The upper half space $\mathbb{R}^3_{**+}$ is the set of points $(a,b,c) \in \mathbb{R}^3$ with $c > 0$. We write (x,y,z) for the standard coordinate functions on the upper half space,

$$(\ x[(a,b,c)] \ , \ y[(a,b,c)] \ , \ z[(a,b,c)] \,) = (a,b,c).$$

This can also be written in the form $(x,y,z) = I$, where I is the identity function on the upper half space.

The stereographic projection $\boldsymbol{P}$ maps the upper half space to itself. We will give its simple definition.

When Z is a positive real number, we use the notation $S[Z]$ for the sphere in the upper half space which sits on the origin $(0,0,0)$ and has radius Z. Its center is at the point $(0,0,Z)$. The only point of $S[Z]$ which is not in the upper half space is $(0,0,0)$.

The formula of $S[Z]$ is $x^2 + y^2 + (z - Z)^2 = Z^2$. This can be simplified to

$$\frac{x^2 + y^2 + z^2}{2z} = Z \, .$$

The left side is well defined on the upper half space. To make certain to have the right geometric picture, we emphasize that the center and equator of the sphere $S[Z]$ are on the plane $z = Z$.

It makes sense to define r to be the length of (x, y, z) given by $r = \sqrt{x^2 + y^2 + z^2}$ so $r^2 = x^2 + y^2 + z^2$. Then the formula of $S[Z]$ becomes

$$\frac{r^2}{2z} = Z.$$

Exercise. Show that $sS[Z] = S[sZ]$. In other words, $S[Z]$ scaled by the factor $s > 0$, equals $S[sZ]$.

It is clear that each point in the upper half space is on the sphere $S[Z]$ for precisely one value of the radius $Z > 0$, because Z is uniquely determined by the coordinates (x, y, z).

Indeed, the point (a, b, c) in the upper half space is on the sphere $S[Z]$ with

$$Z = \frac{a^2 + b^2 + c^2}{2c}.$$

The stereographic projection $\boldsymbol{P}$ takes (a, b, c) to the point of intersection of the straight ray from $(0,0,0)$ to (a, b, c), with the plane $z = Z$.

Lets compute the coordinates of the stereographic projection of (a, b, c). Since the ray can be parameterized as (ta, tb, tc) with $t > 0$, we obtain the intersection with the plane $z = Z$ by plugging in $tc = z = Z$, yielding the intersection point

$$\boldsymbol{P}[(a,b,c)] = \frac{Z}{c}(a,b,c) = \frac{a^2+b^2+c^2}{2c^2}(a,b,c)\,. \quad (1)$$

In particular $\boldsymbol{P}[(a,b,c)]$ exists in the upper half space.

$\boldsymbol{P}[(a,b,c)]$ is a positive scalar multiple of (a,b,c).

Define the components of $\boldsymbol{P}$ by $(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z}) = \boldsymbol{P}$. We can evaluate these component functions by expanding (1):

$$\boldsymbol{P}[(a,b,c)] = (\boldsymbol{X}[(a,b,c)]\,,\boldsymbol{Y}\,[(a,b,c)],\boldsymbol{Z}[(a,b,c)]) =$$

$$\left(\frac{(a^2+b^2+c^2)a}{2c^2},\frac{(a^2+b^2+c^2)b}{2c^2},\frac{(a^2+b^2+c^2)c}{2c^2}\right).$$

Alternatively, we can rewrite (1) by defining

$$g[(a,b,c)] = \frac{a^2+b^2+c^2}{2c^2},$$

So that $\boldsymbol{P}[(a,b,c)] = g[(a,b,c)]\ (a,b,c)$.

Let's summarize our formulas for $\boldsymbol{P}$ by writing them in terms of the coordinate functions.

