Topological data analysis
Homework 1

Sara Bizjak (27202020)

November 1, 2020

1 Theoretical problems

1.1 Exploring different metrics

a) Determing the distances between the points $(2, 1),\ (4, 2),\ (0, 2)$ in metrics $\alpha,\ \beta,\ \gamma$.

- Metric α.

$$\mathrm{d}_\alpha\big((2,1),\ (4,2)\big) = \sqrt{2^2+1^2} + \sqrt{4^2+2^2} = \sqrt{5}+\sqrt{20} = 6.708203932499369.$$

$$\mathrm{d}_\alpha\big((2,1),\ (0,2)\big) = \sqrt{2^2+1^2} + \sqrt{0^2+2^2} = \sqrt{5}+\sqrt{4} = 4.23606797749979.$$

$$\mathrm{d}_\alpha\big((4,2),\ (0,2)\big) = \sqrt{4^2+2^2} + \sqrt{0^2+2^2} = \sqrt{20}+\sqrt{4} = 6.47213595499958.$$

- Metric β.

$$\mathrm{d}_\beta\big((2,1),\ (4,2)\big) = \sqrt{(2-4)^2+(1-2)^2} = \sqrt{4+1} = 2.23606797749979.$$

$$\mathrm{d}_\beta\big((2,1),\ (0,2)\big) = \sqrt{2^2+1^2} + \sqrt{0^2+2^2} = \sqrt{5}+\sqrt{4} = 4.23606797749979.$$

$$\mathrm{d}_\beta\big((4,2),\ (0,2)\big) = \sqrt{4^2+2^2} + \sqrt{0^2+2^2} = \sqrt{20}+\sqrt{4} = 6.47213595499958.$$

- Metric γ.

$$\mathrm{d}_\gamma\big((2,1),\ (4,2)\big) = |2-4| + |1| + |2| = 2+1+2 = 5.$$

$$\mathrm{d}_\gamma\big((2,1),\ (0,2)\big) = |2-0| + |1| + |2| = 5.$$

$$\mathrm{d}_\gamma\big((4,2),\ (0,2)\big) = |4-0| + |2| + |2| = 4+2+2 = 8.$$

b) Draw the open balls $B\big((0,0),1\big)$, $B\big((1,0),2\big)$, $B\big((0,2),6\big)$ in α metric.
Note that in an open ball are all points that are away from the center for $< r$, where r is a radius. Since a center is away from itself for 0, it's always contained in an open ball.

$$\begin{aligned} B\big((0,0),1\big) &= \{(0,0)\} \cup \{(x,y) \in \mathbb{R}^2 : \sqrt{x^2+y^2} + \sqrt{0^2+0^2} < 1\} \\ &= \{(0,0)\} \cup \{(x,y) \in \mathbb{R}^2 : \sqrt{x^2+y^2} < 1\}. \end{aligned}$$

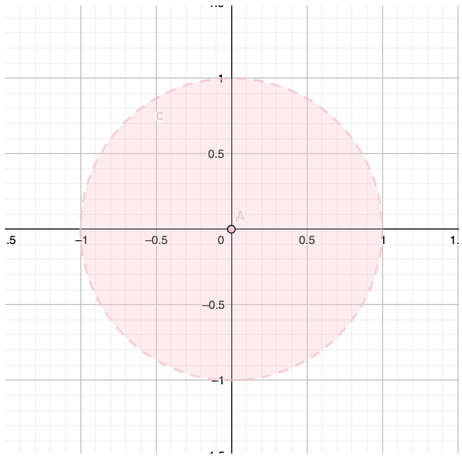

Figure 1: $B\big((0,0),1\big)$ in metric α.

$$\begin{aligned} B\big((1,0),2\big) &= \{(1,0)\} \cup \{(x,y) \in \mathbb{R}^2 \setminus (1,0) : \sqrt{x^2+y^2} + \sqrt{1^2+0^2} < 2\} \\ &= \{(1,0)\} \cup \{(x,y) \in \mathbb{R}^2 \setminus (1,0) : \sqrt{x^2+y^2} < 1\}. \end{aligned}$$

