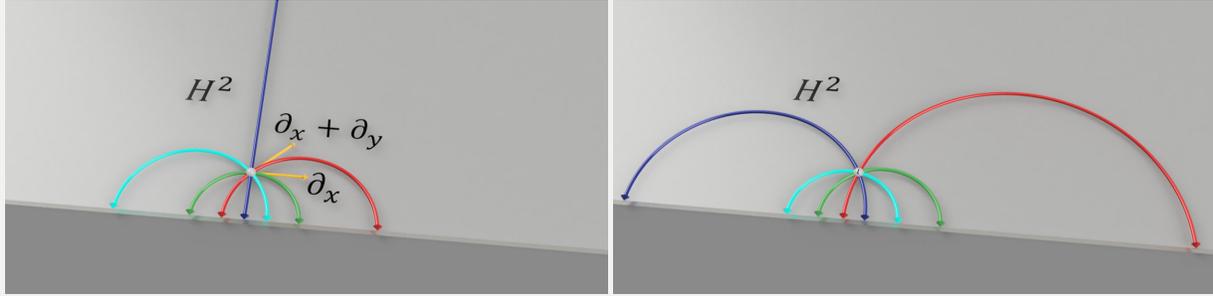
Figure 25: Geodesics on the upper half-space H^2 emanating from the base point $p = i$. On the left, the blue geodesic points in direction ∂_y, the green geodesic is obtained by $\gamma(t) = \exp_i t \cdot \partial_x$.

On the right, these geodesics are transformed via an isometry fixing p. This isometry is an element of the isotropy group and can hence be obtained as: $\exp_{Id_{2 \times 2}}(\phi \cdot \mathfrak{k}) \in SO(2, \mathbb{R})$, where the parameter ϕ determines the rotation angle. Since the geodesics are of infinite length, they approach the boundary $x = 0$, but will never reach it. This is indicated by small arrows.

The Lie algebra of $Sp(2n, \mathbb{R})$

The Lie algebra of the symplectic group is $\mathfrak{sp}(2n, \mathbb{R})$, the set of $(2n \times 2n)$-matrices of the form:

$$\mathfrak{sp}(2n, \mathbb{R}) = \mathfrak{p} \oplus \mathfrak{k} = \begin{pmatrix} A & B \\ B & -A \end{pmatrix} \oplus \begin{pmatrix} C & D \\ -D & C \end{pmatrix}, \quad A, B, D \text{ symmetric}, \ C \text{ skew-symmetric.} \tag{190}$$

It is apparent that the Siegel space SH^1 of dimension 2 is just hyperbolic space H^2:

$$Sp(2 \cdot 1, \mathbb{R}) = SL(2, \mathbb{R}), \quad \mathfrak{sp}(2 \cdot 1, \mathbb{R}) = \mathfrak{sl}(2, \mathbb{R}). \tag{191}$$

To see explicitly that the Lie algebra $\mathfrak{sl}(2, \mathbb{R})$ corresponding to $(H^2)^n$ is contained in $\mathfrak{sp}(2n, \mathbb{R})$ corresponding to SH^n, consider the matrices:

$$
\tilde{M}_1 := \left(\begin{array}{cccccc|cccccc}
\alpha_1 & & & & & & \beta_1 & & & & & \\
& 0 & & & & & & & & & & \\
& & 0 & & & & & & & & & \\
& & & \ddots & & & & & & & & \\
& & & & 0 & & & & & & & \\
\hline
\gamma_1 & & & & & & -\alpha_1 & & & & & \\
& & & & & & & 0 & & & & \\
& & & & & & & & 0 & & & \\
& & & & & & & & & \ddots & & \\
& & & & & & & & & & 0 &
\end{array} \right) , \quad
\tilde{M}_2 := \left(\begin{array}{cccccc|cccccc}
0 & & & & & & 0 & & & & & \\
& \alpha_2 & & & & & & \beta_2 & & & & \\
& & 0 & & & & & & & & & \\
& & & \ddots & & & & & & & & \\
& & & & 0 & & & & & 0 & & \\
\hline
& \gamma_2 & & & & & & -\alpha_2 & & & & \\
& & & & & & & & 0 & & & \\
& & & & & & & & & \ddots & & \\
& & & & & & & & & & 0 &
\end{array} \right) , \quad \dots
$$
(192)

with zeros everywhere else. These $(2n \times 2n)$-matrices are contained in the Lie algebra $\mathfrak{sp}(2n, \mathbb{R})$ which is immediate to check. In fact, the small (2×2)-matrices within the $(2n \times 2n)$-matrices above generated by the numbers $(\alpha_i, \beta_i, \gamma_i)$ are traceless and span the (2×2)-matrices of $\mathfrak{sl}(2, \mathbb{R})$ from eq. 189.

14.7.1 The subspace $(H^2)^n$ as totally geodesic submanifold

In the last section it was shown that $\mathfrak{sl}(n, \mathbb{R})^n \subset \mathfrak{sp}(2n, \mathbb{R})$. Now one wants to know, whether the induced submanifold $(H^2)^n$ is a totally geodesic subspace of SH^n. For this it is easy to verify that the transvection vectors $(\mathfrak{p}_{H^2})^n$ are a Lie triple system within $\mathfrak{p}_{SH^n}$ from eq. 178. For this, recall that $\mathfrak{p}_{H^2}$ is spanned by the first two entries of the Lie algebra $\mathfrak{sl}(2, \mathbb{R})$ in eq. 189. Thus, $(\mathfrak{p}_{H^2})^n$ has the matrix form similar to the span of those matrices of eq. 192 with $\gamma_i = \beta_i$. It is then straight-forward to show:

$$
\left[\left[\begin{pmatrix} D_{\alpha_1} & D_{\beta_1} \\ D_{\beta_1} & -D_{\alpha_1} \end{pmatrix} , \begin{pmatrix} D_{\alpha_2} & D_{\beta_2} \\ D_{\beta_2} & -D_{\alpha_2} \end{pmatrix} \right] , \begin{pmatrix} D_{\alpha_3} & D_{\beta_3} \\ D_{\beta_3} & -D_{\alpha_3} \end{pmatrix} \right] \subseteq \begin{pmatrix} D_{\alpha_4} & D_{\beta_4} \\ D_{\beta_4} & -D_{\alpha_4} \end{pmatrix}
$$
(193)

in block matrix form with D_i a real diagonal $(n \times n)$-matrix.

Results from $(H^2)^n$ being totally geodesic

Since $(H^2)^n$ is a totally geodesic subspace, any geodesic within $(H^2)^n$ is also a geodesic within SH^n. This simplifies finding a geodesic γ connecting two arbitrary points P and Q and calculating the distance:

Due to the transitivity of the symplectic group action, it is possible to find an isometry ϕ mapping P to iId and Q to iD where $D = \text{diag}(q^1, ..., q^n)$ is a diagonal matrix.[34] The isometry ϕ can be understood as a coordinate system change. It is then easy to find the geodesic $\tilde{\gamma}$ connecting $\phi(P)$ and $\phi(Q)$ as $\tilde{\gamma}(t) = \text{diag}(i \exp(q^1 t), ..., i \exp(q^n t))$. The distance between $\phi(P)$ and $\phi(Q)$ just decomposes into the distance calculation for every subspace H^2 as shown in eq. 93.

The transformation ϕ^{-1} maps the computed geodesic $\tilde{\gamma}$ back to γ in the original coordinate system. Since ϕ and thus ϕ^{-1} map geodesics to geodesics, γ is in fact a geodesic since $\tilde{\gamma}$ is a geodesic. It follows $d(P, Q) = d(\phi(P), \phi(Q))$.

Transforming a problem to the subspace $(H^2)^n$ will turn out to be very useful for computing the distance in section 14.11 and the exponential map in section 14.10 below.

14.8 The submanifold $P(n, \mathbb{R}) = GL(n, \mathbb{R})/O(n, \mathbb{R}) \subset SH^n$

The space $P(n, \mathbb{R}) =: P(n)$ is the space of symmetric positive definite matrices with real entries. Looking at the elements of the Siegel space, namely $Z = X + iY$ with Y symmetric and positive definite, one gets an impression that $P(n)$ is a subset of the pureley imaginary part of SH^n. $P(n)$ can be understood as the space of positive definite scalar products, since an "ordinary" scalar product in matrix form is symmetric and positive definite.

---

[34]For explicit computation see section 14.10 about the Riemann exponential map.

The coset formulation of $P(n)$ is $GL(n,\mathbb{R})/O(n,\mathbb{R})$ [Eschenburg, 1997, 1, E.9]. The group $GL(n,\mathbb{R})$ acts as isometries on $P(n)$ as:

$$\theta : GL(n,\mathbb{R}) \to \mathcal{M}aps(P(n); P(n)) \tag{194}$$

$$\theta(g)(Q) = g \cdot Q \cdot g^T, \ \ Q \in P(n) \tag{195}$$

because it preserves the inner product:

$$\langle V, W \rangle_Q = \mathrm{tr}(Q^{-1}VQ^{-1}), \ \ Q \in P(n). \tag{196}$$

This inner product is the restriction of the inner product on the full space SH^n because the term Y_Q^{-1} in eq. 183 is exactly the part Q^{-1}: $P(n)$ can be seen as the imaginary part Y of $Z = X + iY \in SH^n$.

Let us choose $P = Id_n$ as base point.[35] The isotropy group of $P(n)$ at Id_n is $O(n,\mathbb{R})$, since $\theta(k)(Id_n) = k \cdot Id_n \cdot k^T = Id_n$, because k is an orthogonal matrix.

To show that $GL(n,\mathbb{R})/O(n,\mathbb{R})$ is a submanifold of $SH^n = Sp(2n,\mathbb{R})/SO(n,\mathbb{R})$, it is sufficient to show that $GL(n,\mathbb{R}) \subset Sp(2n,\mathbb{R})$. Since $O(n,\mathbb{R}) \subset GL(n,\mathbb{R})$, it follows that $O(n,\mathbb{R}) \subset SpO(2n,\mathbb{R})$.

Consider the inclusion

$$GL(n,\mathbb{R}) \hookrightarrow Sp(2n,\mathbb{R})$$
$$G \mapsto \begin{pmatrix} G & 0 \\ 0 & (G^T)^{-1} \end{pmatrix} \tag{197}$$

and its corresponding inclusion in the Lie algebra of $Sp(2n,\mathbb{R})$:

$$\mathfrak{gl}(n,\mathbb{R}) \hookrightarrow \mathfrak{sp}(2n,\mathbb{R})$$
$$M \hookrightarrow \begin{pmatrix} M & 0 \\ 0 & -M \end{pmatrix}, \ \ M \text{ arbitrary.} \tag{198}$$

The action $GL(n,\mathbb{R}) \curvearrowright P(n)$ fits to the inherited action $Sp(2n,\mathbb{R}) \curvearrowright SH^n : \theta(G)(Z) = GZG^T$. One can see that the space $GL(n,\mathbb{R})/O(n,\mathbb{R})$ thus nicely fits into the coset formulation of SH^n.

In fact, the framework for the $GL(n,\mathbb{R})/O(n,\mathbb{R})$ lays the foundation for the following very powerful general submanifold $SL(n,\mathbb{R})/SO(n,\mathbb{R})$:

14.9 The submanifold $SL(n,\mathbb{R})/SO(n,\mathbb{R})$

Since $GL(n,\mathbb{R})/O(n,\mathbb{R})$ is a submanifold of the Siegel space, it is straight-forward to see that $SL(n,\mathbb{R})/SO(n,\mathbb{R})$ is also a submanifold. In fact, it is also a totally geodesic submanifold: Since the transvection space $\mathfrak{sl}(n,\mathbb{R})$ is the space of symmetric traceless matrices, it is closed under the Lie triple product[36].

The embedding into $SL(n,\mathbb{R})/SO(n,\mathbb{R})$

The space $SL(n,\mathbb{R})/SO(n,\mathbb{R})$ is incredibly versatile. In fact, one can read off one explicit submanifold already, namely $H^2 \cong SL(2,\mathbb{R})/SO(2,\mathbb{R})$. In that sense, it is a generalization of hyperbolic space. In fact, any symmetric space of noncompact type[37] can be found within $SL(n,\mathbb{R})/SO(n,\mathbb{R})$ for adequate n. The natural[38] metric tensor on $SL(n,\mathbb{R})/SO(n,\mathbb{R})$ fits up to a constant to the metric tensor on the starting symmetric space to be embedded. This is a remarkable result for the Siegel space, because it shows that all negatively curved symmetric spaces can be found as totally geodesic subspaces within it. Note that we had already explored the totally geoedesic submanifold of $(H^2)^n$. It was found that H^3 is not a submanifold of $(H^2)^n$. By this result however, H^m *can be* embedded as totally geodesic submanifold. In particular, for any index set I, the product manifold $\left(\bigtimes_I H^I\right) \times \mathbb{R}^r$

---

[35]Keeping in mind that $P(n)$ is the imaginary part of SH^n, this is the same reference point as iId_n in the full Siegel space.

[36]The commutator or two symmetric matrices is an anti-symmetric matrix, the commutator of a symmetric matrix with an anti-symmetric matrix is a symmetric matrix. Since the trace of any commutator is zero (as the trace is cyclic) the statement follows.

[37]Pictorially speaking, this means that the space is infinite in any direction.

[38]The metric tensor induced by the Killing form as described in section 14.12.

can be found as totally geodesic submanifold within $SL(n,\mathbb{R})/SO(n,\mathbb{R})$ for adequate n. In particular, even the Siegel space SH^n itself can be embedded into a submanifold of $SL(2n^2-n,\mathbb{R})/SO(2n^2-n,\mathbb{R})$[39]. A sphere $S^m \cong SO(m+1,\mathbb{R})/SO(m,\mathbb{R})$ however cannot be embedded in $SL(n,\mathbb{R})/SO(n,\mathbb{R})$ as the isometry group $SO(m+1,\mathbb{R})$ lies in the isotropy group K of $SL(n,\mathbb{R})/SO(n,\mathbb{R})$: Any transvection in S^m is in fact killed by K. Let us state the general theorem about the embedding into $SL(n,\mathbb{R})/SO(n,\mathbb{R})$ in a bit more detail:

Embedding theorem in $SL(n,\mathbb{R})/SO(n,\mathbb{R})$ [Eberlein, 1996, Theorem 2.6.5]

For any symmetric space L_m, $\dim(L_m) = m$ of noncompact type there is an (essentially) isometric diffeomorphism F sending L_m to a totally geodesic submanifold M_{emb} of $SL(n,\mathbb{R})/SO(n,\mathbb{R})$ with $n = I_0(L_m)$ [a]. The metric tensor on L_m is equal to the restriction of the metric in $SL(n,\mathbb{R})/SO(n,\mathbb{R})$ to the submanifold up to a positive constant.
Let $p \in L_m$ be the base point of $L_m \cong G/K$, then the map F has the form:

$$F : L_m \to SL(n,\mathbb{R})/SO(n,\mathbb{R}) \tag{199}$$
$$q = g(p) \mapsto \text{Ad}(g) \cdot k, \quad k \in SO(n,\mathbb{R}), \quad g \in G. \tag{200}$$

In particular the base point is mapped to $Id_n \in SL(n,\mathbb{R})/SO(n,\mathbb{R})$.

---

[a]Note that $\dim(I(L_m)) \neq \dim(L_m)$, e.g. $\dim(S^j) = j \neq \dim(SO(j+1)) = \frac{j(j+1)}{2}$ for the j-sphere.

The metric induced on the submanifold of $SL(n,\mathbb{R})/SO(n,\mathbb{R})$ by the pullback of F does not yield *every* metric which one would wish for. The metric tensor is scaled with a positive constant. This is meant by the notion of an essentially isometric map. In particular, not all hyperbolic spaces of any curvature can be found as totally geodesic submanifolds, since the metric on $SL(n,\mathbb{R})/SO(n,\mathbb{R})$ and hence on its submanifolds is fixed. In conclusion, spaces as H_r^j cannot be embedded as totally geodesic submanifolds with correct curvature.

14.10 The Riemann exponential map on the Siegel space

The Riemann exponential map sends a vector W at an arbitrary point Q to a point $\tilde{Q}$ in the Siegel space. The set (Q, W) of starting point and vector uniquely determines a geodesic within the Siegel space. The Riemann exponential map yields exactly that point R such that the geodesic distance between Q and R is the length of the vector W. This map within SH^n can be computed via the geodesic equation eq. 23. However, for this procedure one needs to solve $n^2 + n$ differential equations which one tries to avoid.

Using the knowledge of symmetric spaces, one can find two[40] ways to circumvent these differential equations depicted in the diagram below:

---

[39]The dimension of the isometry group of the Siegel space is $2n^2 - n$ as one can read off from eq. 14.4.