The stereographic projection on the upper half space is

$$\boldsymbol{P} = \frac{x^2+y^2+z^2}{2z^2}(x,y,z).$$

$$\boldsymbol{P} = (\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z}) = \left(\frac{r^2x}{2z^2}\ ,\ \frac{r^2y}{2z^2}\ ,\ \frac{r^2z}{2z^2}\right)$$

$$\boldsymbol{P} = g \cdot (x,y,z)$$

$$\boldsymbol{P} = g \cdot I$$

Where $r^2 = x^2+y^2+z^2$ and we define

$$g = \frac{x^2+y^2+z^2}{2z^2} = \frac{r^2}{2z^2}.$$

We included the dot in the expression $\boldsymbol{P} = g \cdot (x,y,z)$ to emphasize multiplication. When we evaluate $\boldsymbol{P}$ at (a,b,c), the function g produces the real number $g[(a,b,c)]$ and the standard coordinates (x,y,z) are the identity and just produce the point (a,b,c) itself.

The output of the function $\boldsymbol{P} = (\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ at any point (a, b, c) in the upper half space, is a set of three new coordinates. It is however a question whether $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ does indeed provide a system of curvilinear coordinates on the upper half space. To answer this question we will need to check that $\boldsymbol{P}$ is one-to-one and onto, so it has an inverse. Moreover, in order to use these coordinates for differential calculations, we will need to check that $\boldsymbol{P}$ and its inverse are smooth, which just means that the components have partial derivatives of all orders.

Exercise. Show that $\boldsymbol{P}$ is smooth by showing that the components

$$(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}) = \ (gx\ ,\ \ gy\ ,\ \ gz), \qquad g = \frac{x^2 + y^2 + z^2}{2z^2},$$

have partial derivatives of all orders with respect to the coordinate functions (x, y, z), on the upper half space.

Exercise. Notice that the formula for g does not extend continuously to the plane $z = 0$. The mapping $\boldsymbol{P}$ does not extend continuously either. Explain geometrically what stereographic projection does to planes $z = \varepsilon$, as $\varepsilon \to 0$. This exercise will be easier later on.

Now that we have the formulas for $\boldsymbol{P}$, it is a question whether we want to think of the output $\boldsymbol{P}[(a,b,c)]$ as being in the same space as the point (a,b,c). There is another way of thinking often called the dual interpretation, which is that the function $(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z})$ takes coordinates from the upper half space to coordinates in a separate new copy of the upper half space. Thinking of (a,b,c) and $\boldsymbol{P}[(a,b,c)]$ as being in separate copies of the upper half space is just a picture and does not show up in our mathematical formulas. Although we could label the outputs with a tag to say which function they came from, we do not bother to do this. We can imagine (a,b,c) and $\boldsymbol{P}[(a,b,c)]$ to be in separate copies of the upper half space and join them by an arrow, in order to more easily visualize $(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z})[(a,b,c)]$ as giving us new coordinates for (a,b,c), without becoming confused by the fact that we are assigning two different sets of coordinates to each point.

The picture of having different ranges for different coordinate functions is common in the theory of manifolds. On a manifold there is an ambient geometry.

It can be described in different coordinates and is given by different formulas in the different coordinates. For example, for a curved surface in space we can picture the coordinates either as a grid of curves on the surface, or as the standard Euclidean coordinates on the plane with the geometry invisible and encoded in formulas. One case where this becomes trivial is when we are considering the standard coordinates on Euclidean space. It is not necessary for us to visualize (x, y, z) as mapping the upper half space to a separate copy of the upper half space, since the map is just the identity map and does not create an ambiguity in the names of the points.

Let's check whether $\boldsymbol{P} = (\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ given by

$$\boldsymbol{P} = g \cdot (x, y, z), \qquad g = \frac{x^2 + y^2 + z^2}{2z^2} = \frac{r^2}{2z^2}$$

is indeed a curvilinear coordinate system. We still need to check that $\boldsymbol{P}$ is one-to-one and onto and that the inverse is smooth. For a point (a, b, c) in the upper half space, set

$$(A, B, C) = \boldsymbol{P}[(a, b, c)] = g[(a, b, c)] \ (a, b, c).$$

The function g is non-vanishing. Hence the inverse should just be obtained just by dividing by $g[(a,b,c)]$.