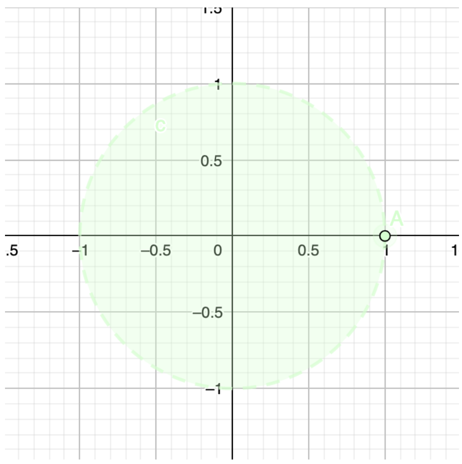

Figure 2: $B\big((1,0),2\big)$ in metric α.

$$B\big((0,2),6\big) = \{(0,2)\} \cup \{(x,y) \in \mathbb{R}^2 \setminus (0,2) : \sqrt{x^2+y^2} + \sqrt{0^2+2^2} < 6\}$$
$$= \{(0,2)\} \cup \{(x,y) \in \mathbb{R}^2 \setminus (0,2) : \sqrt{x^2+y^2} < 4\}.$$

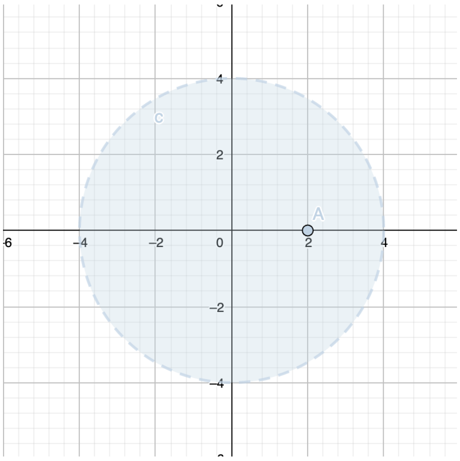

Figure 3: $B\big((0,2),6\big)$ in metric α.

c) Draw the open balls $B\big((0,0),1\big)$, $B\big((1,0),2\big)$, $B\big((2,2),3\sqrt{2}\big)$ in β metric.

$$B\big((0,0),1\big) = \{(x,y) \in \mathbb{R}^2 : \sqrt{(x-0)^2+(y-0)^2} < 1\}$$
$$= \{(x,y) \in \mathbb{R}^2 : \sqrt{x^2+y^2} < 1\}.$$

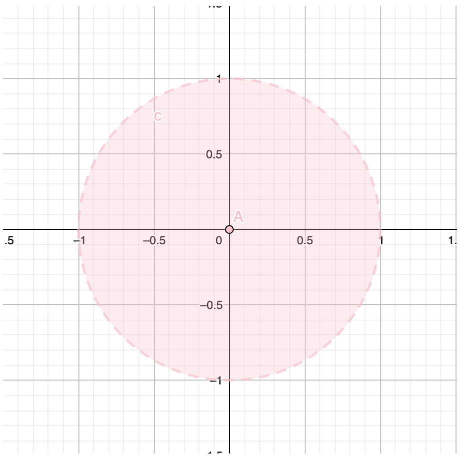

Figure 4: $B\big((0,0),1\big)$ in metric β.

$$\begin{aligned} B\big((1,0),2\big) &= \{(x,0) \in \mathbb{R}^2 : \sqrt{(x-1)^2} < 2\} \cup \{(x,y) \in \mathbb{R}^2,\ y \neq 0 : \sqrt{x^2+y^2} + \sqrt{1^2} < 2\} \\ &= \{(x,0) \in \mathbb{R}^2 : |x-1| < 2\} \cup \{(x,y) \in \mathbb{R}^2,\ y \neq 0 : \sqrt{x^2+y^2} < 1\} \\ &= \{(x,0) \in \mathbb{R}^2 : -1 < x < 3\} \cup \{(x,y) \in \mathbb{R}^2,\ y \neq 0 : \sqrt{x^2+y^2} < 1\}. \end{aligned}$$

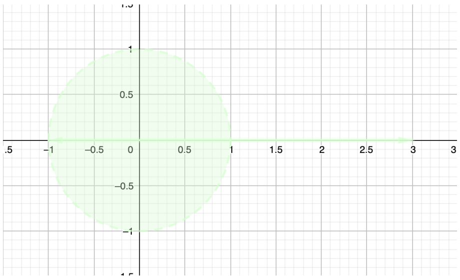

Figure 5: $B\big((1,0),2\big)$ in metric β.