[40]There are of course other ways to do so, such as applying a coordinate transformation to move the emanating point Q and its tangent vector into the totally geodesic submanifold $P(n)$. The following will focus on two procedures depicted in red and orange.
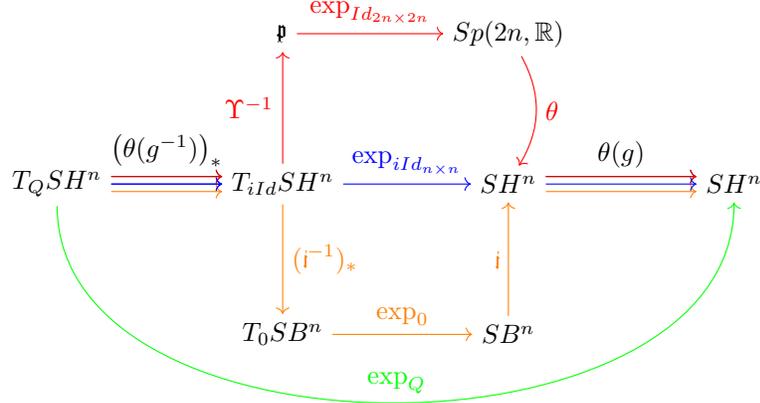
Figure 26: Diagram of the setup for the exponential map in the Siegel space. The green arrow symbolizes the solution to the differential equations. Three alternatives are depicted in red, blue and orange. They represent the paths via the coset formulation, via the Riemann exponential in SH^n and via the ball model SB^n, respectively. All three paths first send the vector at point Q to the base point $P = iId_n$ via $\theta(g^{-1})_*$ for computational reasons. After the exponential mapping, the result is sent back via $\theta(g)$. This procedure can be understood as a coordinate transformation.

14.10.1 Coordinate transformation

Looking at the diagram fig. 26 above, we need the maps $(\theta(g^{-1}))_*$ and $\theta(g)$ in order to shift the computation of the Riemann exponential map to the base point $P = iId_n$. Transporting the problem to the base point makes sense, since the Siegel space is a homogeneous space, i.e. all points can be considered as equal. Since the symplectic group $Sp(2n, \mathbb{R})$ acts transitively on SH^n, one can find an isometry (an element of $Sp(2n, \mathbb{R})$) which sends the base point $P = iId$ to any point $Q = X_Q + Y_Q$:

$$SH^n \to Sp(2n, \mathbb{R})$$
$$Q \mapsto \begin{pmatrix} \sqrt{Y_Q} & X_Q\sqrt{Y_Q^{-1}} \\ 0 & \sqrt{Y_Q^{-1}} \end{pmatrix} =: g, \quad \text{with} \ \ \theta(g)(P) = Q. \tag{201}$$

This map can be split into two symplectic matrices:

$$g = g_{\Re} \cdot g_{\Im} \tag{202}$$

$$\begin{pmatrix} \sqrt{Y_Q} & X_Q\sqrt{Y_Q^{-1}} \\ 0 & \sqrt{Y_Q^{-1}} \end{pmatrix} = \begin{pmatrix} Id_n & X_Q \\ 0 & Id_n \end{pmatrix} \cdot \begin{pmatrix} \sqrt{Y_Q} & 0 \\ 0 & \sqrt{Y_Q^{-1}} \end{pmatrix} \tag{203}$$

which corresponds to concatenating the mappings $g_{\Im} \curvearrowright Z \mapsto \sqrt{Y_Q}Z\sqrt{Y_Q}$ and $g_{\Re} \curvearrowright Z \mapsto Z + X_Q$. The push-forward map $\theta(g)_*$ which sends a vector V at P to a vector $\theta(g)_*(V)$ at $Q = \theta(g)(P)$ then boils down to:

$$\theta(g)_*(V) = \sqrt{Y_Q} \cdot V \cdot \sqrt{Y_Q} \ \text{with "·" as matrix multiplication.} \tag{204}$$

Now one can easily find the maps $\theta(g^{-1})$ and $\theta(g^{-1})_*$ because of the group action property:

$$\theta(g^{-1}) = \theta(g)^{-1} \ \text{ and hence}$$
$$\theta(g^{-1})_*(V) = \sqrt{Y_Q^{-1}} \cdot V \cdot \sqrt{Y_Q^{-1}}. \tag{205}$$

To conclude, the relevant maps are eq. 201 and eq. 205.

14.10.2 The exponential map via the coset formulation

The existence of the correspondence of the coset formulation, i.e. $SH^n \cong Sp(2n, \mathbb{R})/SpO(2n, \mathbb{R})$, has been shown already in eq. 84. However, in order to use the anti-isomorphism Υ^{-1} from eq. 67 between the vector spaces $T_{iId}SH^n$ and $\mathfrak{p}$, this map firstly has to be found explicitly. This enables us to calculate the exponential map within the coset formulation as matrix exponential: An element $\mathfrak{v} \in \mathfrak{p}$ is sent via the matrix exponential to a symplectic matrix S. This matrix corresponds to an isometry on the Siegel space. Letting S act on the base point $P = iId$, one obtains the resulting point $\tilde{R} \in SH^n$. The red path in the diagram 26 above is then explicitly shown below:
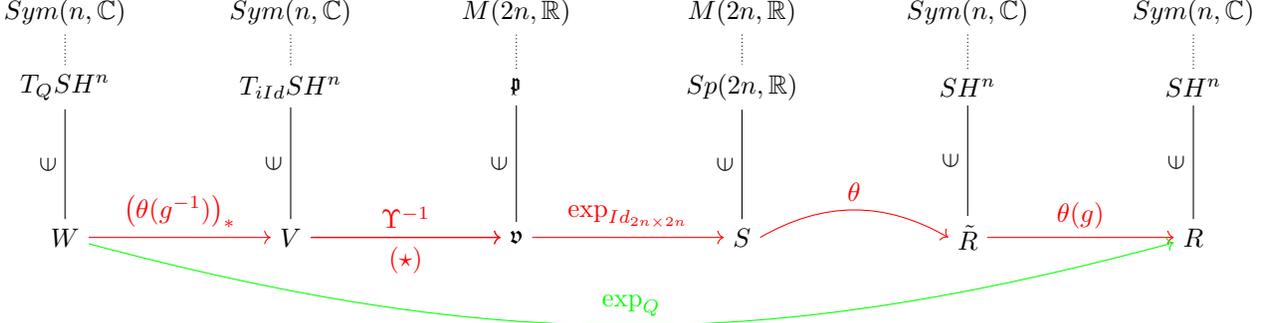
Figure 27: Diagram of the setup for the exponential map in the Siegel space via the coset formulation. The arrow marked by $(\star)$ is not trivial: The space $T_{iId}SH^n$ is a vector space of real dimension $(n^2 - n)$ as is the vector space $\mathfrak{p}$. However, $T_{iId}SH^n$ consists of complex $(n \times n)$-matrices, whereas $\mathfrak{p}$ consists of real $(2n \times 2n)$-matrices.

Let us postpone the construction of the anti-isomorphism Υ^{-1} for a short time. With aid of Υ^{-1}, any vector in $T_{iId}SH^n$ has a unique corresponding element in $\mathfrak{p}$: Any vector can be linearly decomposed into basis vectors $\{V_i\}$, resulting in a linear combination of basis vectors $\{b_i\}$ in $\mathfrak{p}$. The resulting real $(2n \times 2n)$-matrix in $\mathfrak{sp}(2n, \mathbb{R})$ can be exponentiated (this is the $(2n \times 2n)$-matrix exponentiation) which yields a symplectic matrix. Letting this symplectic matrix act on $P = iId$ yields the point $\tilde{R}$. Sending this point via $\theta(g)$ to R undoes the very first coordinate change and is hence the result of the Riemann exponential $\exp_Q(W) = R$.

The correspondence of $T_{iId}SH^n$ and $\mathfrak{p}$

The crucial part of the exponential map via the coset formulation is the connection between the tangent space $T_{iId}SH^n$ and the transvection space $\mathfrak{p}$. The following procedure will construct the anti-ismorphism Υ and hence the needed inverse map Υ^{-1}:

- Find a basis $\{b_i\}_{i=1,\dots,n^2+n}$ of the vector space $\mathfrak{p}$ in terms of $(2n \times 2n)$-matrices. Since $\dim(\mathfrak{p}) = n^2 + n$, there are $(n^2 + n)$ basis matrices. We know from eq. 178:

$$\mathfrak{p} = \begin{pmatrix} A & B \\ B & -A \end{pmatrix}, \quad A, B \text{ symmetric} \tag{206}$$

  which has $(n^2 + n)$ degrees of freedom.

- Parameterize the basis matrix b_i by multiplying with the parameter t.

- Use the matrix exponential at the identity $Id_{2n \times 2n}$ of that parameterized basis matrix. This yields a real symplectic $(2n \times 2n)$-matrix $S_i(t) := \exp(t \cdot b_i)$.

- Let the resulting matrix $S_i(t)$ act on $Z \in SH^n$, this yields a flow ϕ_t on SH^n parameterized by t.

- The velocity vector at $t = 0$ of the flow ϕ_t is the corresponding Killing vector field

$$\left(\frac{d}{dt} \theta(S_i(t))(Z) \right)\Big|_{t=0} =: X(Z) \in \mathcal{X}(SH^n). \tag{207}$$

62

- Evaluating the Killing vector field at $Z = iId_{n \times n}$ yields the basis vector $V_i \in T_{iId}SH^n$ we are searching:

$$X(iId_{n \times n}) =: V_i \in T_{iId}SH^n. \tag{208}$$

- The repetition for all basis matrices b_i of $\mathfrak{p}$ yields a basis $\{v_i\}_{i=1,\ldots,n^2+n}$ of $T_{iId}SH^n$. The anti-isomorphism Υ is hence determined. Since this map is just an assignment map between $\{b_i\}$ and $\{V_i\}$, the inverse map Υ^{-1} is automatically found.

A basis $\{b_i\}_{i=1,\ldots,n^2+n}$ of the vector space $\mathfrak{p}$ can be found easily when looking at its matrix form:

$$\mathfrak{p} \cong \begin{pmatrix} A & B \\ B & -A \end{pmatrix} \text{ with } A, B \in Sym(n, \mathbb{R}). \tag{209}$$

One can find four sets of basis vectors which correspond to:

1. the diagonal part of the block matrix A with n elements,

2. the diagonal part of the block matrix B with n elements,

3. the off-diagonal part of the block matrix A with $\frac{n^2-n}{2}$ elements,

4. the off-diagonal part of the block matrix B with $\frac{n^2-n}{2}$ elements.

The above mentioned procedure yields after some tedious calculations:

Figure 28: The basis vectors $\{b_1, \ldots, b_n\} \in \mathfrak{p}$ represent the diagonal part of the block matrix A in eq. 209. The n corresponding vectors in $T_{iId}SH^n$ represent the diagonal part of the imaginary part of the vector space $T_{iId}SH^n \cong Sym(n, \mathbb{C})$.

$$b_{n+1} := \left(\begin{array}{cccc|cccc} & & & 1 & & & \\ & \vdots & & & 0 & & \\ \cdots & 0 & \cdots & & & \ddots & \\ & \vdots & & & & & 0 \\ \hline & & & & & & & 0 \\ 1 & & & & & & \\ & 0 & & & & \vdots & \\ & & \ddots & & \cdots & 0 & \cdots \\ & & & 0 & & \vdots & \\ & & & & 0 \end{array} \right), b_{n+2} := \left(\begin{array}{cccc|cccc} & & & 0 & & & \\ & \vdots & & & 1 & & \\ \cdots & 0 & \cdots & & & 0 & \\ & \vdots & & & & \ddots & \\ \hline 0 & & & & & & & 0 \\ & & & & & & \\ 1 & & & & & \vdots & \\ & 0 & & & \cdots & 0 & \cdots \\ & & \ddots & & & \vdots & \\ & & & 0 \end{array} \right), \ldots$$

Figure 29: The n basis vectors $b_{n+1}, ..., b_{2n} \in \mathfrak{p}$ represent the diagonal part of the block matrix B in eq. 209. The n corresponding vectors in $T_{iId}SH^n$ represent the diagonal part of the real part of the vector space $T_{iId}SH^n \cong Sym(n, \mathbb{C})$.

$$V_{n+1} := \left(\begin{array}{ccc} 2 & & \\ 0 & \vdots & \\ \cdots & 0 & \cdots \\ & \vdots & \end{array} \right) + i \left(\begin{array}{ccc} & \vdots & \\ \cdots & 0 & \cdots \\ & \vdots & \end{array} \right), \ V_{n+2} := \left(\begin{array}{ccc} 0 & & \\ 2 & \vdots & \\ \cdots & 0 & \cdots \\ & \vdots & \end{array} \right) + i \left(\begin{array}{ccc} & \vdots & \\ \cdots & 0 & \cdots \\ & \vdots & \end{array} \right), \ldots$$

After having found the first $2n$ basis vectors, we investigate their nature, in particular which geodesics they induce. When looking at section 14.7 about $(H^2)^n \subset SH^n$, one easily sees that the vectors $\{b_1, ..., b_{2n}\}$ generate all geodesics within this subspace. In particular, eq. 192 shows that $\{b_1, ..., b_n\}$ correspond to those n geodesics which each point to i-direction of the hyperbolic upper half-space H^2. In Euclidean coordinates, this direction corresponds to the y-coordinate. The vectors $\{b_{n+1}, ..., b_{2n}\}$ thus correspond to those n geodesics which each point to $+1$-direction of the hyperbolic upper half-space, i.e. in Euclidean x-direction. The corresponding geodesics form halfcircles:
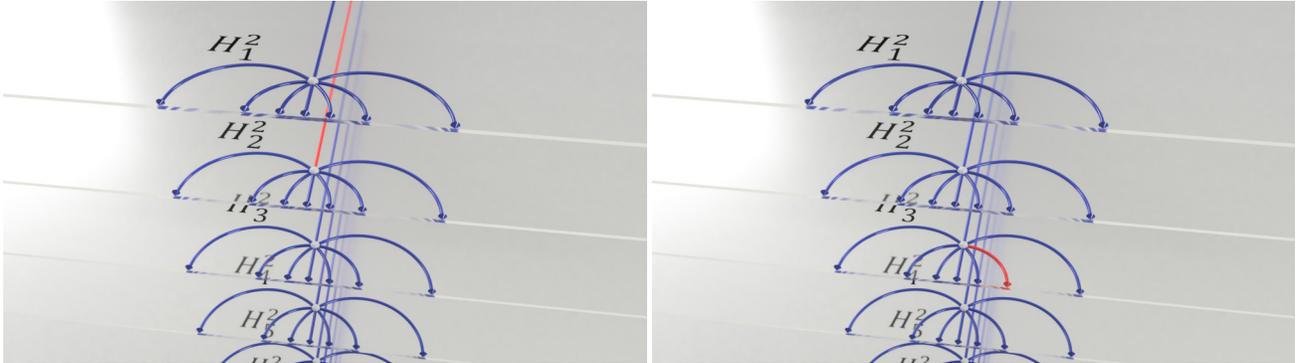
Figure 30: On the left: Illustration of the vector V_2 from the set $\{V_1, ..., V_n\}$. These vectors generate transvections, i.e. geodesics in i-direction within a subspace H^2. In particular, the vector V_2 generates the geodesic indicated by red color.
On the right: Illustration of the vector V_{n+2} from the set $\{V_{n+1}, ..., V_{2n}\}$. These n vectors generate transvections in the real direction in the complex upper half-plane. The geodesic which is marked by red color is generated by the vector V_{n+3}.
Note that the n hyperbolic spaces H^2 stand perpendicularly on each other, which cannot be represented accurately in three dimensions.

$$b_{2n+1} := \left(\begin{array}{cc|cc} 0 & 1 & & \\ 1 & 0 & \vdots & \cdots & 0 & \cdots \\ & \cdots & 0 & \cdots & & \vdots \\ & \vdots & \ddots & & & \\ \hline & & 0 & -1 & & \\ & \vdots & -1 & 0 & \vdots & \\ & \cdots & 0 & \cdots & \cdots & 0 & \cdots \\ & \vdots & & & \vdots & \ddots \end{array} \right), ..., \quad b_{2n+1+k} := \left(\begin{array}{c|c} \begin{matrix} & & 0 & 0 \\ \vdots & 0 & 1 \\ \cdots & 0 & \cdots \\ 0 & 0 & \vdots \\ 0 & 1 \end{matrix} & \begin{matrix} \vdots \\ \cdots & 0 & \cdots \\ \vdots \end{matrix} \\ \hline \begin{matrix} \vdots \\ \cdots & 0 & \cdots \\ \vdots \end{matrix} & \begin{matrix} 0 & 0 \\ \vdots & 0 & -1 \\ \cdots & 0 & \cdots \\ 0 & 0 & \vdots \\ 0 & -1 \end{matrix} \end{array} \right), ...$$

$$V_{2n+1} := \left(\begin{matrix} & \vdots & \\ \cdots & 0 & \cdots \\ & \vdots & \end{matrix} \right) + i \left(\begin{matrix} 0 & 2 & \\ 2 & 0 & \vdots \\ & \cdots & 0 & \cdots \\ & & \vdots \end{matrix} \right), ..., \quad V_{2n+1+k} := \left(\begin{matrix} & \vdots & \\ \cdots & 0 & \cdots \\ & \vdots & \end{matrix} \right) + i \left(\begin{matrix} & & 0 & 0 \\ & \vdots & 0 & 2 \\ \cdots & 0 & \cdots \\ 0 & 0 & \vdots \\ 0 & 2 \end{matrix} \right), ...$$