$$(a,b,c) = \frac{1}{g[(a,b,c)]}\,(A,B,C)\,.$$

However, it looks as though we need to know (a,b,c) to compute the scale factor $1/g[(a,b,c)]$, whereas the inverse is supposed to compute (a,b,c) from (A,B,C). This problem is easily remedied because

g is scale invariant meaning that for every $t > 0$ and point (a,b,c) in the upper half plane,

$$g[(ta,tb,tc)] = g[(a,b,c)].$$

Equivalently,

$$\frac{(ta)^2+(tb)^2+(tc)^2}{2(tc)^2} = \frac{a^2+b^2+c^2}{2c^2}.$$

Now we note that, (A,B,C) is a scalar multiple of (a,b,c). The scale factor happens to be $g[(a,b,c)]$. By the scale invariance, g has the same value at (a,b,c) and (A,B,C), and so $g[(a,b,c)] = g[(A,B,C)]$. Thus

$$(a,b,c) = \frac{1}{g[(A,B,C)]}\ (A,B,C)\ .$$

In summary, we used

$$g[(a,b,c)] = \frac{a^2+b^2+c^2}{2c^2} = \frac{A^2+B^2+C^2}{2C^2}$$
$$= \ g[(A,B,C)],$$

To obtain

$$(a,b,c) = \frac{1}{g[(a,b,c)]}\ (A,B,C)$$
$$= \ \frac{2C^2}{A^2+B^2+C^2}\ (A,B,C).$$

However, this uniquely recovers the point (a,b,c) from the value (A,B,C) of $\boldsymbol{P}$ and so $\boldsymbol{P}$ is one-to=one.

Now conversely, for any choice of (A,B,C) in the upper half space, set

$$(a,b,c) = \frac{2C^2}{A^2+B^2+C^2}\ (A,B,C).$$

Then $\quad (A,B,C) = \dfrac{A^2+B^2+C^2}{2C^2}(a,b,c),$

$$= \frac{a^2+b^2+c^2}{2c^2}(a,b,c) = \boldsymbol{P}[(a,b,c)].$$

Where again, we used the scale invariance of g. We have found a point (a, b, c) in the upper half space, with $(A, B, C) = \boldsymbol{P}[(a, b, c)]$, and so $\boldsymbol{P}$ is onto.

This completes the proof that $\boldsymbol{P}$ one-to-one, and onto, and is thus invertible.

Geometrically, the fact that g is scale invariant means that it is constant on rays through the origin. From this we see that the map $\boldsymbol{P}$ is acting just by multiplying the points of each ray by a uniform positive scale factor. Such a scaling is one-to-one and onto for each individual ray. Thus $\boldsymbol{P}$ it is one-to-one and onto on the whole upper half space because the upper half space is the disjoint union of the rays.

We have exhibited the inverse formula

$$(a, b, c) = \frac{2C^2}{A^2 + B^2 + C^2}\,(A, B, C).$$

Denoting the inverse of $\boldsymbol{P}$ by Q, for (A, B, C) in the upper half space,

$$Q[(A, B, C)] = \frac{2C^2}{A^2 + B^2 + C^2}\,(A, B, C).$$

This is the point where we come to an unpleasant realization, which is that in order to write Q in terms of coordinate functions we need to plug the point (A, B, C) which was an output of $\boldsymbol{P}$, into the function Q. We can write Q in terms of the coordinate functions.

$$Q = \frac{2z^2}{x^2 + y^2 + z^2}(x, y, z) = \frac{1}{g} \cdot (x, y, z).$$

Indeed, we can check that this formula works for (A, B, C).

Exercise. Show that the components of Q have partial derivatives of all orders, so Q is smooth. This completes the proof that $\boldsymbol{P}$ is a smooth coordinate map from the upper half space to the upper half space

Exercise. Define a function f on the slab $\{(a, b, c) :$
$1 \leq c \leq 2\}$, by $f[(a, b, c)] = (3 - \mathrm{c}) \cdot (a, b, c)$.

