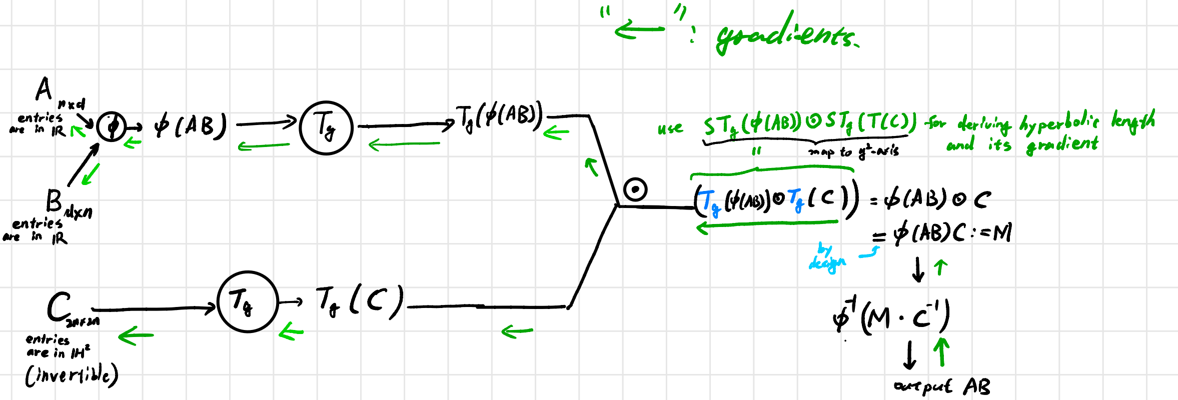
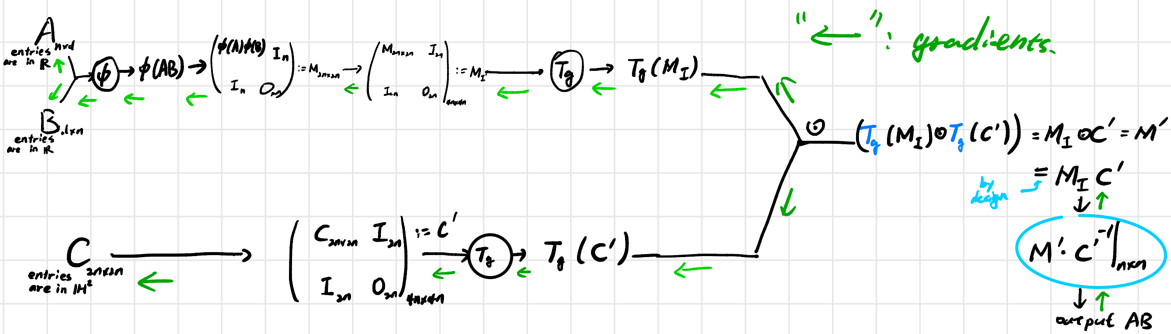
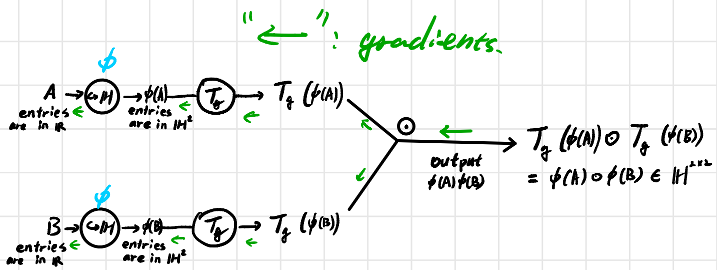
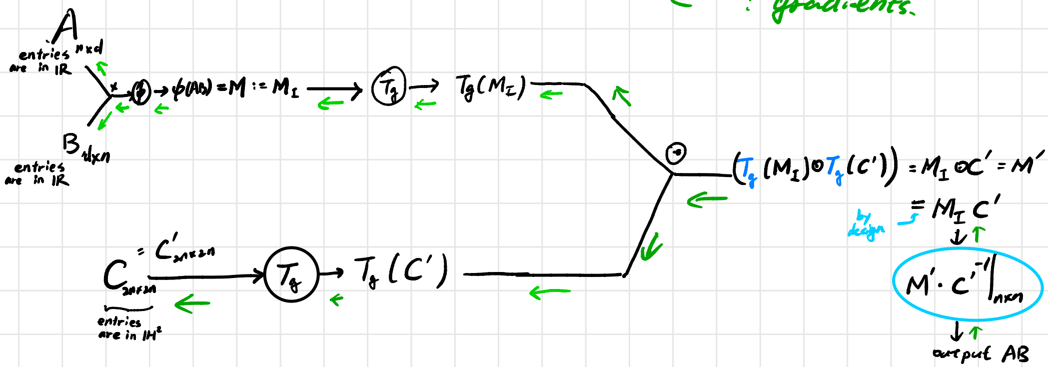