Figure 31: The basis vectors $\{b_{2n+1}, ..., b_{\frac{n(n+3)}{2}}\} \in \mathfrak{p}$ represent the off-diagonal part of the block matrix A in eq. 209. The $\frac{n^2-n}{2}$ corresponding vectors in $T_{iId}SH^n$ represent the off-diagonal part of the imaginary part of the vector space $T_{iId}SH^n \cong Sym(n, \mathbb{C})$.

$$b_{\frac{n(n+3)}{2}+2} := \left(\begin{array}{c|c} \begin{matrix} \vdots \\ \cdots & 0 & \cdots \\ \vdots \end{matrix} & \begin{matrix} 0 & 1 \\ 1 & 0 & \vdots \\ \cdots & 0 & \cdots \\ \vdots \end{matrix} \\ \hline \begin{matrix} 0 & 1 \\ 1 & 0 & \vdots \\ \cdots & 0 & \cdots \\ \vdots \end{matrix} & \begin{matrix} \vdots \\ \cdots & 0 & \cdots \\ \vdots \end{matrix} \end{array} \right), ..., \quad b_{\frac{n(n+3)}{2}+2k} := \left(\begin{array}{c|c} \begin{matrix} \vdots \\ \cdots & 0 & \cdots \\ \vdots \end{matrix} & \begin{matrix} 0 & 0 \\ \vdots & 0 & 1 \\ \cdots & 0 & \cdots \\ 0 & 0 & \vdots \\ 0 & 1 \end{matrix} \\ \hline \begin{matrix} 0 & 0 \\ \vdots & 0 & 1 \\ \cdots & 0 & \cdots \\ 0 & 0 & \vdots \\ 0 & 1 \end{matrix} & \begin{matrix} \vdots \\ \cdots & 0 & \cdots \\ \vdots \end{matrix} \end{array} \right), ...$$

$$v_{\frac{n(n+3)}{2}+2} := \left(\begin{matrix} 0 & 2 & \\ 2 & 0 & \vdots \\ & \cdots & 0 & \cdots \\ & & \vdots \end{matrix} \right) + i \left(\begin{matrix} & \vdots & \\ \cdots & 0 & \cdots \\ & \vdots & \end{matrix} \right), ..., \quad v_{\frac{n(n+3)}{2}+2k} := \left(\begin{matrix} & & 0 & 0 \\ & \vdots & 0 & 2 \\ \cdots & 0 & \cdots \\ 0 & 0 & \vdots \\ 0 & 2 \end{matrix} \right) + i \left(\begin{matrix} & \vdots & \\ \cdots & 0 & \cdots \\ & \vdots & \end{matrix} \right), ...$$

Figure 32: The basis vectors $\{b_{\frac{n(n+3)}{2}+2}, ..., b_{n^2+n}\} \in \mathfrak{p}$ represent the off-diagonal part of the block matrix B in eq. 209. The $\frac{n^2-n}{2}$ corresponding vectors in $T_{iId}SH^n$ represent the off-diagonal part of the real part of that vector space. Note that the combination of all vectors $\{b_i\}_{i=1,..,n^2+n}$ allows to control independently real and imaginary part of the vector space $T_{iId}SH^n \cong Sym(n, \mathbb{C})$ which is spanned by the vectors $\{V_i\}_{i=1,..,n^2+n}$.

14.10.3 The exponential map via the Siegel ball model

The second path to calculate the exponential map in SH^n is to transfer the vectors to the Siegel ball model. The following figure examines the orange path in fig. 26 in greater detail. For the transfer to the ball model the push-forward $\left(i^{-1}\right)_*$ is needed. In the Siegel ball model the exponential map is in general also hard to find.

However, one can make a coordinate transformation in which the exponential map is simple. This coordinate transformation can be found via the Takagi factorization. The ball model is particularly well-suited to this approach, since the coordinate transformation has a very simple form due to the simple action of the isometry group G_0 on SB^n.
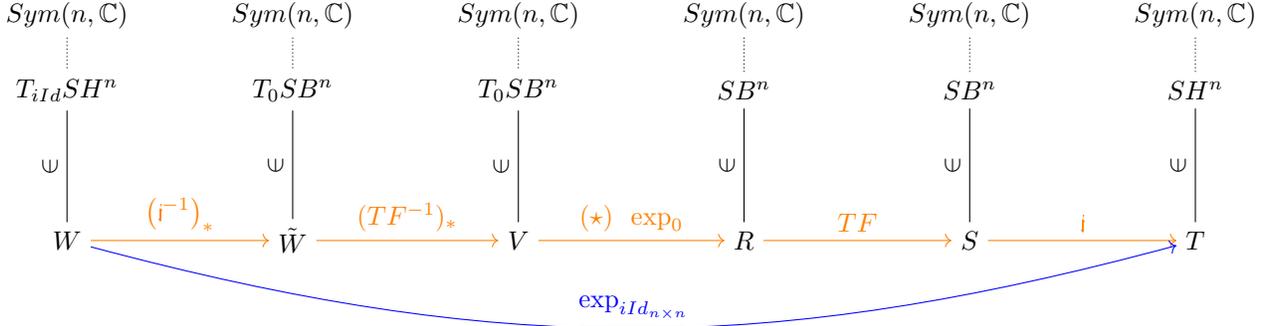
Figure 33: Diagram of the setup for the exponential map in the Siegel space as ball model. The arrow marked by $(\star)$ is trivial since the vector V corresponds to a diagonal matrix after the map TF^{-1} which uses the Takagi factorization. All matrices for this exponential map have the form $(n \times n)$ with complex entries.

The push-forward $\left(i^{-1}\right)_*$

The models SH^n and SB^n of the Siegel space are equivalent. This also means that their corresponding tangent bundles are closely related: As the vector space $T_{iId}SH^n$ has a representation as complex matrices, so does the vector space T_0SB^n. However, the corresponding isomorphism which takes the underlying models into account needs to be the push-forward map $(i^{-1})_*$, which is the derivative of i^{-1}.

To construct this map, one can use the decomposition of a generalized Möbius transform:

$$f(Z) = (AZ + B)(CZ + D)^{-1} = \left(f_5 \circ f_4 \circ f_3 \circ f_2 \circ f_1\right)(Z) \text{ with} \tag{210}$$

- $f_1(Z) = CZ$

- $f_2(Z) = Z + D$

- $f_3(Z) = Z^{-1}$ [41]

- $f_4(Z) = (B - AC^{-1}D)Z$ [42]

- $f_5(Z) = Z + AC^{-1}$.

This decomposition is similar in spirit to the well-known decomposition of Möbius transformations. Due to the non-commutativity of matrices, the decomposition is a little bit more complicated [43]

The derivative of an action as in eq. 210 can be found via the chain rule as:

$$f_*\Big|_Z(V) = \left(B - AC^{-1}D\right)\left(-(CZ + D)^{-1}CV(CZ + D)^{-1}\right). \tag{211}$$

For $f = i^{-1}$ with $g_{i^{-1}} = \frac{1}{2i}\begin{pmatrix} Id_n & -iId_n \\ Id_n & iId_n \end{pmatrix}$ and the evaluation point $Z = iId$ the resulting map is:

$$\left(i^{-1}\right)_*\Big|_{Z=iId} : T_{iId}SH^n \to T_0SB^n \tag{212}$$

$$V \mapsto -\frac{i}{2}V. \tag{213}$$

---

[41] This inversion in well-defined (i.e. $(CZ + D)^{-1}$ exists) because the action of $Sp(2n,\mathbb{R})$ on SH^n is well-defined [Siegel, 1943, II.5].

[42] In general, the matrix C need not be invertible [Siegel, 1943, II,5], although the matrix $(CZ + D)$ is invertible (for C, D block matrices from a symplectic matrix, $Z \in SH^n$). As this decomposition of a symplectic transform will be evaluated at $C = Id$, the inverse of C exists in a neighbourhood of C, as the determinant is a continuous function on $M(n,\mathbb{R})$.

[43] Note that the maps f_i are no isometries.

The Takagi factorization

The Takagi factorization (also known as Autonne-Takagi factorization due to LÉON AUTONNE and TEJI TAK-AGI) allows to decompose any complex symmetric matrix as $V = U\Lambda U^T$ with $\Lambda \geq 0$ real, diagonal and U unitary matrix[44].

The vector $\left(\mathsf{i}^{-1}\right)_*(V)$ can be represented as matrix. However, the Riemann exponential map on SB^n is still difficult. If the matrix V had diagonal form, the Riemann exponential map could be easily computed as combination of the single Riemann exponential map in every entry: The jth diagonal entry represents a vector in the jth subspace $\left(H^2\right)^n \subset SB^n$.

This can be done via a coordinate change: One needs to find an isometry which fixes the origin of SB^n and whose derivative maps the vector V to a diagonal matrix. This transformation (labelled TF^{-1}) can be found with the Takagi factorization. Its derivative is labeled $(TF^{-1})_*$. In this coordinate system one conducts the Riemann exponential map. After that, the resulting point is transformed back to the original coordinate system via TF.

The isotropy group G_0

Since the coordinate change can be understood as a rotation, one has to investigate those isometries, which fix the point 0, i.e. the isotropy group at 0. Since the isotropy group at $iId \in SH^n$ is $SpO(2n, \mathbb{R})$, the isometry group at $0 = \mathsf{i}^{-1}(iId)$ is $\left(\mathsf{i} \circ SpO(2n, \mathbb{R}) \circ \mathsf{i}^{-1}\right)$. This yields the isotropy group G_0 as:

$$G_0 = \begin{pmatrix} U & 0 \\ 0 & \bar{U} \end{pmatrix} \cong U(n, \mathbb{C}) \text{ with } \bar{U} \text{ as the complex conjugate of the unitary matrix } U. \qquad (214)$$

Its action on SB^n is in particular simple as

$$\begin{aligned} G_0 : U(n, \mathbb{C}) \times SB^n &\to SB^n \\ (U, P) &\mapsto UPU^T \text{ as matrix multiplication with } P \in SB^n, \ U \in G_0. \end{aligned} \qquad (215)$$

It is immediate that G_0 fixes 0. Additionally, the push-forward of an isotropy group action maps a vector $V \in T_0 SB^n$ as follows:

$$\begin{aligned} \left(G_0\right)_*\Big|_{Z=0} : T_0 SB^n &\to T_0 SB^n \\ V &\mapsto UVU^T \text{ as matrix multiplication.} \end{aligned} \qquad (216)$$

To find the coordinate transformation, one thus has to find the Takagi factorization of the vector V as $V = U\Lambda U^T$ with Λ as diagonal matrix. The coordinate transformation is then its inverse, i.e.

$$\begin{aligned} \left(TF^{-1}\right)_* : T_0 SB^n &\to T_0 SB^n \\ V = U\Lambda U^T &\mapsto U^\dagger(V)\bar{U} = U^\dagger\left(U\Lambda U^T\right)\bar{U} = \Lambda \end{aligned} \qquad (217)$$

After this coordinate transformation the Riemann exponential map is easy to perform:

---

[44]The Takagi factorization is a special case of the well-known singular value decomposition applied on a complex, symmetric square matrix. In fact, AUTONNE discovered this decomposition as special case of the singular value decomposition.

> **The Riemann exponential in the ball model**
> The matrix Λ above is a real diagonal matrix. It is thus sufficient to consider real diagonal matrices. As the vector Λ completely lies in the subspace $(H^2)^n \subset SB^n$, the exponential map decomposes into n exponential maps on each subspace H^2 in the Poincaré disk model.
> As Λ is real, the vectors in each Poincaré disk are vectors only pointing in real direction. The Riemann exponential map for a real vector $v \in H^2$ is:
>
> $$\exp_0(v) = \tanh(v) \in \mathbb{R} \subset H^2. \tag{218}$$
>
> This yields:
> $$\exp_0(\Lambda) = \operatorname{diag}\big(\tanh(v_1), ..., \tanh(v_n)\big) \subset SB^n. \tag{219}$$

To transform the resulting point back to the original coordinate system, one applies

$$
\begin{aligned}
TF : SB^n &\to SB^n \\
R &\mapsto URU^T.
\end{aligned}
\tag{220}
$$

Note that the maps TF and $(TF^{-1})_*$ in principle act on different spaces, namely the Siegel space and its tangent space. This procedure is only possible since the action of the isotropy group G_0 and its push-forward coincide as seen in eq. 215 and eq. 216.
To obtain the resulting point on the upper half-space model, one just sends the point on SB^n to SH^n via the inverse of the generalized Cayley transform $\mathfrak{i}^{-1}$ from eq. 168.

14.11   The Riemannian distance on SH^n

The distance between two points in SH^n is in general defined as the length of the shortest path connecting these points. This notion is not very helpful to calculate the distance explicitly. We make use of section 14.10 which shows that it is possible to make such a coordinate change that two points Q and R both lie in the same subspace $(H^2)^n$. Since $(H^2)^n$ is a totally geodesic submanifold, the distance within this submanifold coincides with the distance in SH^n. As $(H^2)^n$ is just a Cartesian product of hyperbolic spaces, it is in particular easy to calculate the distance as:

$$d_{(H^2)^n}(Q, R) = \sqrt{\sum_i^n d_i(Q_i, R_i)^2}, \quad Q_i, R_i \text{ coordinates in the space } H_i^2. \tag{221}$$

The procedure is hence as follows: Consider Q and R in the upper half-space model of SH^n. First, one applies a transformation $\theta(g^{-1})$ which maps Q to the base point iId_n and R to a different point in SH^n. Afterwards, both points are mapped to the ball model via $\mathfrak{i}^{-1}$. Q is hence mapped to the origin, while R is mapped to a point in SB^n. Finally, one applies the Takagi factorization $(TF)^{-1}$ for the coordinate form of the position R now has. This coordinate change via the Takagi factorization thus moves R to the subspace $(H^2)^n$ and leaves the origin fixed. The distance can now be calculated within $(H^2)^n$. The procedure is shown below:

Figure 34: Diagram to illustrate the procedure to compute the distance between the points Q and R in the Siegel upper half-space model.

Now we want to find an explicit formula for the distance in terms of R and Q. As shown in the diagram above, it is possible to compute the distance after the last step within the Siegel ball model, since all transformations

68

applied before are isometries, i.e. do not alter the distance. Because one point (namely Q in our notation) is mapped to the origin in SB^n, we can use the following formula for the distance within the 2-dimensional Poincaré disk:

$$d_{\text{Poincaré}}(0, (r, \phi)^T) = 2 \ \text{arctanh}(r) = \ln\left(\frac{1+r}{1-r}\right). \tag{222}$$

Plugging this in eq. 221 for the space $(H^2)^n$ one obtains:

$$d_{(H^2)^n}(Q, R) = \sqrt{\sum_i^n \ln^2\left(\frac{1+r_i}{1-r_i}\right)} \tag{223}$$

with the radii r_i of the coordinates in n Poincaré disks.

To use this formula, it is hence only necessary to find $(TF)^{-1}\big(\mathsf{i}^{-1}\big(\theta(g^{-1})R\big)\big)$[45]. This is the diagonal matrix obtained by a coordinate change via the Takagi factorization.

To compute $\mathsf{i}^{-1}\big(\theta(g^{-1})R\big)$ one makes use of formula eq.201 and obtains:

$$
\begin{aligned}
\mathsf{i}^{-1} \circ \theta(g^{-1})(R) &= \theta\left(\frac{1}{2i}\begin{pmatrix} Id_n & -iId_n \\ Id_n & iId_n \end{pmatrix} \cdot \begin{pmatrix} \sqrt{Y_Q^{-1}} & -\sqrt{Y_Q^{-1}}X_Q \\ 0 & \sqrt{Y_Q} \end{pmatrix}\right)(R) \\
&= \theta\left(\frac{1}{2i}\begin{pmatrix} \sqrt{Y_Q^{-1}} & -\sqrt{Y_Q^{-1}}X_Q - i\sqrt{Y_Q} \\ \sqrt{Y_Q^{-1}} & -\sqrt{Y_Q^{-1}}X_Q + i\sqrt{Y_Q} \end{pmatrix}\right)(R) \\
&\overset{\star}{=} \theta\left(\begin{pmatrix} Id_n & -X_Q - iY_Q \\ Id_n & -X_Q + iY_Q \end{pmatrix}\right)(R) = \theta\left(\begin{pmatrix} Id_n & -Q \\ Id_n & -\overline{Q} \end{pmatrix}\right)(R) \\
&= (R - Q)(R - \overline{Q})^{-1}, \quad Q = X_Q + iY_Q \in SH^n
\end{aligned}
\tag{224}
$$

where the equality marked with "$\star$" is because overall factors cancel in the action.