$$\begin{aligned} B\big((2,2),3\sqrt{2}\big) &= \{(x,x) \in \mathbb{R}^2 : \sqrt{2\cdot(x-2)^2} < 3\sqrt{2}\} \cup \{(x,y) \in \mathbb{R}^2,\ x \neq y : \sqrt{x^2+y^2} + \sqrt{2\cdot 2^2} < 3\sqrt{2}\} \\ &= \{(x,x) \in \mathbb{R}^2 : \sqrt{2}\cdot|x-2| < 3\sqrt{2}\} \cup \{(x,y) \in \mathbb{R}^2,\ x \neq y : \sqrt{x^2+y^2} + 2\sqrt{2} < 3\sqrt{2}\} \\ &= \{(x,x) \in \mathbb{R}^2 : -1 < x < 5\} \cup \{(x,y) \in \mathbb{R}^2,\ x \neq y : \sqrt{x^2+y^2} < \sqrt{2}\}. \end{aligned}$$

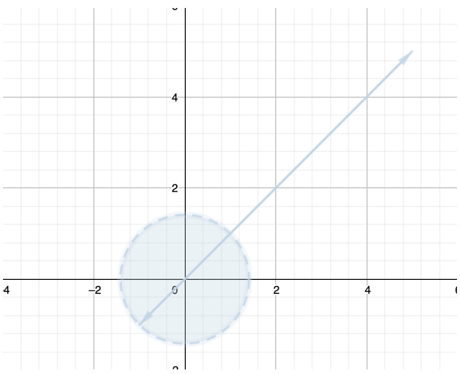

Figure 6: $B\big((2,2),3\sqrt{2}\big)$ in metric β.

d) Draw the open balls $B\big((0,0),1\big)$, $B\big((1,0),2\big)$, $B\big((2,0),3\big)$ in γ metric.

$$\begin{aligned} B\big((0,0),1\big) &= \{(0,y) \in \mathbb{R}^2 : |y-0| < 1\} \cup \{(x,y) \in \mathbb{R}^2,\ x \neq 0 : |x-0| + |y-0| < 1\} \\ &= \{(0,y) \in \mathbb{R}^2 : -1 < y < 1\} \cup \{(x,y) \in \mathbb{R}^2 : -1 < x < 1,\ -1 < y < 1\}. \end{aligned}$$

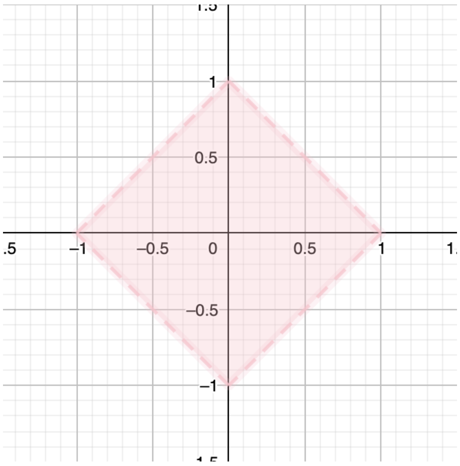

Figure 7: $B\big((0,0),1\big)$ in metric γ.

$$
\begin{aligned}
B\big((1,0),2\big) &= \{(1,y)\in\mathbb{R}^2 : |y-0|<2\}\cup\{(x,y)\in\mathbb{R}^2,\ x\neq 1 : |x-1|+|y-0|<2\}\\
&= \{(1,y)\in\mathbb{R}^2 : -2<y<2\}\cup\{(x,y)\in\mathbb{R}^2,\ x\neq 1 : -1<x<3,\ -2<y<2\}.
\end{aligned}
$$

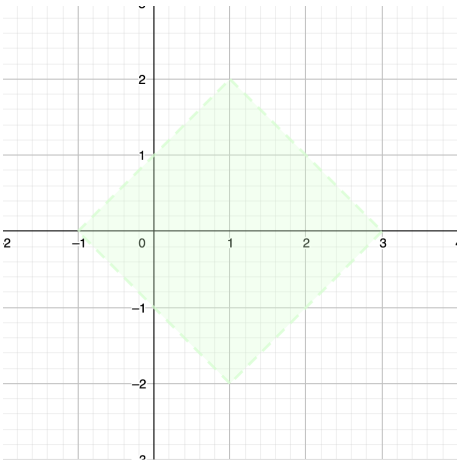

Figure 8: $B\big((1,0),2\big)$ in metric γ.

$$
\begin{aligned}
B\big((2,0),3\big) &= \{(2,y)\in\mathbb{R}^2 : |y-0|<3\}\cup\{(x,y)\in\mathbb{R}^2,\ x\neq 2 : |x-2|+|y-0|<3\}\\
&= \{(2,y)\in\mathbb{R}^2 : -3<y<3\}\cup\{(x,y)\in\mathbb{R}^2,\ x\neq 2 : -1<x<5,\ -3<y<3\}.
\end{aligned}
$$