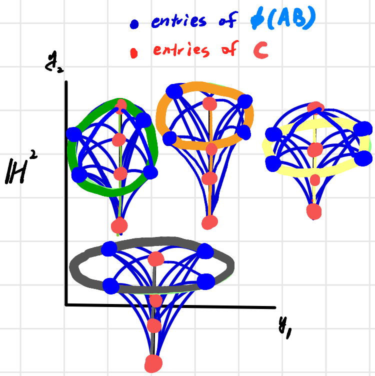


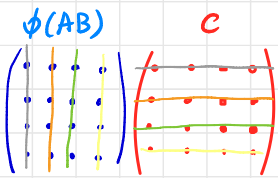


Case 2. Method 3.

"←": gradients.



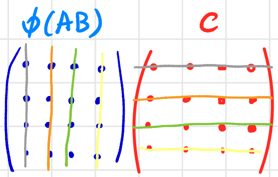




$= \phi(AB)\, C$

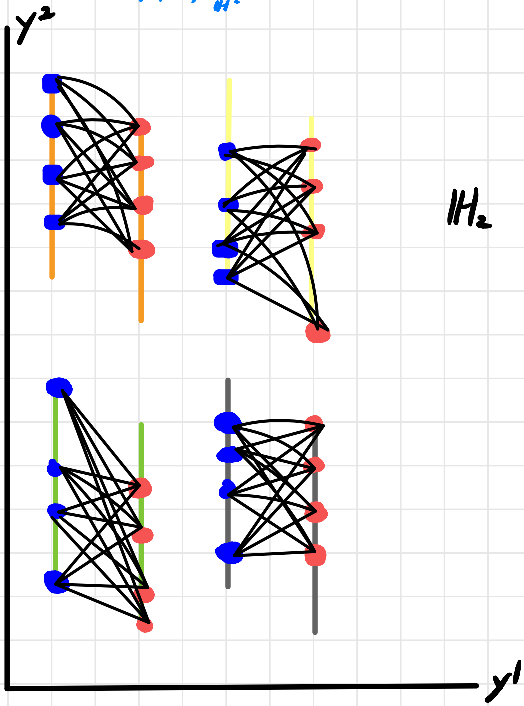
$= \phi(AB) \odot_{\mathbb{H}^2} C$

1. map M_I' entries to $\mathbb{H}^2$ randomly.
2. randomly general M'' (invertible and real)
3. $C' := M_I'^{-1} \cdot M''$



$$= \phi(AB)\, C$$

$$= \phi(AB) \odot_{\mathbb{H}^2} C$$



$x \in T_m(M_I')$ $\quad$ $x, y \in \mathbb{H}^2$ and they are entries of $T_m(M_I')$

$y \in T_m(C')$ $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad T_m(C')$.

$$d_H(x,y) = \underbrace{d_H(S(x), S(y))}_{\text{on } y_2\text{-axis}} = \ln\left(\frac{S(x)}{S(y)}\right).$$

$$\frac{\partial d_H}{\partial y} = \frac{\partial}{\partial y}\ln\left(\frac{Sx}{Sy}\right)$$

$$= \frac{Sy}{\cancel{Sx}}\frac{(-S\cancel{x})\left(\frac{\partial S}{\partial y}\right)}{(Sy)^2}$$

$$= \frac{-1}{Sy}\frac{\partial S}{\partial y} = \frac{-1}{Sy}\cdot\frac{1}{(cy+d)^2} = \boxed{\frac{-1}{(ay+b)(cy+d)}}$$

$$\frac{\partial d_H}{\partial x} = \frac{Sy}{Sx}\frac{\frac{\partial S}{\partial x} - 0}{(Sy)^2}$$