Recall that the Takagi factorization $V = U\Lambda U^T$ is a decomposition of the matrix V into a product of matrices with $\Lambda = \text{diag}(\sqrt{\lambda_1}, ..., \sqrt{\lambda_n})$ with λ_i the eigenvalues of $V\overline{V}$ which are real and exist since $V\overline{V}$ is hermitian. From Λ one can read off the radii of the n Poincaré disks of $(H^2)^n$ as diagonal elements. It is hence sufficient to calculate all eigenvalues of the matrix

$$\mathsf{i}^{-1}\theta(g^{-1})(R) \cdot \overline{\mathsf{i}^{-1}\theta(g^{-1})(R)}. \tag{225}$$

Plugging the result eq. 224 into eq. 225 leads to the formula:

$$d_{SH^n}(Q, R) = \sqrt{\sum_i^n \ln^2\left(\frac{1+\sqrt{\lambda_i}}{1-\sqrt{\lambda_i}}\right)} \tag{226}$$

where λ_i are the eigenvalues of

$$(R - Q)(R - \overline{Q})^{-1} \cdot \overline{(R - Q)(R - \overline{Q})^{-1}}. \tag{227}$$

Distance in the coset picture

Since every point in the Siegel space corresponds in a one-to-one fashion to an equivalence class in the coset formulation $Sp(2n, \mathbb{R})/SpO(2n, \mathbb{R})$, it is sensible to consider computing the distance of two points in the Siegel space in the coset form:

Let the Q, R be points in the Siegel space and $g_Q, g_R \in Sp(2n, \mathbb{R})$ such that $\theta(g_Q)(iId) = Q$ and for R accordingly. Then one can compute the distance between Q and R as [Freitas, 1999, Corollary 2.3.2]:

$$d(Q, R)^2 = \sum_{j=1}^{2n} \left| \ln \sigma_j(g_Q^{-1}g_R) \right|^2, \quad \sigma_j \text{ singular values.} \tag{228}$$

The equivalence of the two formulations can be checked easily when considering that Q and R can always be mapped via a coordinate change to iId and $i\text{diag}(\lambda_1, ...\lambda_n)$, respectively.

---

[45]Note that this term is also dependent on the original Q via $\theta(g^{-1})Q = iId$.

14.12   The sectional curvature of the Siegel space

So far, submanifolds of SH^n and their sectional curvatures have been explored. The sectional curvature of the whole space however has not yet been investigated. Instead of calculating the sectional curvature via the Riemann curvature tensor in eq. 28, one can use the correspondence of the tangent space $T_{iId}SH^n$ with the subset $\mathfrak{p} \subset \mathfrak{g}$ of the Lie algebra of the isometry group as pointed out in section 7. This allows to compute the Riemann tensor via the commutator on $\mathfrak{p}$. With use of the Killing form on $\mathfrak{g}$, it is possible to compute the sectional curvature quite easily.

The Killing form

Any Lie group G is naturally equipped with a so-called Killing form $B(\cdot, \cdot)$ on its Lie algebra $\mathfrak{g}$. The Killing form is a symmetric bilinear form using the *adjoint map*:

$$\mathrm{ad} : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$$
$$(x, y) \mapsto [x, y] =: \mathrm{ad}_x(y). \tag{229}$$

Obviously, the map $\mathrm{ad}(y) = [\cdot, y]$ is an element of $\mathcal{M}aps(\mathfrak{g}, \mathfrak{g})$ which is linear in y. Since the trace of a linear map maps to a scalar, the Killing form

$$B(x, y) = \mathrm{tr}\big(\mathrm{ad}(x) \circ \mathrm{ad}(y)\big), \quad x, y \in \mathfrak{g} \tag{230}$$

actually is a bilinear form on $\mathfrak{g}$. Its symmetry follows from the trace being cyclic.

It is possible to calcuate the Killing form explicitly for a given Lie algebra by using the fact that $\mathfrak{g}$ is a vector space. This enables to choose a basis of $\mathfrak{g}$ and represent the concatenation of ad-maps with respect to that basis. For this map one can then calculate the trace. Due to the linearity of the Killing form, this suffices since any elements x, y can be decomposed into the chosen basis. As the Riemannian metric and the Killing form are both K-invariant, they can only differ by a factor λ.

The Killing form gives an important insight into the sectional curvature of a space, since:

$$g_P(R(U, V)V, U) = \lambda^{-1} B(\upsilon^{-1}([U, V], \Upsilon^{-1}([U, V]), \quad U, V \in T_P M. \tag{231}$$

The parameter λ hence determines if the sectional curvature is nonnegative ($\lambda < 0$), nonpositive ($\lambda > 0$) or zero ($\lambda = 0$). The left hand side of the equation above is in fact the formula for the sectional curvature eq. 28 without the denominator. However, due to the Schwarz-inequality, the denominator cannot change the sign of λ, because it is nonnegative.

The statement above can be shown using $B([x, y], z) = -B(y, [x, z])$ and eq. 73:

$$\lambda g_P(R(U, V)V, U) = B\big(\Upsilon^{-1}(U), \Upsilon^{-1}(R(U, V)V, U)\big) = B\big(\Upsilon^{-1}(U), [\Upsilon^{-1}(V), [\Upsilon^{-1}(U), \Upsilon^{-1}(V)]]\big)$$
$$= B\big([\Upsilon^{-1}(U), \Upsilon^{-1}(V)], [\Upsilon^{-1}(U), \Upsilon^{-1}(V)]\big) = B\big(\Upsilon^{-1}([U, V]), \Upsilon^{-1}([U, V])\big). \tag{232}$$

This can be applied for the Siegel space SH^n. Its Killing form on $\mathfrak{sp}(2n, \mathbb{R})$ can be found as:

$$B(V, W) = (2n + 2)\mathrm{tr}(VW), V, W \in \mathfrak{sp}(2m, \mathbb{R}). \tag{233}$$

Now again, the explicit anti-isomorphism Υ which has been dealt with in section 14.10 relating $\mathfrak{p}$ and the space of Killing vector fields on SH^n comes into play. To relate the two scalar products to each other, it is sufficient to consider the tangent space at $P = iId$. After all, the map Υ has been computed for $T_{iId}SH^n \cong \mathfrak{p}$. At iId, the Riemann metric tensor from eq. 183 reduces to:

$$\mathrm{tr}(iId \cdot V \cdot iId \cdot W) = -\mathrm{tr}(VW), \quad V, W \in T_{iId}SH^n. \tag{234}$$

The relation between the Killing form on $\mathfrak{p}$ and the metric tensor on SH^n is hence:

$$-\mathrm{tr}(V_i V_j) = \lambda(2n + 2)\mathrm{tr}\big(\Upsilon^{-1}(V_i) \cdot \Upsilon^{-1}(V_j)\big) = \lambda(2n + 2)\mathrm{tr}(b_i, b_j), \quad V_i \ T_{iId}SH^n, \ b_i \in \mathfrak{p}. \tag{235}$$

One can compute this explicitly for all combinations b_i, b_j the trace of its product. This can just be read off from the matrix relations in section 14.10. This yields:

$$\text{tr}(b_i, b_j) = \frac{1}{2}\text{tr}(V_i, V_j). \tag{236}$$

Plugging this result into eq. 235 yields:

$$-\text{tr}(V_i V_j) = \lambda(2n+2)\frac{1}{2}\text{tr}(V_i, V_j)$$
$$\Rightarrow \lambda = \frac{1}{n+1} > 0. \tag{237}$$

This ultimately shows that SH^n is a space of nonpositive curvature which is called *of noncompact type*. Of course, all sectional curvatures are nonpositive. SH^n is diffeomorphic to $\mathbb{R}^{n^2+n}$, which means that it has the same "shape" as $\mathbb{R}^{n^2+n}$ but of course not its geometry. It just extends to infinty in $n^2 + n$ directions.

14.13   Implementation considerations for computations on the Siegel space

Graph embeddings in a manifold are mostly performed using the Riemann gradient descent method. The structure of such an algorithm is carried out in detail in section 16, so this section is kept in a synthetic formulation.
The Riemann gradient descent method requires the three key operations:

- computing the distance,

- computing the Riemann gradient from the Euclidean gradient,

- computing the Riemann exponential map.

This is why these operations have been treated explicitly above.
There is another constraint to the implementation, namely that the computation of the distance has to be machine-differentiable. To make use of an automatic differentiation (such as *autograd*), all steps to compute the distance have to be machine-differentiable (i.e. all operations need to have an implemented machine differentiation).
The conversion to the Riemann gradient and the computation of the Riemann exponential map do not have to be machine-differentiable.

The model of choice for the Siegel space is the upper half-space model as coordinates within the subspace $(H^2)^n$ are complex numbers with positive imaginary part. This comes in handy as the machine *float* number system is very dense close to zero which improves precision. Additionally, the values of the coordinates in the upper half-space model can be very large which also supports relative precision.
This section proposes an implementation procedure orientated at the off-the-shelf *pytorch* framework:

14.13.1   Computing the distance and Riemannian gradient

The formula eq. 226 for the distance in the Siegel space requires the computation of n eigenvalues of a complex[46] $(n \times n)$-matrix or equivalently the computation of the singular values of a complex $(n \times n)$-matrix. In fact, calculating the eigenvalues of the full matrix eq. 227 is equivalent to computing the singular values of the first part of eq. 227. Since it is computationally difficult to compute eigenvalues of a complex matrix *machine differentiably*, it is sensible to consider computing the distance in the coset picture via eq. 228: This method requires computing singular values of a real $(2n \times 2n)$-matrix which is easier to implement differentiably in *pytorch*.
To compute the matrices $g_Q, g_R \in Sp(2n, \mathbb{R})$ from eq. 228, one can make use of eq. 201 which is an explicit formula in terms of the matrix representation of Q and R. The matrix $g_Q^{-1} g_R$ becomes:

$$g_Q^{-1} g_R = \begin{pmatrix} \sqrt{Y_Q^{-1} Y_R} & \sqrt{Y_Q^{-1}} X_R \sqrt{Y_R^{-1}} - \sqrt{Y_Q^{-1}} X_Q \sqrt{Y_R^{-1}} \\ 0 & \sqrt{Y_Q Y_R^{-1}} \end{pmatrix} \quad \text{with } Z = X_Z + iY_Z \in Sp(2n, \mathbb{R}). \tag{238}$$

---

[46]The matrix need not be real, since $(X+iY)\overline{X-iY}$ is in general not real since X and Y do not commute.

This requires computing the inverse and the square root of real symmetric positive-definite $(n \times n)$-matrices which is feasible[47] within the *pytorch* framework. As mentioned, the singular value computation of a real matrix is a standard computation and already implemented within *pytorch*.

The algebraic operations as taking the square or the logarithm to finally computing the distance are straight-forward and do not require any further discussion.

It is hence possible to compute the gradient with respect to the coordinates of a point $Q \in SH^n$. This gradient however is computed with respect to the Euclidean gradient, since the automatic differentiation is carried out within the space $\mathbb{R}^{2n \cdot 2n}$.

The Euclidean gradient can be converted to the true Riemann gradient via formula 186. This matrix computation is straight-forward. One does not have to take care of the implementation of the derivative as it is just a conversion of a vector to the same manifold governed by a non-Euclidean metric.

14.13.2 Computing the Riemann exponential map

Two different procedures have been explained in detail in the section above. The computation of the Riemann exponential map does not have to be machine-differentiable. One thus is not constrained by the implemented *pytorch* framework. In fact, the action of the symplectic group on the Siegel space has to be computed with use of a complex matrix inversion[48]. Additionally, the Takagi decomposition is not implemented on *pytorch* yet, but there are algorithms at hand such as [Chebotarev and Teretenkov, 2014] or [Bunse-Gerstner and Gragg, 1988]. The matrix exponential which is needed for the exponential map on the Lie algebra, can be computed with several algorithms such as those summarized in [Moler and Van Loan, 2003].

There are hence no fundamental obstacles for the implementation of the gradient descent method on the Siegel space via *pytorch*.

Looking at the diagram below which shows the proposed procedures to compute the Riemann exponential map, one may note that there are shortcuts which may facilitate the computation. It turns out however that this is not the case:

- Let us have a look at the orange path first. The first shortcut does not save any computation, since the maps $\left(\theta(g^{-1})\right)_*$ (eq. 205) and $(i^{-1})_*$ (eq. 212) are very easy to compute whereas the derivative of the concatenation of these maps is computationally intensive.

- The map $\theta(g)$ is easy to compute, since $\theta(g)(Z) = \sqrt{Y_Q} Z \sqrt{Y_Q} - X_Q$ in the notation of eq. 201. The square root and inverse have been computed already for the map $\left(\theta(g^{-1})\right)_*$. A concatenation of the maps i and $\theta(g)$ is in fact computationally more demanding as it involves more matrix multiplications. The same holds for the shortcut in the red path via the coset formulation.

Speed and accuracy difference between these paths depend on the chosen implementations of the matrix exponential and the Takagi factorization and have to be figured out experimentally.

---

[47]The machine-differentiable matrix square root computation has been added very recently.

[48]It is possible to transfer the complex matrix inversion to several real matrix inversions, but this is computationally more expensive [Ehrlich, 1970].
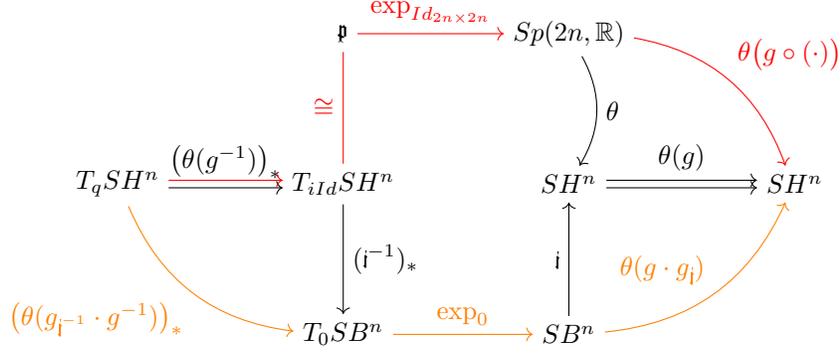
Figure 35: Diagram of the setup for the Riemann exponential map in the Siegel space. The paths colored in red and orange show procedures via the Lie group quotient model and the Siegel ball model, respectively. The "shortcuts" turn out to be computationally more difficult.

15 Data representation capabilities of symmetric spaces

A focus on representing data in spaces of mixed curvature lies on interpretability of the embedding. A faithful embedding should hence represent the graph distance of any two vertices by the distance of the embedded vertices in the embedding space.

A big advantage of spaces of mixed curvature is that subspaces of certain curvature are more suited to incorporate certain concepts (e.g. hyperbolic subspaces representing hierarchy, spherical subspaces representing loops and clusters): Vertices which can be distinguished by a difference within one concept should be positioned in the same subspace with different coordinates. Because the distance within that subspace should represent the different role within that concept, it is necessary that this subspace is also totally geodesic, i.e. the distance within the whole embedding space is equal to the distance within the subspace representing the conceptual difference.

As illustrated in fig. 18, there are in general lots of submanifolds which are not totally geodesic. For instance, it is not very promising to embed a spherical concept in Euclidean space, although the sphere can be embedded within Euclidean space, because this submanifold carries the Riemannian flat structure and is hence not positively curved.

To make use of the properties of mixed-curvature spaces, it has been proposed to represent data in Cartesian products [Gu et al., 2019]. As described in section 13.6, these spaces do indeed possess a variety of sectional curvature. It should be recalled, that the sectional curvature is a local statement, i.e. only in a tiny neighbourhood, the subspace resembles a surface of that sectional curvature. Consequently, the totally geodesic submanifolds of product spaces are not very versatile and hence limit the embedding capability.

In contrast, with the example of the Siegel space, it has been shown that irreducible symmetric spaces can possess a variety of totally geodesic submanifolds. In particular, the aforementioned product spaces are contained within the class of symmetric spaces.

Certainly, with the same number of dimensions, more concepts can be represented due to versatility and quantity of totally geodesic submanifolds[49]. In that sense, irreducible symmetric spaces possess superior data representation capabilities.

On the downside, these spaces are in general more difficult to implement as shown in section 14.13. With the growing computational power, it is however reasonable to expect that this disadvantage can be overcome.

---

[49]E.g. the rank of a space is the maximal dimension of a Euclidean subspace and hence the maximal dimension of grid-like concepts which can be represented faithfully in that manifold.

Graph embedding in symmetric spaces

Overview

A meaningful network embedding reflects properties of the network by the layout in the embedding space, i.e. the structure of the network becomes accessible and combinatorial problems translate to geometric problems.

The idea to use curved spaces to embed into has gained attention due to analytical embeddings in hyperbolic space as in the seminal paper [Kleinberg, 2007] or in [Sarkar, 2011]. However, these analytical embeddings are carried out with perfect trees and can hence not be applied easily to more complicated networks.

Using a numerical approach, there are many algorithms which map networks to an Euclidean embedding space. Notable are the *Deep Walk* algorithm [Perozzi et al., 2014] or the *LLE* algorithm [Roweis and Saul, 2000]. However, they cannot be adapted easily to embed a network in a general (curved) manifold.

It has been tried to embed a network in hyperbolic space as proposed in [Boguñá et al., 2010] or [De Sa et al., 2018]. However, this is not versatile enough for large datasets, complex graph structure and complicated embedding space. The problem of network embedding is hence widely considered as an optimization problem over the embedding space which can be addressed with state-of-the-art numerical algorithms.