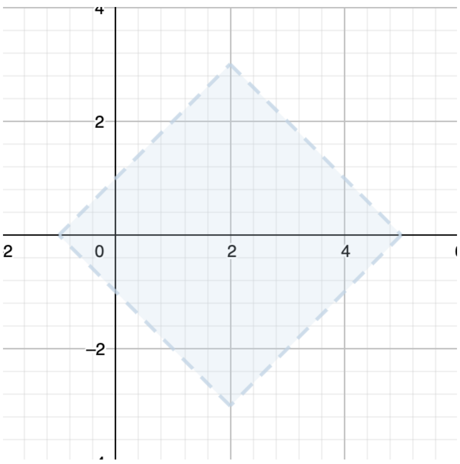

Figure 9: $B\big((2,0),3\big)$ in metric γ.

1.2 Discrete metric

The discrete metric on a space X is defined as $d: \ X \times X \to \mathbb{R}$, where $d(x,y) = 0$ if $x = y$ and 1 otherwise.
a) Let $X = \mathbb{N}$. Describe $B\big(1, \frac{1}{2}\big)$ and $B\big(2, 1\big)$.

$$B\left(1, \tfrac{1}{2}\right) = \{x \in \mathbb{N} : d(1,x) < \tfrac{1}{2}\} = \{x = 1 : d(1,1) = 0\} = \{1\}.$$
$$B(2,1) = \{x \in \mathbb{N} : d(2,x) < 1\} = \{x = 2 : d(2,2) = 0\} = \{2\}.$$

b) The triangle with vertices at distinct integers a, b, c is always equilateral, because the distance between distinct points in discrete metric is always 1.

1.3 Homeomorphic spaces

Let $X = S^{n-1} \times [0,1] \subset \mathbb{R}^{n+1}$ and $Y = \{(x_1, \dots, x_n) \in \mathbb{R}^n \ : \ 1 \leq x_1^2 + \cdots + x_n^2 \leq 4\}$. What we want to do is to prove that X and Y are homeomorphic. Therefore, we need to define continuous functions $f : X \to Y$ and $g : Y \to X$ and show that $g = f^{-1}$. We do that by calculating $f \ \circ g$ and $g \circ f$ and show that they are identities.
To get an idea, we first look at the sketches of spaces X and Y for $n = 1$ and $n = 2$, denoted as X_1, X_2 and Y_1, Y_2.

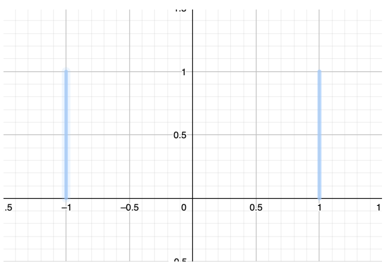

Figure 10: $X_1 = S^0 \times [0,1] \subset \mathbb{R}^2$.

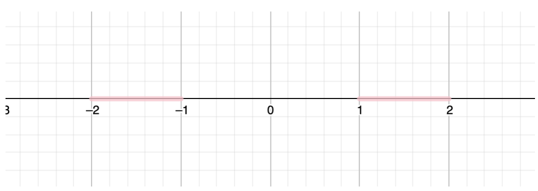

Figure 11: $Y_1 = \{x_1 \in \mathbb{R} : 1 \leq x_1^2 \leq 4\}$.

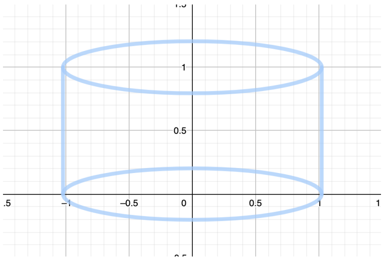

Figure 12: $X_2 = S^1 \times [0,1] \subset \mathbb{R}^3$.

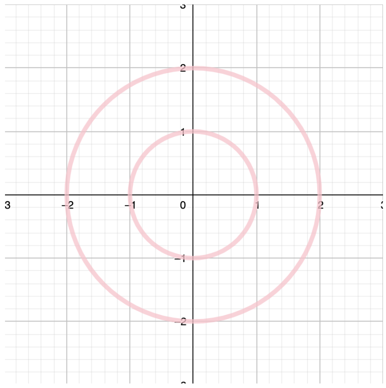

Figure 13: $Y_2 = \{(x_1, x_2) \in \mathbb{R}^2 : 1 \leq x_1^2 + x_2^2 \leq 4\}$.