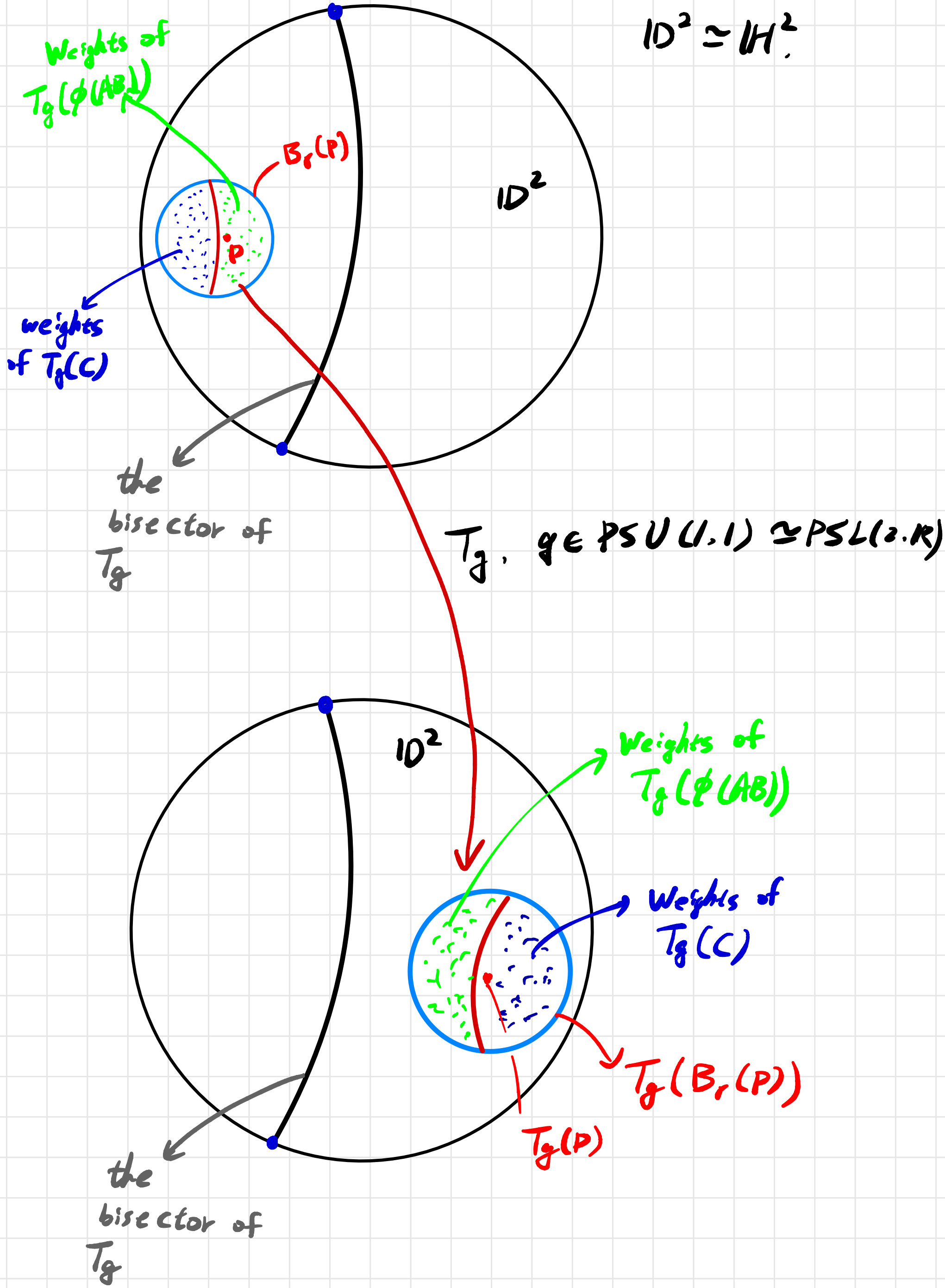
$$= \frac{1}{Sx\,Sy}\cdot\frac{\partial S}{\partial x} = \frac{1}{SxSy}\frac{1}{(cx+d)^2} = \frac{1}{\left(\frac{ax+b}{cx+d}\right)\left(\frac{ay+b}{cy+d}\right)(cx+d)^2}$$