16   The gradient descent optimization

A widely used optimization algorithm is the so-called *gradient descent method*. This optimization approach has as big advantage that it can be modified easily to fit to almost any underlying manifold. It is neither the most elegant nor it is suited best to a specific manifold, but its flexibility is a key feature. The gradient descent algorithm has become much more powerful over the last years, mostly due to the advances in machine learning, which extensively use the gradient descent method for all kinds of tasks. Using the machine learning framework for this optimization thus promises a fast program execution.

Simply put, the gradient descent algorithm iteratively tries to modify the embedding of the graph in such a way that a scalar function $\mathcal{L}$ is minimized. This scalar function $\mathcal{L}$ is called *loss function* and is a key part of the algorithm. It determines the criteria according to which the embedding is optimized. There are different objectives such as:

- Minimizing the stretch between the graph distance and the embedding distance as in [Gu et al., 2019]

- Forming clusters of the vertices in the embedding space as in [Nickel and Kiela, 2017]

- Finding an embedding which enables a stable routing algorithm as in [Boguñá et al., 2010].

After every epoch, the gradient of the scalar function is calculated with respect to the coordinates of the vertices. To minimize the loss function, one then slightly moves the vertices in such a way that the value of the loss function is diminished. Since the loss function is dependent on the coordinates of the points in the embedding space, the positive gradient shows for every point the direction in which that point has to be moved to increase the loss function. By following the negative gradient, the value of the loss function is hence diminished.

16.1   Machine differentiability

The update of the embedding is performed using of the gradient of the loss function. The loss function is in general an algebraic expression dependent on the mutual distances of all vertices. The gradient of this function is hence cumbersome to compute symbolically. Additionally, the loss function may vary (as described in section 16.4) to improve computational speed. Thanks to numerical differentiation (such as *autograd*) the gradient can be constructed on the fly while computing the loss function. This is in principle done by multiple application

of the chain rule. The symbolic derivatives of each block (such as the derivative of the distance function or the algebraic expression of the loss function) are then merged to compute the gradient with respect to the whole embedding.

The computation of the distance thus has to be computed in such a way that it is machine-differentiable. This is not the case by default as stressed in section 14.13.1 for the Siegel space: The distance computation via the eigenvalues of a complex matrix is (not yet) machine differentiable.

16.2 Gradient descent algorithm for graph embedding
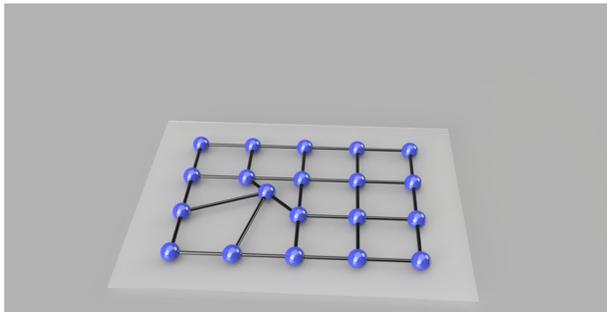
(a) Initial situation. An embedding in the plane with one (obviously) distorted vertex.
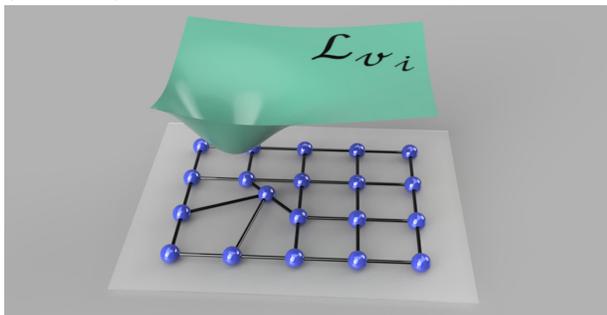
(b) Scalar loss function with respect to the distorted vertex at θ_i.
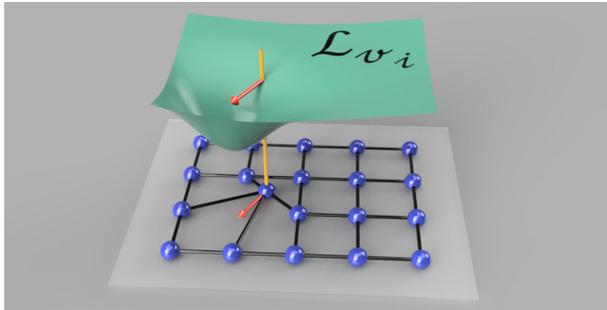
(c) The negative gradient of the loss function is symbolized by a red arrow.
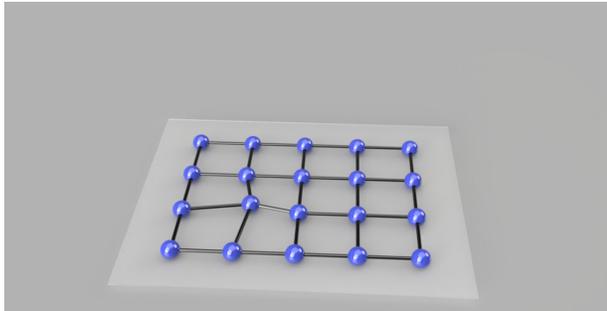
(d) Embedding after the application of a step in direction of the negative gradient.

Figure 36: Illustration of the gradient descent algorithm for a graph embedding.

The gradient descent algorithm is an algorithm to step-wise improve the existing embedding. One can assume that there is already a (possible very bad) embedding to start with[a]. This situation is depicted in fig. 36a.

The algorithm can be split in three parts: Calculating the loss function, computing the negative gradient and updating the embedding. The pictures on the left illustrate these steps with respect to one vertex only[b].

An embedding of a graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ with n vertices (i.e. $\#\mathcal{V} = n$) can be represented as $\{\theta_{i,t}\}_{v_i \in \mathcal{V}}$ where the $\theta_{i,t}$ are d-dimensional coordinates of the embedding manifold at iteration step t.

The procedure of the algorithm is:

1. The scalar loss function with respect to vertex v_i is calculated; it is denoted by $\mathcal{L}_{v_i}$[c]. In the example, the loss function depicted in fig. 36b is chosen to yield an isometric embedding. It is minimized if all lengths of the connecting edges are equal, i.e. if the vertices lie on a grid.

2. Because the loss function is a scalar function, one can compute the gradient of that function at evaluation point θ_i of the vertex v_i as:

$$\nabla \mathcal{L}_{v_i}(\theta_{i,t}). \qquad (239)$$

As the loss function should be minimized, reversing the sign and hence the direction of the vector yields the negative gradient. Note that the gradient is a tangent vector of the embedding manifold, not a tangent vector of the surface formed by the loss function. This can be seen in fig. 36c, as the arrow is tangent to the embedding manifold, but not tangent to the green surface.

3. The coordinates $\theta_{i,t}$ of the vertex v_i are updated by moving the coordinates in direction of the negative gradient:

$$\theta_{i,t+1} = \theta_{i,t} - \eta \nabla \mathcal{L}_{v_i}(\theta_{i,t}), \eta \in \mathbb{R}^+. \qquad (240)$$

The parameter η is the so-called *learning rate*. This scaling (usually much smaller than 1) is introduced to prevent overshooting.

---

[a]In fact, the initial embedding is often chosen to be a random embeddings, i.e. random coordinates are assigned to every vertex.

[b]A so-called "batch" version of the algorithm alters the coordinates of all vertices at once by parallel processing. For simplicity, let us consider one iteration at one vertex only, since the full algorithm is easy to infer.

[c]An algebraic expression (usually the sum) of the set $\{\mathcal{L}_{v_i}\}_{v \in \mathcal{V}}$ is referred to as *the* loss function $\mathcal{L}$. It is a scalar function of the whole embedding $\{\theta_{i,t}\}_{v_i \in \mathcal{V}}$. In the example, only one term - namely $\mathcal{L}_{v_i}$- of the complete loss function is considered.

16.3 Adapting the algorithm to curved spaces

The gradient descent method is commonly used for Euclidean spaces of high dimension. Using this method for a general Riemannian manifold requires some adaptations due to the underlying metric tensor of the embedding manifold. The groundwork of these adaptations has been carried out by [Bonnabel, 2013].

The coordinates $\{\theta_{i,t}\}_{v_i \in \mathcal{V}}$ are represented as "ordinary" numbers, but operations such as computing the gradient or the addition of points are different. One can carry them out with aid of the metric tensor, which can be thought of as measure how the true geometry differs from Euclidean geometry [50].

The ordinary gradient descent algorithm needs to be altered at three points. The resulting algorithm is called Riemannian gradient descent:

1. Calculation of the loss function:
   The loss function is dependent on the distances between embedded vertices. This distance $d(\theta_i, \theta_j)$ has to be calculated within the embedding manifold. Because of the distortion, one cannot simply compute the norm of the difference between the coordinates assigned to the vertices as in Euclidean space. In general, the distance is defined as the minimum of all line integral along a geodesic which connect θ_i and θ_j which is an optimization problem on its own. One thus needs to find an expression to calculate the distance right from their coordinates without having to find the shortest path.

2. Calculating the gradient:
   The loss function $\mathcal{L}_{v_i}$ is a scalar function living on a the embedding manifold, since the embedded vertices live in that manifold. The gradient of this function is hence different from the Euclidean gradient. Since the metric tensor g can be understood as a conversion between the Riemannian manifold and Euclidean space, the Euclidean gradient (i.e. the standard gradient with respect to the coordinates) needs to be scaled with the metric tensor. This yields the "true" gradient, which an observer within the manifold would measure. In components this vector field becomes:

$$\text{grad}_{Rie}\mathcal{L}_{v_i} = g^{\alpha\beta}\partial_\beta\mathcal{L}_{v_i}\partial_\alpha,$$

whereas $\text{grad}_{Euc}\mathcal{L}_{v_i} = \delta^{\alpha\beta}\partial_\beta\mathcal{L}_{v_i}\partial_\alpha$ with δ as Kronecker delta.
   (241)

3. Subtracting the gradient from the previous embedding:
   In a general manifold, adding a vector to a point is not defined. This works in Euclidean space as the parallel transport is trivial, i.e. points and vectors are basically equivalent.
   The generalization of this addition/subtraction to a general manifold is the so-called Riemann exponential map[51]. It can be understood as following the direction of the vector for the length of the magnitude of this vector.

In conclusion, the formula for an update becomes:

$$\theta_{i,t+1} = \exp_{\theta_{i,t}}\left(-\eta\, g^{\alpha\beta}\partial_\beta\mathcal{L}_{v_i}(\theta_{i,t})\, \partial_\alpha\right).$$
   (242)

16.4 Stochastic Gradient Descent and other variants

The batch gradient descent as described above computes the full loss function $\mathcal{L}$ with respect to all vertices at the same time. This procedure is mathematically very sound but can become computationally very demanding for a large graph.

One can address this by approximating the loss function $\mathcal{L}$. The complete loss function is an algebraic expression of all loss functions $\{\mathcal{L}_{v_i}\}_{v \in \mathcal{V}}$. Computing the loss function with only a small subset of $\{\mathcal{L}_{v_i}\}_{v \in \mathcal{V}}$ results in a much faster computation of the gradient of the loss function.

To improve the speed of the algorithm even further, one can approximate the constituent functions $\mathcal{L}_{v_i}$:
As the function $\mathcal{L}_{v_i}$ is dependent on the distance to all other $n-1$ vertices, the computation of the gradient can become computationally expensive. To address that, one can compute an approximation of the complete

---

[50]More on that can be found in section 2.
[51]This map is introduced formally in section 2.7.

loss function $\mathcal{L}_{v_i}$ by only considering the distances to a subset of all other $n-1$ vertices. These few vertices are chosen randomly for every iteration. The computation of the corresponding gradient can be tremendously faster.

These two approximations both increase the speed by orders of magnitude and hence easily compensate the loss of accuracy per step by many more iterations. The algorithm is thus overall much faster as shown in [Gu et al., 2019] or [Nickel and Kiela, 2017].
There are also lots of other variants of the gradient descent algorithm which mostly vary the updating procedure such that the algorithm overcomes saddle points and local minima e.g. by dynamically changing the learning rate or introducing a momentum vector which is added to the gradient vector. A comprehensive overview is given in [Ruder, 2016].

17   Experiments

The theoretically superior data representation capabilities of symmetric spaces have been shown above. However, the interpretability of an embedding strongly depends on the algorithm's ability to distinguish different structures of a graph.
Before extending the embedding space to general symmetric spaces, it has to be made sure that the approach of existing algorithms is capable of preserving the structure of the datasets, i.e. structures have to be placed faithfully according to the curvature of the subspaces.
Otherwise, making sense of a representation which does not align graph structures according to the curvature of the respective subspaces looses its theoretical foundation.
In particular, test data (which theoretically fits well to the chosen embedding space) should be spread out according to the underlying curvature while preserving the overall structure. More precisely, the embedding should display the properties of the graph in order to make a distinction of different structures even possible.

Starting from the machine-learning code from [Nickel and Kiela, 2017] and [Nickel and Kiela, 2018] experiments are conducted whether the structure of data can be discovered by current algorithms.
To enable evaluation of experimental results, at first common measures of fidelity are revisited:

17.1   Measures of fidelity

To quantify the quality of the embedding, there are plenty of fidelity measures which capture different properties. The following measures are widely used:

- The average distortion D_{avg} displays the average stretch of the edges in the embedding spaces compared to the weight of the graph.

- The worst-case distortion D_{wc} is the product of maximal compression and maximal expansion of edges within the embedding space with respect to the graph metric.

- The mean average precision mAP indicates how well the local structure of the graph is maintained by examining neighborhoods within the embedding space.

17.1.1   The average distortion D_{avg}

The distortion of a pair of points $u, v \in \mathcal{V}$ for an embedding $\phi : \mathcal{G} \to M, \ v \in \mathcal{V} \mapsto \phi(v)$ is

$$\frac{|d_{\mathcal{G}}(u,v) - d(\phi(u), \phi(v))|}{d_{\mathcal{G}}(u,v)}, \tag{243}$$

where the distance measures $d_{\mathcal{G}}(\cdot, \cdot)$ and $d(\cdot, \cdot)$ are the graph distance and the distance in the embedding space M, respectively.
The distortion relates the graph distance with the distance in the embedding space. The average distortion D_{avg} is defined as the average of the distortion over all pairs of all points, i.e. these points are not necessarily connected by an edge. It is a measure which examines the embedding on all length-scales of the graph, i.e. the distance fidelity is not only examined for small patches of the graph, but globally.
If the embedding is not distorting at all, then $D_{avg} = 0$.

17.1.2 The worst-case distortion D_{wc}

The worst-case distortion is defined as

$$D_{wc} = \left(\max_{u,v \in \mathcal{V}, u \neq v} \frac{d\big(\phi(u), \phi(v)\big)}{d_{\mathcal{G}}(u,v)} \right) \cdot \left(\max_{u,v \in \mathcal{V}, u \neq v} \frac{d_{\mathcal{G}}(u,v)}{d\big(\phi(u), \phi(v)\big)} \right) \tag{244}$$

$$= \left(\max_{u,v \in \mathcal{V}, u \neq v} \frac{d\big(\phi(u), \phi(v)\big)}{d_{\mathcal{G}}(u,v)} \right) : \left(\min_{u,v \in \mathcal{V}, u \neq v} \frac{d\big(\phi(u), \phi(v)\big)}{d_{\mathcal{G}}(u,v)} \right). \tag{245}$$

This measure gives a very broad impression on the relative distortion of the embedding: It is a measure of the width of the distribution of relative distortion. A perfectly distance-preserving embedding yields the worst case distortion $D_{wc} = 1$. Note that outliers (e.g. only one vertex placed extremely badly) increase D_{wc} tremendously.

17.1.3 The mean average precision mAP

Consider the node $v \in \mathcal{V}$ and its neighborhood $N_v = \{u \in \mathcal{V} | d_{\mathcal{G}}(v,u) = 1\}$ which is the set of all vertices directly connected to v. It has the size $\deg(v)$.

Now consider the analogue neighbourhood in the embedding space M, namely $\overline{B_{\phi(v)}(\epsilon)}$ which is the standard closed ϵ-ball around $\phi(v)$. The smalles neighborhood containing the node $\phi(u)$ is labelled $B_{v,u}$.

The average precision of the embedding of the neighborhood of x is

$$\frac{1}{\deg(v)} \sum_{i}^{|N_v|} \frac{|N_v \cap B_{v,v_i}|}{|B_{x,v_i}|} \text{ and consequently}$$

$$\mathrm{mAP}(\mathcal{G}) = \frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} \frac{1}{\deg(v)} \sum_{i}^{|N_v|} \frac{|N_v \cap B_{v,v_i}|}{|B_{v,v_i}|}. \tag{246}$$

The mean average precision is a measure which examines the embedding only on small length-scales of the graph and is hence a local fidelity measure. If the structure of the graph is preserved perfectly in the embedding space, then $\mathrm{mAP} = 1$.