We imagine space X_2 as an union of disjointed lines connecting upper and lower circle edges. Similarly, we imagine space Y_2. Let's present this idea in the following pictures.

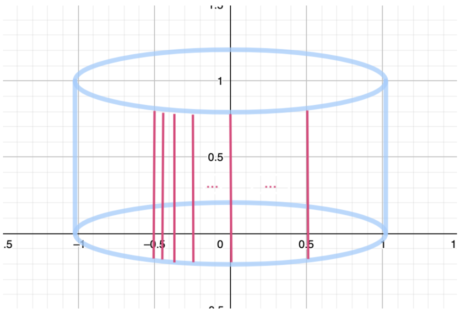

Figure 14: X_2 as an union of disjoint lines.

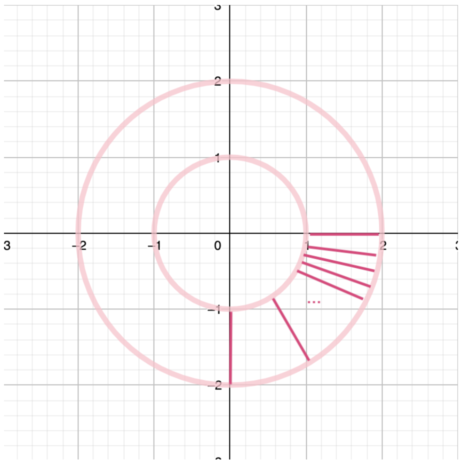

Figure 15: Y_2 as an unoin of disjoint lines.

We see that we can go from space X_2 to space Y_2 by "flipping" lines down on the plane. And equivalently, from Y_2 to X_2 we do the same thing but in reverse.
Lines in X_2 have the form of $(1-t)\cdot(x_1, x_2, 0) + t \cdot (x_1, x_2, 1)$ and lines in Y_2 have the form of $(1-t)\cdot(x_1, x_2) + 2t \cdot (x_1, x_2)$, where $t \in [0,1]$.
To get from X_2 to Y_2 we have to project lines to x_1x_2-plane (by skipping the third coordinate) and then stretch them with factor $(1+t)$, where $t \in [0,1]$. On the other hand, to get from Y_2 to X_2 we have to squeeze the lines to the unit circle (normalization) and then stretch them up from 0 to 1, so we write the new coordinate, denoted by t, as a function of x_1 and x_2. Following this idea, functions $f_2 : X_2 \to Y_2$ and $g_2 : Y_2 \to X_2$ are:
$f_2(x_1, x_2, t) = (x_1 \cdot (1+t), x_2 \cdot (1+t))$ and $g_2(x_1, x_2) = \left(\frac{x_1}{\sqrt{x_1^2+x_2^2}}, \frac{x_2}{\sqrt{x_1^2+x_2^2}}, \sqrt{x_1^2+x_2^2} - 1\right)$, where $t \in [0,1]$.

Let's get back to X and Y spaces and use our idea for general n. Let $f : X \to Y$ and let $g : Y \to X$. Function f is:

$$f(x_1,\dots,x_n,t) = (x_1\cdot(1+t),\dots,x_n\cdot(1+t)) = (1+t)\cdot(x_1,\dots,x_n).$$

Now, we need to verify that $(1+t)\cdot(x_1,\dots,x_n) \in Y$. Because $(x_1,\dots,x_n,t) \in X$, then $x_1^2+\dots+x_n^2 = 1$ and $t\in[0,1]$. So $((1+t)\cdot x_1)^2+\dots+((1+t)\cdot x_2)^2 = (1+t)^2\cdot(x_1^2+\dots+x_n^2)^2 = (1+t)^2$ and $1 \le (1+t)^2 \le 4$, so the condition is satisfied.