$$= \boxed{\frac{(cx+d)}{(ax+b)(ay+b)(cx+d)}}$$

$$Sx = \frac{ax+b}{cx+d} \qquad \frac{\partial S}{\partial x} = \frac{a(\cancel{cx}+d) - c(\cancel{ax}+b)}{(cx+d)^2}$$

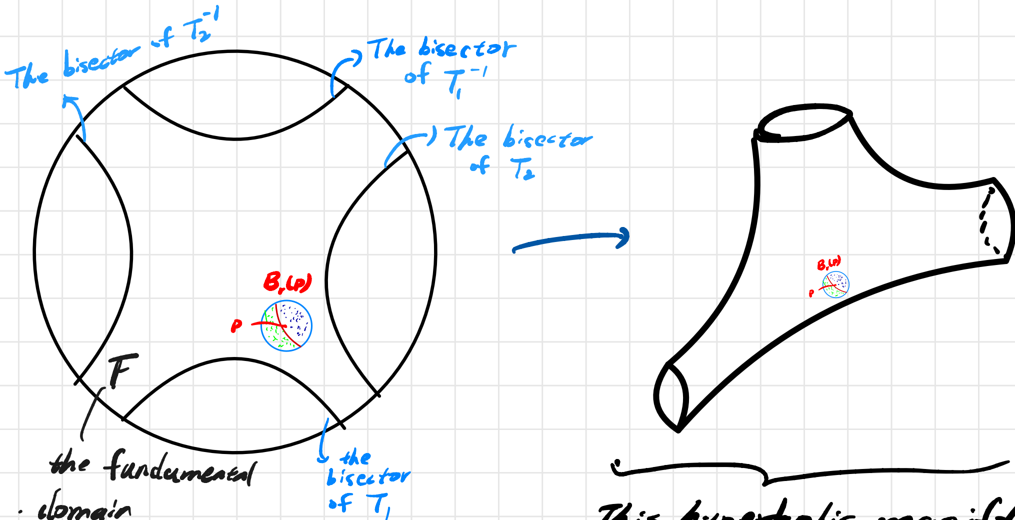
$$Sy = \frac{ay+b}{cy+d}. \qquad = \frac{ad-bc}{(cx+d)^2} = \frac{1}{(cx+d)^2}$$

$$= \frac{1}{(cy+d)^2}$$



$\mathbb{D}^2 \simeq \mathbb{H}^2$.
Weights of $T_g(\phi(AB))$
$B_r(P)$
P
$\mathbb{D}^2$
weights of $T_g(C)$
the bisector of T_g
T_g, $g \in PSU(1,1) \simeq PSL(2,\mathbb{R})$
$\mathbb{D}^2$
Weights of $T_g(\phi(AB))$
Weights of $T_g(C)$
$T_g(B_r(P))$
$T_g(P)$
the bisector of T_g

Assume $T_g = T_1 \circ T_2$. Let T_1 and T_2 have the following configuration on their bisectors. Consider a subgroup of $PSU(1,1) \simeq PSL(2,\mathbb{R})$ that is generated by T_1 and T_2: $\langle T_1, T_2 \rangle$.



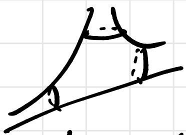
This hyperbolic manifold is corresponding to a point in its moduli space.

Using elements in $\langle T_1, T_2 \rangle$ to map $B_r(p)$ to a new place in $\mathbb{F}$. to generate an equivalence of a given converged model. Using $B_r(p)$ as a device to detect the property of its corresponding loss surface.