17.2 Datasets

The test datasets comprise a tree with branching factor four, a two-dimensional grid graph and the subset of mammals of the dataset *wordnet*. These are all unweighted graphs.

The tree and lattice graphs are artificial test data to examine if the algorithm correctly discovers the underlying hierarchical and grid-like structure correctly. These datasets are relatively small in order to easily analyze the properties of the embedded graph and enable visualization.

The mammals dataset however is not artificially designed to suit perfectly to an embedding manifold. It consists of word pairs which can be connected with "is a" relations such as (blue whale, whale). Due to that ordering it is highly hierarchical, but at the same time possesses loops such as: (white elephant, elephant), (elephant, mammal), (white elephant, mammal). This dataset hence examines the algorithm's capability to deal with different structures at once.

17.3 Outline of the chapter

Subsequently, the results of modified embedding algorithms are depicted, followed by a short qualitative evaluation. Quantitative results and a discussion are presented in sections 17.8 and 18.

The pictures are organized as follows: The embeddings in hyperbolic space are shown in the Poincaré ball model indicated by a light blue half-shell. The lattice is embedded in three-dimensional Euclidean space.

The edges are drawn with color corresponding to the distortion of that edge as in eq. 243. Note that the color bar is scaled to enable a good visualization of the distortion of all edges. On the upper right, a histogram shows the distribution of distortion of all edges of the graph. Below that, the distortion of all $(n^2 - n)/2$ possible pairs is depicted.

17.4 Unchanged embedding algorithm

In order to further improve the optimization algorithm, the properties of the starting algorithm by [Nickel and Kiela, 2017] are revisited:

Their loss function of an embedding ϕ is:

$$\mathcal{L}(\phi) = \sum_{(u,v)\in\mathcal{E}} \ln \exp d\big(\phi(u), \phi(v)\big) - \ln \Big(\sum_{u'\in\big(\mathcal{N}(u)\big)^c} \exp -d\big(\phi(u'), \phi(v)\big)\Big) \tag{247}$$

where the first term can be understood as moving connected points close to each other whereas the second term drives unconnected points away from each other.

The initial embedding to optimize is random, i.e. the coordinates of all vertices assigned by values within the domain $(-0.001, 0.001)$, independent of the embedding space.

17.4.1 Results following [Nickel and Kiela, 2017]:

The embedding capabilities of the aforementioned algorithm can be evaluated by considering the following figures:
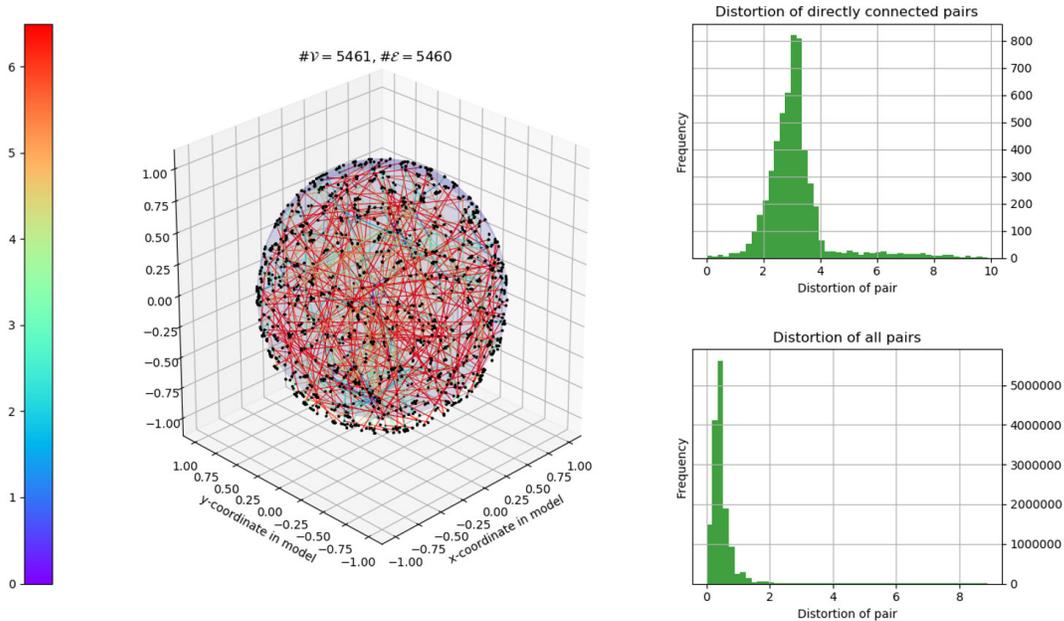
Figure 37: Tree with branching factor four in 3-dimensional Poincaré ball. The tree structure of the graph is not visible. The entanglement within the graph can be seen clearly by highly distorted edges crossing the whole model.
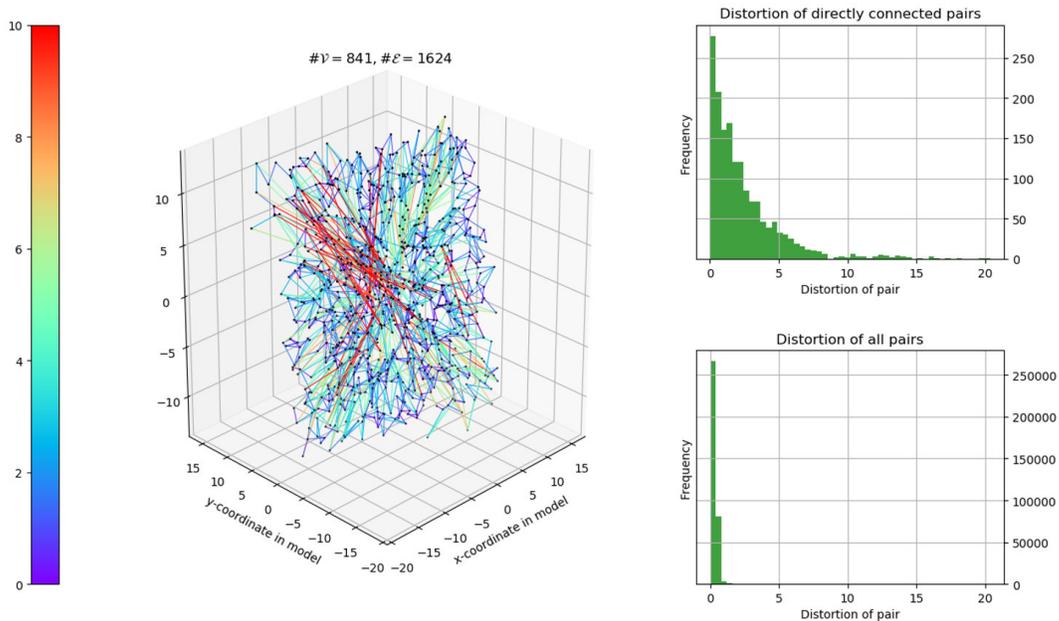
Figure 38: Lattice embedded in 3-dimensional Euclidean space. The intrinsic 2-dimensionality of the lattice can be surmised. The structure of the lattice is not recognizable.
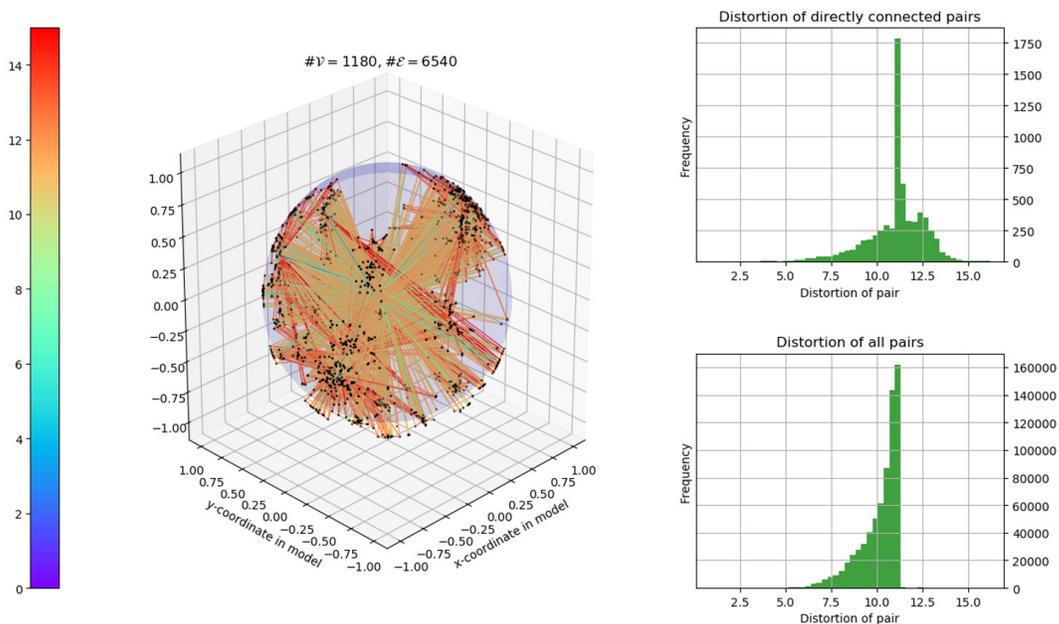
Figure 39: Mammals wordnet in the 3-dimensional Poincaré ball. The hierarchical structure is visible. Nevertheless, the local neighborhoods are not well preserved as visible by long crossing edges with high distortion.

81

Evaluation

Keeping in mind that the datasets are small test datasets, the performance of the algorithm as proposed by [Nickel and Kiela, 2017] is disappointing: The structure of the tree and lattice graphs is not discovered. The Mammals wordnet is shown to be hierarchical, but the embedding is highly distorting.

The loss function eq. 247 is designed to maximize mAP measure which in particular works well in hyperbolic space due to the exponential growth of the circumference of a circle.

Since minimizing the distortion is not considered by the algorithm, the high distortion is explainable. This algorithm hence does not aim at making use of the property of hyperbolic space to yield very small distortion for trees as shown by [Sarkar, 2011].

17.5   Initial embedding using the Fruchterman-Reingold algorithm

As seen in figures 37, 38, 39, the structure of the graph was not discovered by the algorithm. To address that, a way to find a better initial embedding was considered. The initial layout can be obtained with the *Fruchterman-Reingold algorithm* [Fruchterman and Reingold, 1991]. This algorithm iteratively embeds a graph in Euclidean space: Unconnected vertices repel each other, whereas connected vertices attract each other. Repulsion and attraction act as counterparts such that the embedding distance be equal to the graph distance. In that sense, it can be understood as an optimization algorithm to isometrically embed data in Euclidean space.

Its complexity and hence its runtime scale as $\mathcal{O}(n^2)$ which makes only few (in the experiments below 100) iterations possible.

Since the Poincaré ball model is bounded within the space $\mathbb{R}^n$, the embedding in Euclidean space has to be shrunk to fit into the model. Additionally, the distances in Euclidean space and hyperbolic space are different which drastically increases the distortion in hyperbolic space.

However, the benefit of the warm-start is telling:
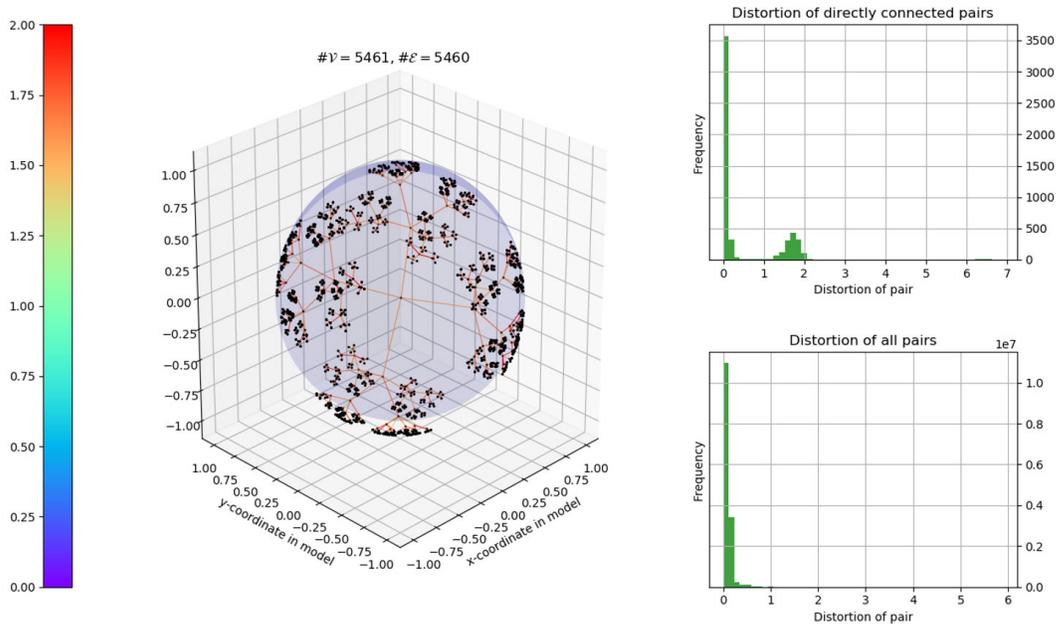
Figure 40: Tree in Poincaré ball model. The structure of the tree is clearly visible. Additionally, the distortion has decreased significantly compared to fig. 37.
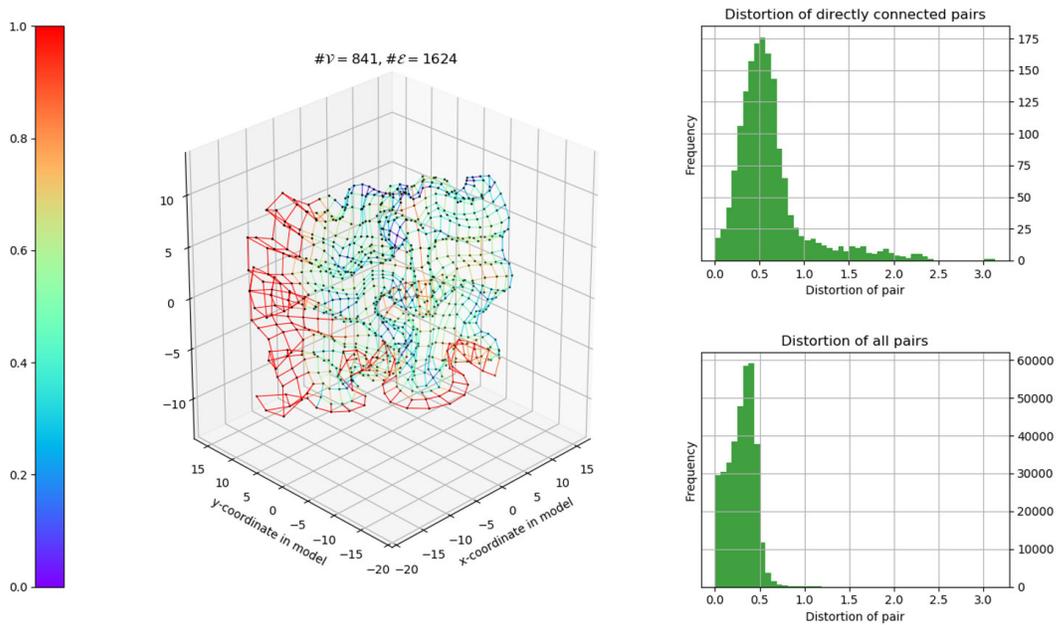
Figure 41: Lattice in Euclidean space. The structure of the graph can be seen, but the lattice is convolved and distorted at the edges.
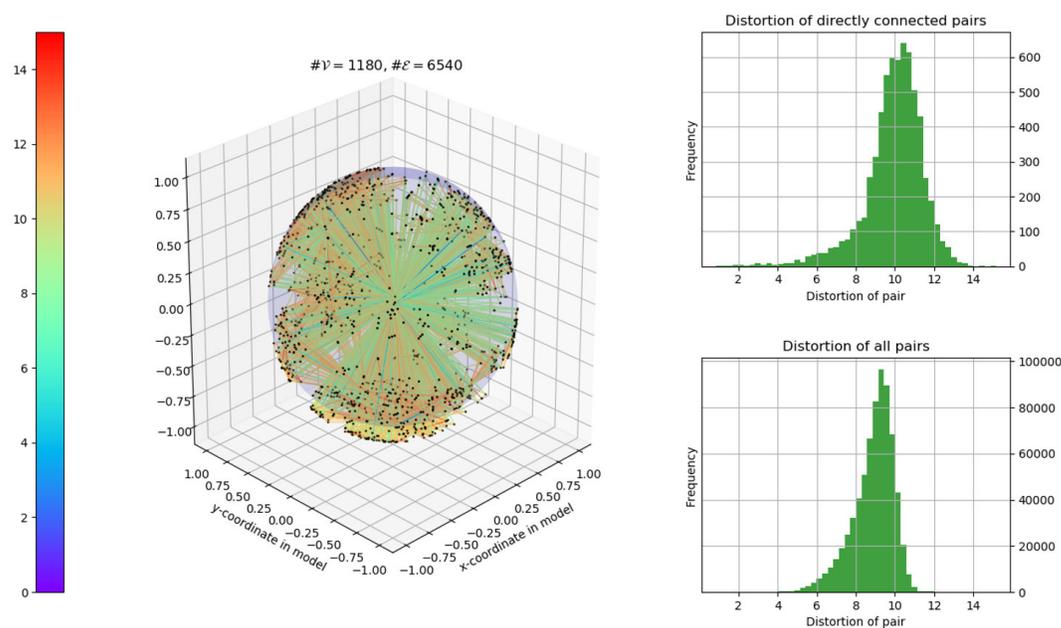
Figure 42: Mammals wordnet in Poincaré ball. The hierarchical structure is clearly visible with less prominent crossing edges and less distortion than in fig. 39.