For function g we need to express the new coordinate t as a function of $(x_1,\dots,x_n)$. For $t\in[0,1]$ the relation $t = \sqrt{x_1^2+\cdots x_n^2}-1$ is obvious.

$$g(x_1,\dots,x_n) = \left(\frac{x_1}{\sqrt{x_1^2+\dots+x_n^2}},\dots,\frac{x_n}{\sqrt{x_1^2+\dots+x_n^2}},\sqrt{x_1^2+\dots+x_n^2}-1\right).$$

Similarly as we did for function f, we now need to verify that $\left(\frac{x_1}{\sqrt{x_1^2+\dots+x_n^2}},\dots,\frac{x_n}{\sqrt{x_1^2+\dots+x_n^2}},\sqrt{x_1^2+\dots+x_n^2}-1\right) \in X$. Because $(x_1,\dots,x_n)\in Y$ follows $1 \le x_1^2+\dots+x_n^2 \le 4$. Since $\left(\frac{x_1}{\sqrt{x_1^2+\dots+x_n^2}}\right)^2+\dots+\left(\frac{x_n}{\sqrt{x_1^2+\dots+x_n^2}}\right)^2 = \frac{x_1^2+\dots+x_n^2}{x_1^2+\dots+x_n^2} = 1$ and $t = \sqrt{x_1^2+\dots+x_n^2}-1$ is equivalent to $t\in[0,1]$, the condition is satisfied.

Clearly, both functions are continuous.
All there is left to do is to calculate both compositums. Both compositums are also continuous.

$$(f\circ g): Y \to X \to Y$$

$$\begin{aligned}
(f\circ g)(x_1,\dots,x_n) &= f\left(\frac{x_1}{\sqrt{x_1^2+\dots+x_n^2}},\dots,\frac{x_n}{\sqrt{x_1^2+\dots+x_n^2}},\sqrt{x_1^2+\dots+x_n^2}-1\right)\\
&= \left(\frac{x_1}{\sqrt{x_1^2+\dots+x_n^2}}\cdot\left(\sqrt{x_1^2+\dots+x_n^2}\right),\dots,\frac{x_n}{\sqrt{x_1^2+\dots+x_n^2}}\cdot\left(\sqrt{x_1^2+\dots+x_n^2}\right)\right)\\
&= (x_1,\dots,x_n) = \mathrm{id}_Y
\end{aligned}$$

$$(g\circ f): X \to Y \to X$$

$$\begin{aligned}
(g\circ f)(x_1,\dots,x_n,t) &= g\left(x_1\cdot(1+t),\dots,x_n\cdot(1+t)\right)\\
&= (1+t)\cdot\left(\frac{x_1}{\sqrt{x_1^2+\dots+x_n^2}},\dots,\frac{x_n}{\sqrt{x_1^2+\dots+x_n^2}},\sqrt{x_1^2+\dots+x_n^2}-1\right)\\
&= (x_1,\dots,x_n,t) = \mathrm{id}_X
\end{aligned}$$

$\implies X\cong Y.$

1.4 Homeomorphic spaces

Let $S_+^n = \{(x_1, \ldots, x_{n+1}) \in S^n\ ;\ x_{n+1} \geq 0\}$ and $B^n = \{(x_1, \ldots, x_n) \in \mathbb{R}^n\ ;\ x_1^2 + \cdots x_n^2 \leq 1\}$. We want to prove that S_+^n and B^n are homeomorphic. To get an idea it's sufficient to look at sketches for $n = 1$.

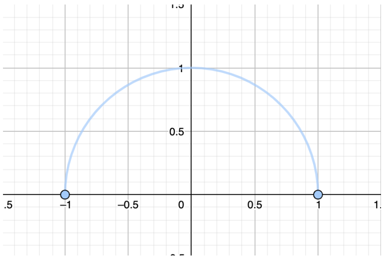

Figure 16: $S_+^1 = \{(x_1, x_2) \in S^1\ ;\ x_2 \geq 0\}$.

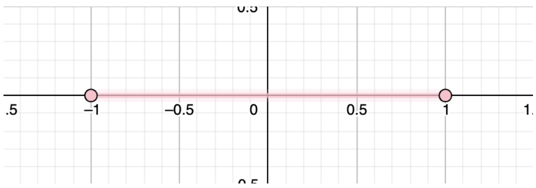

Figure 17: $B^1 = \{x_1 \in \mathbb{R}\ ;\ {}_1^2 \leq 1\}$.

To get from left to right, the idea is to project the circle down on the line (so we no longer use coordinate x_2 and coordinate x_1 stays the same), and from right to left to stretch the line up to circle (we add the new coordinate x_2 as a function of x_1).