Different configurations can give different hyperbolic manifolds. Examples are as follows:

Let M be a 2d Riemannian (hyperbolic) manifold with curvature -1.

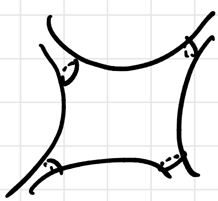
For $\mathrm{Vol}(M) = 2\pi$:

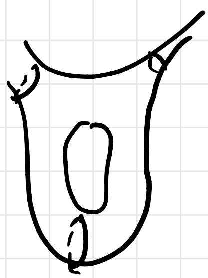


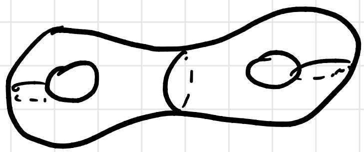
sphere with 3 punctures

torus with a puncture.

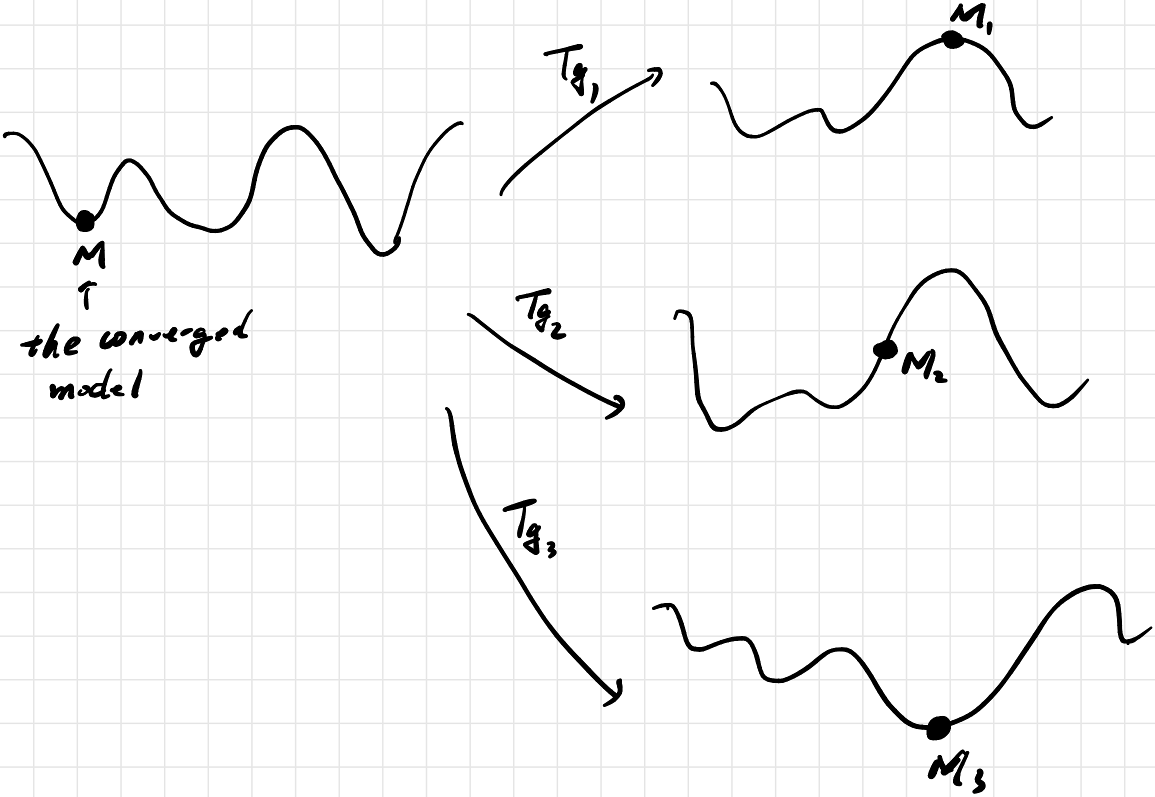
For $\mathrm{Vol}(M) = 4\pi$:







Possible scenarios of loss surfaces:



Each possible scenario is corresponding to different location of $B_r(p)$ in the hyperbolic space and manifold.