Evaluation

The warm-start by first embedding with the Fruchterman-Reingold algorithm has clearly improved the results in terms of structure of the embedding. This property is necessary to extract graph information from the embedding such as levels of hierarchy or intrinsic data dimension.

Looking more closely at fig. 40, one notices that the distortion histogram does not seem to fit to the coloring of the edges. This is due to the depth of the tree: Every visible point in the diagram in fact corresponds to another tree layer. The distortion of their edges corresponds to the high peak in the distortion histogram.

The lattice is folded by the optimization algorithm according to eq. 247 to increase the distance between non-adjacent vertices: The expansion can be fulfilled best by also using the third dimension, although the intrinsic dimension is only two.

The embedding of the mammals wordnet expresses the hierarchical structure of the data well. The loops appear to be shrunk together: High distortion edges crossing the center of the model as in fig. 39 are much less frequent. The reason is that connected vertices have already been grouped together in the initial embedding. This can be observed when looking at the initial embedding.

17.6   Using a distance-preserving loss function

The objective to minimize the overall distortion has already been incorporated in [Fruchterman and Reingold, 1991]. It has been suggested to also adapt the loss function for a machine learning framework to that objective e.g. [Vidnes, 2010]. The variant of that loss function (which has been used in the following examples) reads:

$$\mathcal{L} = \sum_{u,v \in \mathcal{V}} \frac{\left| d(u,v) - d_{\mathcal{G}}(u,v) \right|}{d_{\mathcal{G}}(u,v)}. \tag{248}$$

The following experimental results have been obtained by randomly placing data in the embedding space and optimizing with eq. 248.
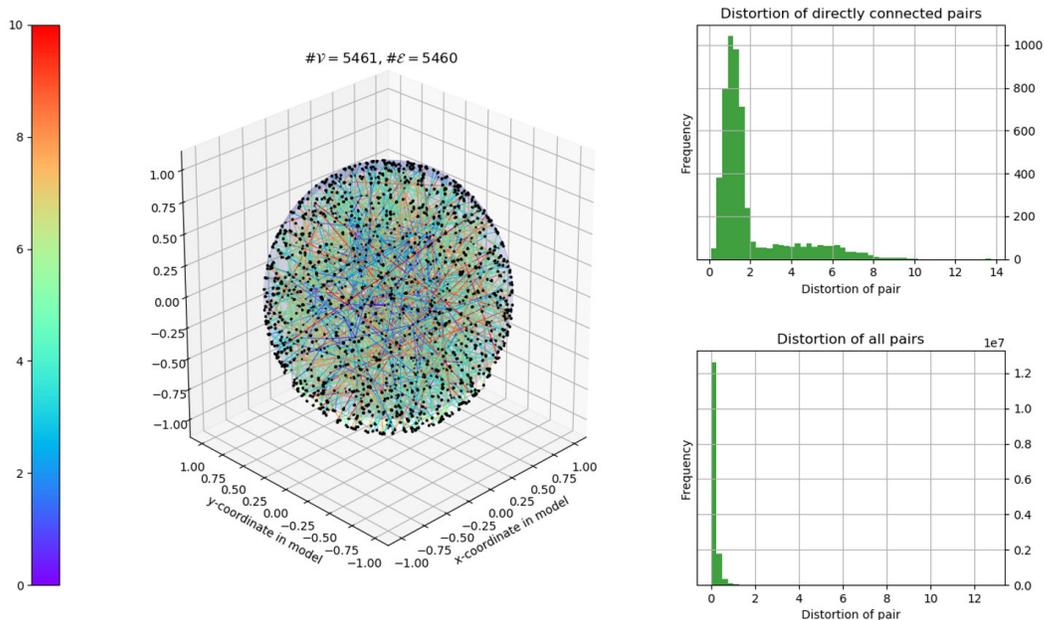
Figure 43: Tree embedded in Poincaré ball. The tree structure is not visible. The overall distortion is smaller than in the initial algorithm from fig. 37, but much bigger than with initial embedding as in fig. 40.
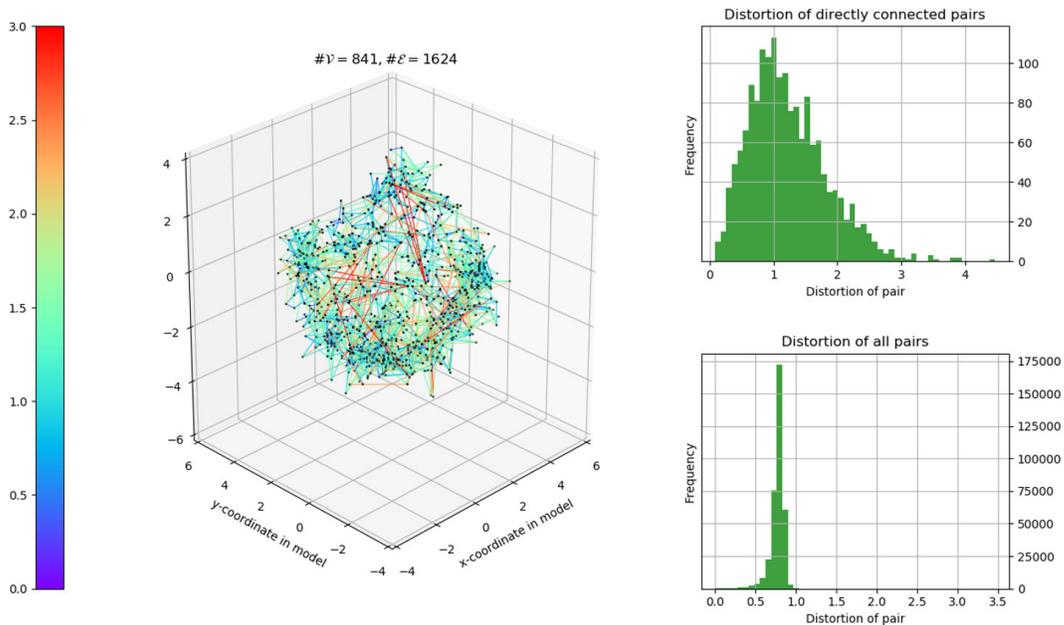
Figure 44: Lattice embedded in Euclidean space. The intrinsic 2-dimensionality of the data can be surmised. However, the embedding algorithm has arrived in a local minimum whose result does not reflect the structure of the graph.
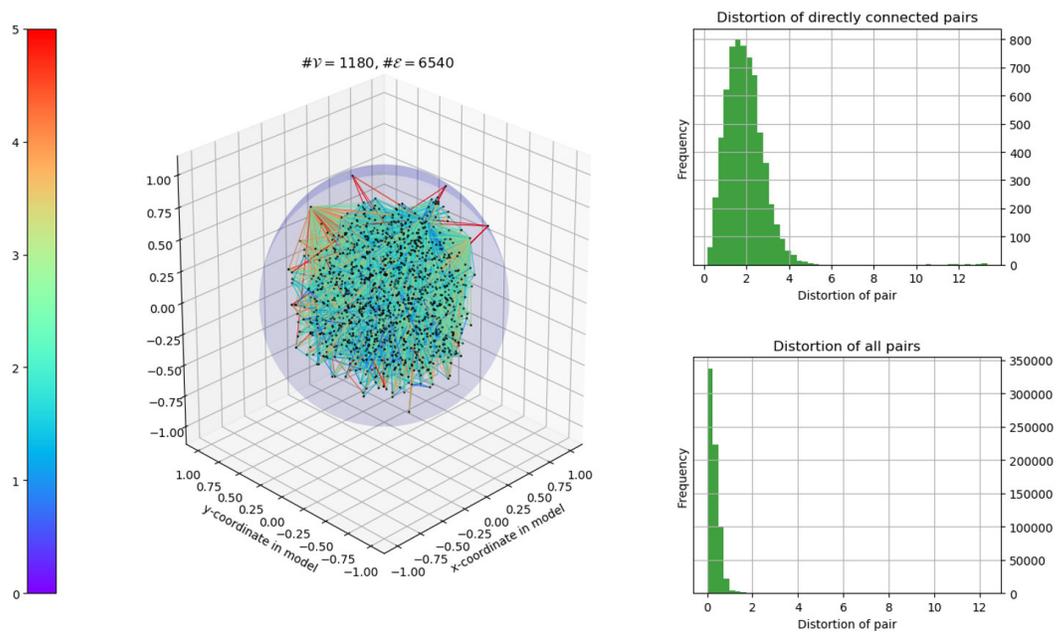
Figure 45: Mammals wordnet embedded in Poincaré ball. The hierarchical structure is not represented in the embedding. However, the overall distortion is much smaller than in the previous figures 39 and 42.

Evaluation

The results obtained for the tree and lattice graphs are tremendously inferior to the results obtained by the expanding loss function with initial Fruchterman-Reingold embedding regarding the overall distortion and the adaptation to the underlying geometry.

The results for the Mammals wordnet graph are in terms of distortion much better than the previously obtained results. This is not surprising, since the distance-preserving loss function eq. 248 explicitly minimizes the distortion. As this loss function does not aim at obtaining a high mAP, the hyperbolicity of the graph is not shown as in the previous algorithms. In that sense, the hierarchical structure of the embedding is diminished by the algorithm.

17.7 Combining Fruchterman-Reingold initial embedding with a distance-preserving loss function

Since the results from the previous experiments have been promising, combining the updated loss function with a warm-start using the Fruchterman-Reingold embedding is carried out:
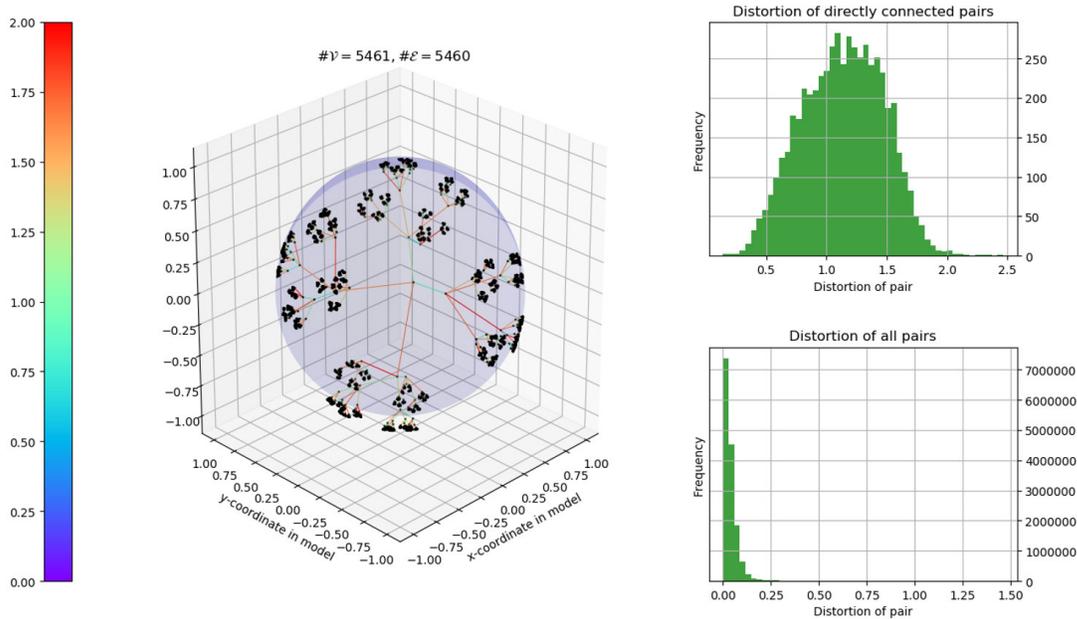
Figure 46: Tree in Poincaré ball. Note the difference between the length of the edges from this experiment to fig. 40: As the vertices are not driven to the boundary of the model (as for the expanding loss function), the full depth of the tree is visible when looking closely. Naturally, the distortion of this embedding with the length-preserving loss function is less compared to the embedding with expanding loss function shown in fig. 40.
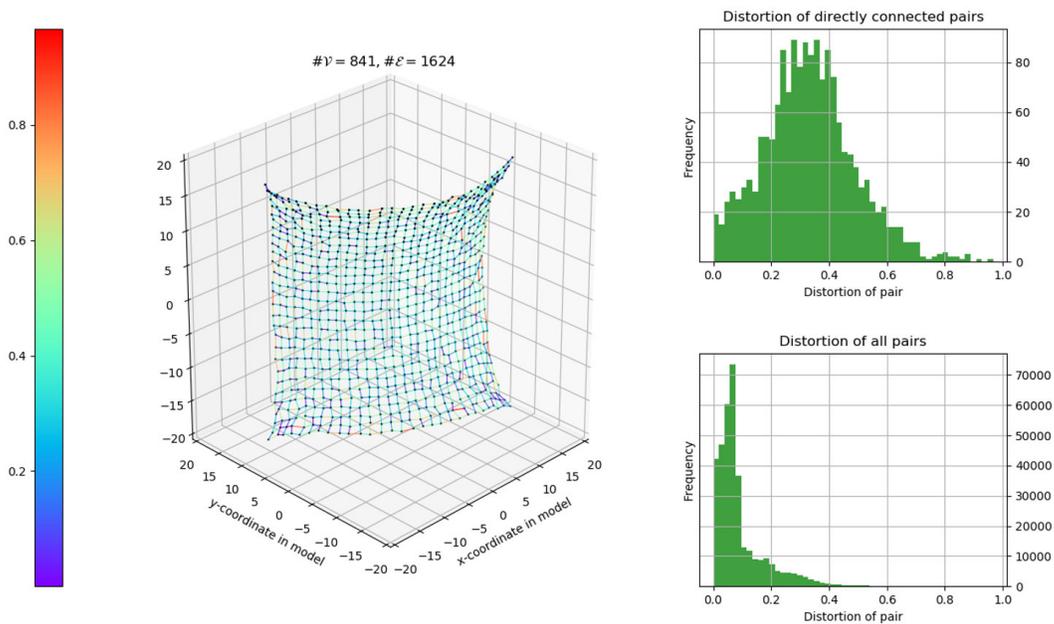
Figure 47: Lattice in 3-dimensional Euclidean space. The structure of the graph is very well visible with overall low distortion.
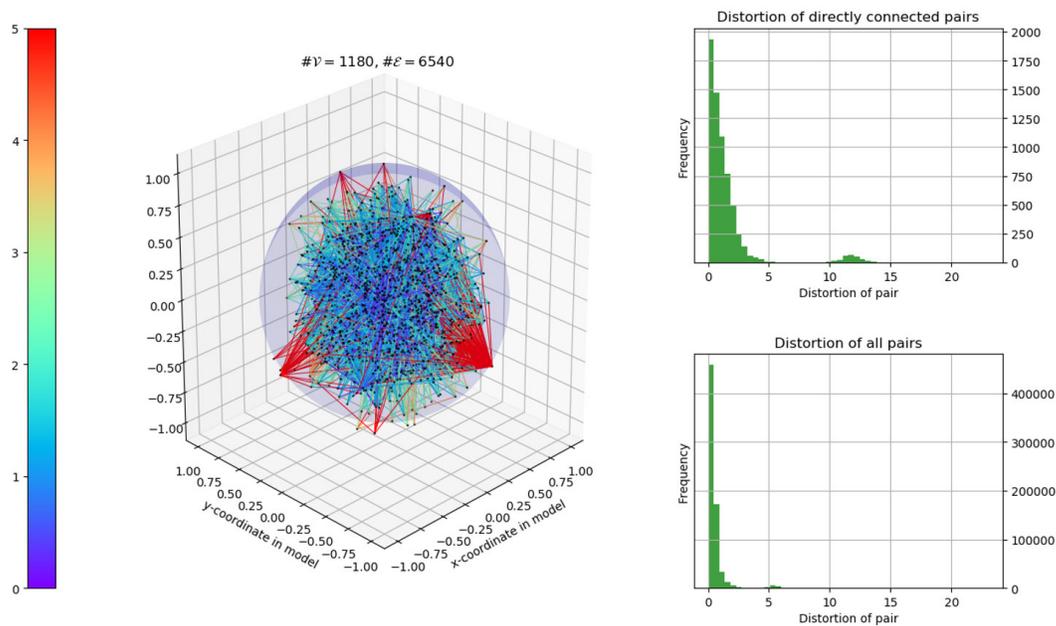
Figure 48: Mammals wordnet in Poincaré ball. Due to the distance-preserving loss function, only the remainders of the hierarchical structure are visible. The loops within the graph contract the graph.