Dividing both sides of this equation by $(3 - c)$, gives

$$(a, b, c) = \frac{1}{(3 - \mathrm{c})} \cdot f[(a, b, c)].$$

Does this prove the map f is invertible and produce an inverse?

As a subtle point for the enthusiast, suppose we had been given the formula

$$\boldsymbol{P} = \frac{x^2 + y^2 + z^2}{2z^2}(x, y, z)$$

And been told that whilst (x, y, z) are coordinate functions, they might be curvilinear, whereas the standard coordinate functions are actually (u, v, w). We would have gone through all our arguments using the coordinates (x, y, z), and our conclusion would still be that

$$Q = \frac{2z^2}{x^2 + y^2 + z^2}(x, y, z)$$

Gets us back from the output of $\boldsymbol{P}$ to the (x, y, z) coordinates of a point. It does not however get us to the standard coordinates, which would necessitate us composing with the inverse of the coordinate map (x, y, z).

We perhaps need more warnings against this kind of ambiguity. For example, now we have the curvilinear coordinates $\boldsymbol{P}$, we always need to keep track of whether we are working with the standard coordinates or the coordinates $\boldsymbol{P}$. For example, the sphere $S[Z]$ with the origin $(0,0,0)$ deleted, which is defined for each fixed positive value of Z, has the formula $r^2/2z = Z$. This formula can be expressed in terms of the new coordinate functions $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ as $\boldsymbol{Z} = Z$. We need to agree that this is not a plane, but a sphere with a point deleted. The problem occurs when we try to describe a surface without being clear whether we are talking about the surface itself or the representation of the surface in non-standard coordinates. As we just noted,

$\{ (a,b,c) : \boldsymbol{Z}[(a,b,c)] = Z \}$,

Is a sphere in the upper half space, whereas

$\{ (\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})[(a,b,c)] : \boldsymbol{Z}[(a,b,c)] = Z \}$

is its image $\boldsymbol{P}[S[Z]]$, which is a horizontal plane in the upper half space. Thus we refer to $\boldsymbol{P}[S[Z]]$ as the $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ coordinates of the surface $\boldsymbol{Z} = Z$. We refer to

$$\{ (a,b,c) : \boldsymbol{Z}[(a,b,c)] = Z \}$$

as just the surface $\boldsymbol{Z} = Z$, or if we want to emphasize the names of the coordinates it is the (x,y,z) coordinates of the surface $\boldsymbol{Z} = Z$. Both the plane

$\boldsymbol{P}[S[Z]]$ and the sphere $S[Z]$ minus the origin, are surfaces in the upper half space. We cannot differentiate between them by them existing in different spaces, because they exist in the same space.

In summary, if we have a subset of space we can try to describe it in terms of specific coordinates, and it makes sense to clearly identify any functions we use.

For example, we already mentioned at least twice that the sphere $S[Z]$ is the (x, y, z) coordinates of $gz = Z$, where $g = (x^2 + y^2 + z^2)/(2z^2)$.

For any real number $t > 0$, define $C[t]$ to be the (x, y, z) coordinates of the level set $g = t$. $C[t]$ is the cone

$$x^2 + y^2 = (2t - 1)z^2.$$

In particular, $C[1]$ is the cone $x^2 + y^2 = z^2$. The reason we get a cone is because g is scale invariant as we proved earlier.

$$g[(sa, sb, sc)] = g[(a, b, c)].$$

However, the $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ coordinates of the the level set $g = t$ is the image under the map $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ of the points (x, y, z), the cone $C[t]$. The scale invariance of g implies that a point is on $C[t]$ if and only if

$$g[(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})(a, b, c)] = g[(a, b, c)] = g[(a, b, c)\ ] = t.$$

Thus $C[t]$ is also the a level set of $g \circ \boldsymbol{P}$ where there small circle denotes composition of operators, and is given by

$$\boldsymbol{X}^2 + \boldsymbol{Y}^2 = \ (2t - 1)\boldsymbol{Z}^2.$$

This is one occasion where we cannot become confused.