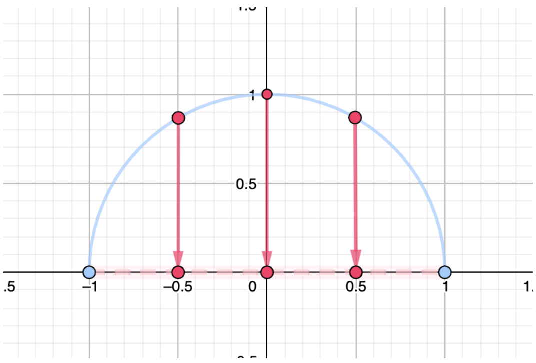

Figure 18: The idea for function $S_+^1 \to B^1$.

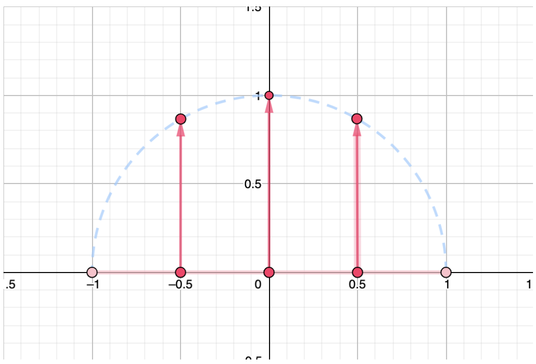

Figure 19: The idea for function $B^1 \to S_+^1$.

Using that idea, we write the regulation for general n.
Let $f : S_+^n \subset \mathbb{R}^{n+1} \to B^n \subset \mathbb{R}^n$ and let $g : B^n \in \mathbb{R}^n \to S_+^n \in \mathbb{R}^{n+1}$.

As said before, to determine function f, we just skip the last coordinate.

$$f(x_1, x_2, \ldots, x_{n+1}) = (x_1, x_2, \ldots, x_n).$$

We need to check if $(x_1, \ldots, x_n) \in B^n$. For $(x_1, x_2, \ldots, x_{n+1}) \in S_+^n$ the relations $x_1^2 + \cdots + x_n^2 + x_{n+1}^2 = 1$ and $x_{n+1} \in [0, 1]$ are valid and that implies that $x_1^2 + \cdots + x_n^2 = 1 - x_{n+1}^2 \leq 1$, so the condition is satisfied.

For function g we need to express the new coordinate x_{n+1} as a function of $(x_1, \ldots, x_n)$. We get $x_{n+1} = \sqrt{1 - x_1^2 - \cdots - x_n^2}$ (we take the positive root because we are in the S_+^n).

$$g(x_1,\ldots,x_n) = \left(x_1,\ldots,x_n,\sqrt{1-x_1^2-\cdots-x_n^2}\right).$$

We need to check if $\left(x_1,\ldots,x_n,\sqrt{1-x_1^2-\cdots-x_n^2}\right) \in S^n_+$. For $(x_1,\ldots,x_n) \in B^n$ the relation $x_1^2+\cdots x_n^2 \leq 1$ is true and $x_1^2+\cdots+x_n^2+x_{n+1}^2 = x_1^2+\cdots+x_n^2+1-x_1^2-\cdots-x_n^2 = 1$, so $x_{n+1} \geq 0$ and the condition is satisfied.
Clearly, both function are continuous.
All there is left to do is to calculate both compositums, which are continuous, too.

$$(f\circ g): B^n \to S^n_+ \to B^n$$

$$(f\circ g)(x_1,\ldots,x_n) = f\left(x_1,\ldots,x_n,\sqrt{1-x_1^2-\cdots-x_n^2}\right) = (x_1,\ldots,x_n) = \mathrm{id}_{B^n}$$

$$(g\circ f): S^n_+ \to B^n \to S^n_+$$

$$(g\circ f)(x_1,\ldots,x_n,x_{n+1}) = g(x_1,\ldots,x_n) = \left(x_1,\ldots,x_n,\sqrt{1-x_1^2-\cdots-x_n^2}\right) = (x_1,\ldots,x_n,x_{n+1}) = \mathrm{id}_{S^n_+}$$

$\Longrightarrow S^n_+ \cong B^n$.

2 Programming problems

2.1 Deciding connectivity

Let G be a simple graph with n vertices and m edges. We write a simple algorithm that returns the connected components of a simple graph. All function that we used are in the attached file `graphcomponents.py`.
Algorithm input:

- a list $V = [1,2,\ldots,n]$ of n vertices,
- a list of m $2-$tuples $E = [(v_1,v_2),\ldots]$ that represent the m edges.