Evaluation

The results for the tree and lattice graphs of this variant of the algorithm are salient: The structure of the graph is nicely represented by an embedding with low distortion.
The mammals graph however has not been embedded according to its hierarchical structure. The loss function aiming at low distortion has not -at the same time- preserved the hierarchy of the graph.

17.8 Quantitative results

Now the quality of the embeddings is quantified using the fidelity measures which have been revisited in sec. 17.1:
As mentioned before, the worst-case distortion D_{wc} is very sensitive to outliers[52]. The explanatory power of D_{wc} is hence very limited. To account for that, the fidelity measure RD_{avg} as the average relative distortion of an embedding ϕ is proposed as:

$$RD_{avg} = \frac{1}{|\mathcal{V}|} \sum_{u,v \in \mathcal{V}} \frac{d\big(\phi(u), \phi(v)\big)}{d_{\mathcal{G}}(u,v)}. \tag{249}$$

The results with respect to these fidelity measures are displayed in the following table:

Table 2: Experimental results comparing different optimization algorithms. Highlighted cells indicate best results. The worst-case distortion measure is deprecated due to its high sensitivity to outliers.

Dataset	fidelity measure	expanding loss function	expanding loss function (FR)	distance-preserving loss function	distance-preserving loss function (FR)
Tree graph	D_{avg}	0.455	0.092	0.143	0.039
	RD_{avg}	1.450	0.960	0.992	0.995
	mAP	0.652	0.851	0.620	0.798
Lattice graph	D_{avg}	0.329	0.293	0.770	0.090
	RD_{avg}	0.754	0.757	0.234	1.034
	mAP	0.191	0.473	0.151	0.982
Wordnet subset 'Mammals'	D_{avg}	10.151	8.888	0.323	0.496
	RD_{avg}	11.152	9.888	0.956	1.207
	mAP	0.970	0.973	0.017	0.018

---

[52]This can be seen explicitly when comparing the worst-case distortions from fig. 42 ($D_{wc} = 99.12$) to the worst-case distortion from fig. 45 ($D_{wc} = 4405.78$). The overall distortion however clearly shows the opposite: The embedding in fig. 45 is much less distorted than the embedding from fig. 42.
The worst-case distortion is hence not very informative for experimental data, although a worst-case distortion close to the ideal value 1 implies small overall distortion.

Evaluating the results on the mammals data set
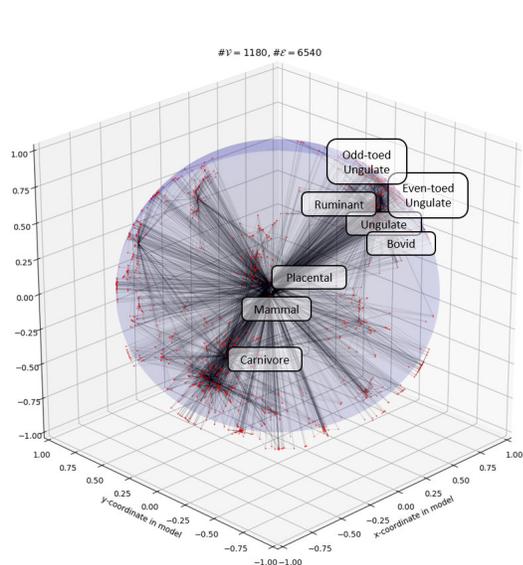
(a) Mammals dataset optimized using the unchanged algorithm and without initial embedding. The hierarchical structure has been discovered decently.
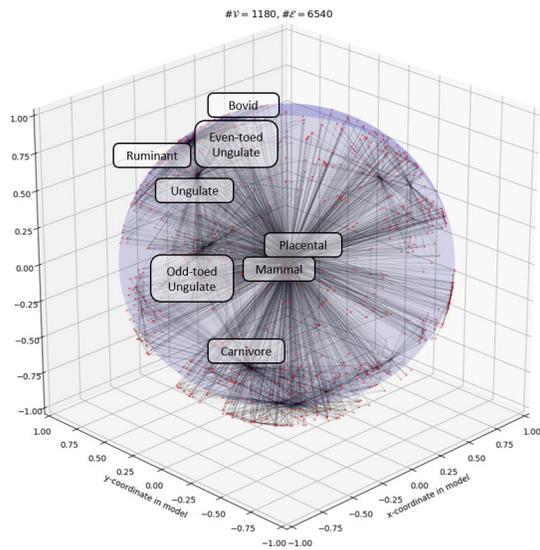
(b) Mammals dataset with expanding loss function starting with an initial embedding obtained by the Fruchterman-Reingold algorithm. The mAP score of 0.973 is higher than the one obtained by [Nickel and Kiela, 2017] in $d = 5$ dimensions.
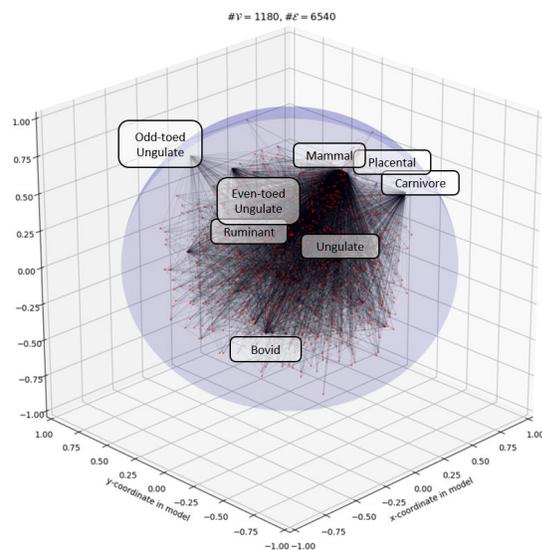
(c) Mammals dataset with distance-preserving loss function without initial embedding. One can see that the high-degree nodes have moved to the boundary of the embedding. The hierarchical structure is not stable with this loss function, as the loops compress the embedding.
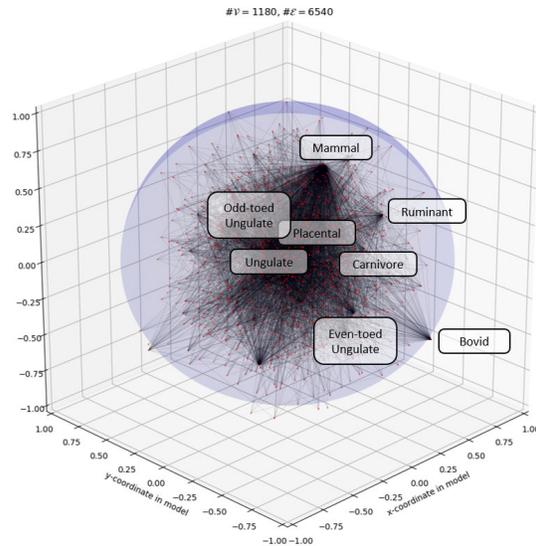
(d) Mammals dataset optimized using the distance-preserving loss function with an initial embedding obtained by the Fruchterman-Reingold algorithm. The initial hierarchical embedding is not stable, this is in particular visible at the by far highest-degree node "Mammal".

Figure 49: Mammals wordnet graph embedded in Poincaré ball using different algorithms. Eight high-degree nodes are highlighted.
Real world graphs do not fit perfectly to the proposed embedding space. The mammals data set hence determines the algorithm's stability.

17.9 Summary of the experimental results

In the experiments the variation of the loss function and its interplay with an initial embedding was examined. To obtain a faithful starting embedding, the Fruchterman-Reingold algorithm has been used. It is conceptually similar to an optimization algorithm using a mixture of a distance-preserving and expanding loss function because it has been created to visualize, i.e. stretch out graphs. It hence serves as a effective disentanglement algorithm on small datasets.

On simple test graphs such as a tree or a lattice, the approach using a distance-preserving loss function trained on a previously (partly) disentangled embedding yields respectable results.

It has been shown that this combination is crucial, since both loss functions could not overcome a random starting layout even for small graph size.

A faithful embedding possesses mAP$= 1$ and $D_{avg} = 0$. Loss functions as proposed in [Nickel and Kiela, 2017] and [Gu et al., 2019] aim at these evaluation indicators, respectively. The respective proposed loss function consequently yield best results for the fidelity measure they are designed for. However, approaching the optimization by only considering one fidelity measure does not yield overall satisfactory results. This has been shown in the experiments using the mammals wordnet graph.

18 Discussion and outlook

The idea to step back and investigate current algorithms on spaces of constant curvature arose from unsuccessful experiments in product spaces: The embeddings have not been able to represent the different structures of graphs, i.e. structures were not arranged within the corresponding subspaces. Even the proof-of-principle experiments by [Gu et al., 2019] do not contain examples of a structure-preserving embedding in a product space for data sets larger than 50 vertices.

Because of that flaw of current embedding algorithms it should be a first goal to manage faithful embeddings of artificial data sets in spaces of constant sectional curvature.

It has been pointed out that a mayor problem is initial entanglement which results in a bad local minimum of the optimization algorithm. One might argue that choosing a high dimension of the embedding space facilitates disentangling the graph structure. However, the reason to embed into curved spaces is to reduce the necessary dimension of the embedding space. Although high dimensional embeddings can preserve the structure and minimize the distortion even in Euclidean space, this rules out the possibility to learn the intrinsic dimension of the data and to detect structures of the graph. Nevertheless, I tried out a disentanglement in high dimensional space followed by a projection. This approach however has not yielded presentable results, yet.

Another approach to tackle entanglement is to consecutively optimize on different length scales. Clustering algorithms with respect to the graph distance could reduce the intrinsic dimensionality of the graph and hence allow to faithfully embed the "graph of clusters". In a subsequent step the clusters could be unfolded and embedded without disturbing the bigger structure.

Another promising approach could be to cut the graph into small patches. Due to their small size, they can be processed quickly by a disentangling algorithm in the spirit of the Fruchterman-Reingold algorithm. At the same time these patches can be analyzed in terms of their intrinsic structure and could hence be placed accordingly into the embedding space.

Additionally, even nicely preprocessed data is not sure to converge to a meaningful embedding as seen in fig. 41, 48. Both loss functions optimize with respect to a fidelity measure, in particular the mean average precision and the average distortion. Optimizing one does not necessarily also optimize the other fidelity measure, although a ground-truth embedding receives perfect scores in regard of both.

It is hence proposed to adapt the loss function such that it involves both fidelity measures.

19 Conclusion

In order to familiarize with the concept of symmetric spaces, a comprehensible introduction to the theory of symmetric spaces was given. As a low-distortion embedding of graphs stems from the suiting curvature of the embedding space, the focus lies on examining the sectional curvature and totally geodesic submanifolds. These determine the sub-structures which can be embedded faithfully in that manifold.

With the example of the Siegel upper half-space, the data representation capability of an irreducible symmetric space was examined in terms of totally geodesic submanifolds and their sectional curvature. A machine-differentiable implementation of the gradient descent algorithm was proposed.

It is argued that the class of symmetric spaces in general possesses superior data representation capabilities. It is hence likely that they outperform Cartesian product spaces on data representation tasks.

In the experimental section, different variants of a gradient descent algorithm have been investigated. As the two examined loss functions only optimize with respect to one graph feature only, it is not surprising that the results regarding the respective other objective are less convincing. A loss function combining these graph features could address that challenge.

It is furthermore proposed to carefully pre-process data before optimizing the algorithm using machine-learning techniques. Although hand-crafted algorithms are often regarded as inferior to purely unsupervised optimization algorithms, it is argued that the combination of approaches experimentally leads to promising results. In the language of machine-learning, this corresponds to initializing the optimization problem closer to the global minimum.

The interpretability of data representations is highly dependent on their property of being structure-preserving. In its current state the high expectations of interpretability cannot be fulfilled, yet. It is however to be expected that improved embedding algorithms will eventually be able to exploit the full richness of data representation capabilities of symmetric spaces by reliably finding unified, structure-preserving embeddings.

20 References

[Adcock et al., 2013] Adcock, A. B., Sullivan, B. D., and Mahoney, M. W. (2013). Tree-like structure in large social and information networks. In *2013 IEEE 13th International Conference on Data Mining*, pages 1–10. IEEE.

[Boguná et al., 2010] Boguná, M., Papadopoulos, F., and Krioukov, D. (2010). Sustaining the internet with hyperbolic mapping. *Nature communications*, 1(1):1–8.

[Bonnabel, 2013] Bonnabel, S. (2013). Stochastic gradient descent on riemannian manifolds. *IEEE Transactions on Automatic Control*, 58(9):2217–2229.

[Bunse-Gerstner and Gragg, 1988] Bunse-Gerstner, A. and Gragg, W. B. (1988). Singular value decompositions of complex symmetric matrices. *Journal of Computational and Applied Mathematics*, 21(1):41–54.

[Chebotarev and Teretenkov, 2014] Chebotarev, A. M. and Teretenkov, A. E. (2014). Singular value decomposition for the takagi factorization of symmetric matrices. *Applied Mathematics and Computation*, 234:380–384.

[De Sa et al., 2018] De Sa, C., Gu, A., Ré, C., and Sala, F. (2018). Representation tradeoffs for hyperbolic embeddings. *Proceedings of machine learning research*, 80:4460.

[Eberlein, 1996] Eberlein, P. (1996). *Geometry of nonpositively curved manifolds*. University of Chicago Press.

[Ehrlich, 1970] Ehrlich, L. W. (1970). Complex matrix inversion versus real. *Communications of the ACM*, 13(9):561–562.

[Eschenburg, 1997] Eschenburg, J.-H. (1997). Lecture notes on symmetric spaces. *preprint*.

[Freitas, 1999] Freitas, P. J. (1999). *On the action of the symplectic group on the Siegel upper half plane*. PhD thesis, University of Illinois at Chicago.

[Fruchterman and Reingold, 1991] Fruchterman, T. M. and Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and experience*, 21(11):1129–1164.

[Gromov, 1987] Gromov, M. (1987). Hyperbolic groups. In *Essays in group theory*, pages 75–263. Springer.

[Gu et al., 2019] Gu, A., Sala, F., Gunel, B., and Ré, C. (2019). Learning mixed-curvature representations in product spaces. In *International Conference on Learning Representations*.

[Gugelmann et al., 2012] Gugelmann, L., Panagiotou, K., and Peter, U. (2012). Random hyperbolic graphs: degree sequence and clustering. In *International Colloquium on Automata, Languages, and Programming*, pages 573–585. Springer.

[Helgason, 1979] Helgason, S. (1979). *Differential geometry, Lie groups, and symmetric spaces*. Academic press.

[Iozzi, 2018] Iozzi, A. (2018). Introduction to symmetric spaces. *preprint*.

[Kleinberg, 2007] Kleinberg, R. (2007). Geographic routing using hyperbolic space. In *IEEE INFOCOM 2007-26th IEEE International Conference on Computer Communications*, pages 1902–1909. IEEE.

[Moler and Van Loan, 2003] Moler, C. and Van Loan, C. (2003). Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM review*, 45(1):3–49.

[Myers and Steenrod, 1939] Myers, S. B. and Steenrod, N. E. (1939). The group of isometries of a riemannian manifold. *Annals of Mathematics*, pages 400–416.

[Nakahara, 2003] Nakahara, M. (2003). *Geometry, topology and physics*. CRC Press.

[Nickel and Kiela, 2017] Nickel, M. and Kiela, D. (2017). Poincaré embeddings for learning hierarchical representations. In *Advances in neural information processing systems*, pages 6338–6347.

[Nickel and Kiela, 2018] Nickel, M. and Kiela, D. (2018). Learning continuous hierarchies in the lorentz model of hyperbolic geometry. *arXiv preprint arXiv:1806.03417*.

[Perozzi et al., 2014] Perozzi, B., Al-Rfou, R., and Skiena, S. (2014). Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 701–710.

[Ravasz and Barabási, 2003] Ravasz, E. and Barabási, A.-L. (2003). Hierarchical organization in complex networks. *Physical review E*, 67(2):026112.

[Roweis and Saul, 2000] Roweis, S. T. and Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *science*, 290(5500):2323–2326.

[Ruder, 2016] Ruder, S. (2016). An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*.

[Sakai, 1996] Sakai, T. (1996). *Riemannian geometry*, volume 149. American Mathematical Soc.

[Sarkar, 2011] Sarkar, R. (2011). Low distortion delaunay embedding of trees in hyperbolic plane. In *International Symposium on Graph Drawing*, pages 355–366. Springer.

[Siegel, 1943] Siegel, C. L. (1943). Symplectic geometry. *American Journal of Mathematics*, 65(1):1–86.

[Theis, 2005] Theis, F. (2005). Gradients on matrix manifolds and their chain rule. *Neural Information Processing LR*, 9:1–13.

[Torgerson, 1952] Torgerson, W. S. (1952). Multidimensional scaling: I. theory and method. *Psychometrika*, 17(4):401–419.

[Vidnes, 2010] Vidnes, N. (2010). Gradient descent algorithm incorporating stochastic pointlocation schemes and its application in multidimensional scaling analysis. Master's thesis, University of Agder.

Erklärung:

Ich versichere, dass ich diese Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

31.05.2020

Heidelberg, den   . . . .