Algorithm output:

- a list $[C_1,C_2,\ldots,C_k]$ of all components, where each component C_1 is a list of vertices $[v_1,v_2,\ldots,v_k]$

We get the connected components of a graph using DFS (Depth Forst Search) algorithm. We could also use BFS algorithm and will get the same result with same time complexity ($\mathrm{O}(|V|+|E|)$). DFS is an algorithm for graph "research", generally used for connectivity questions – in our case to determine connected components. We first select (any random) vertex to start and then explore as far as possible in a branch and then come back to a fixed point. We keep track if we have visited the vertices connected to it. When we search the graph, we use a stack with LIFO (last in first out) feature. We also keep a list of all the vertices we have visited

since we have to visit each vertex only once. So we will add a vertex to the stack only if it has not been visited. With visiting a particular vertex, we remove it from the stack. Finally, we'll end up visiting all the vertices and then the stack will be empty.
In the attached code, the input data is pre-processed into a dict (`makeDictGraph(V, E)`), where the keys are all of the vertices and the values are all vertices that are connected to the key vertex. With that output and dsf algorithm function (`dfs(graph, start)`) we calculate the connected components of a graph in main function `findComponents(V, E)`.
Let's present the outputs of our algorithm for few examples.

Example 1:
Input:
`V = [1,2,3,4,5,6,7,8,9]`
`E = [(1,2),(1,3),(1,8),(3,7),(4,5),(4,6),(4,9),(5,6),(5,9),(7,8)]`
Output:
`[[1, 2, 3, 7, 8], [4, 5, 6, 9]]`

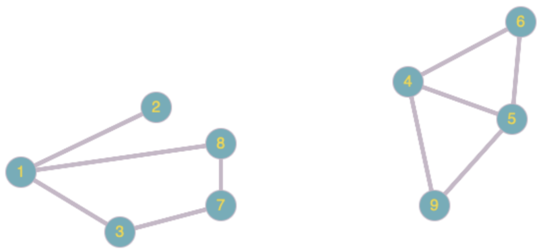

Figure 20: Visualization of graph from example 1.

Example 2:
Input:
`V = [1,2,3,4,5,6,7,8,9,10,11,12]`
`E = [(4,5),(5,4),(1,2),(1,6),(2,1),(2,3),(2,6),(3,2),(3,9),(6,1),(6,2),(6,8),`
`(8,6),(8,9),(9,3),(9,8),(10,11),(11,10),(11,12),(12,11)]`
Output:
`[[1, 2, 3, 6, 8, 9], [4, 5], [7], [10, 11, 12]]`

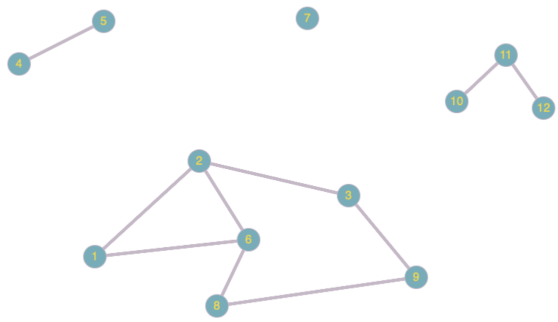

Figure 21: Visualization of graph from example 2.

Example 3:

Input:

```
V = [1,2,3,4,5,6,7,8,9,10,11,12, 13, 14, 15]
E = [(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),(2,3),(3,1),(3,2),(3,4),(3,5),(4,1),
(4,3),(4,7),(5,1),(5,3),(5,6),(5,7),(6,1),(6,5),(6,7),(7,4),(7,5),(7,6),(8,9),
(9,8),(9,10),(9,11),(10,9),(11,9),(12,13),(13,12)]
```

Output:

```
[[1, 2, 3, 4, 5, 6, 7], [8, 9, 10, 11], [12, 13], [14], [15]]
```

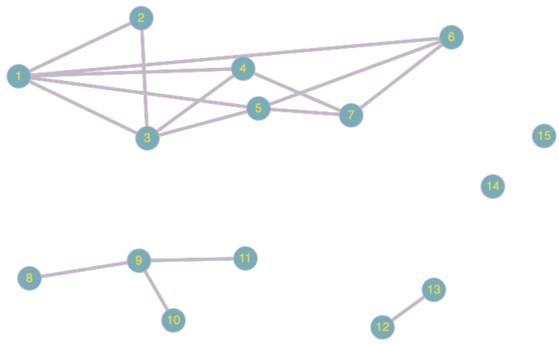

Figure 22: Visualization of graph from example 3.

2.2 Shelling disks