Poincaré ResNet
Supplementary material

Max van Spengler[1], Erwin Berkhout[2], Pascal Mettes[1]

[1] VIS Lab, Informatics Institute, University of Amsterdam,
[2] Department of Oral Radiology, Academic Center for Dentistry, University of Amsterdam & VU Amsterdam

1. Gradient formulations

Here we provide the formulations of the manually derived gradient expressions which are used for backpropagation. For each, we provide the Jacobians ($J.$) with respect to its input variables, left-multiplied by the gradient ($\boldsymbol{u}$) of its output.

1.1. Möbius addition

The Möbius addition operation is defined as

$$\boldsymbol{x} \oplus_c \boldsymbol{y} = \frac{(1 + 2c\langle\boldsymbol{x}, \boldsymbol{y}\rangle + c||\boldsymbol{y}||^2)\boldsymbol{x} + (1 - c||\boldsymbol{x}||^2)\boldsymbol{y}}{1 + 2c\langle\boldsymbol{x}, \boldsymbol{y}\rangle + c^2||\boldsymbol{x}||^2||\boldsymbol{y}||^2}. \tag{1}$$

Its Jacobians, left-multiplied by the output gradient, can be written as

$$\boldsymbol{u}^T J_{\boldsymbol{x}}(\boldsymbol{x} \oplus_c \boldsymbol{y}) = \frac{a}{d}\boldsymbol{u}^T - \frac{2c}{d}\Big(\boldsymbol{u}^T\boldsymbol{y} + \frac{\theta c||\boldsymbol{y}||^2}{d}\Big)\boldsymbol{x}^T + \frac{2c}{d}\Big(\boldsymbol{u}^T\boldsymbol{x} - \frac{\theta}{d}\Big)\boldsymbol{y}^T, \tag{2}$$

$$\boldsymbol{u}^T J_{\boldsymbol{y}}(\boldsymbol{x} \oplus_c \boldsymbol{y}) = \frac{b}{d}\boldsymbol{u}^T + \frac{2c}{d}\Big(\boldsymbol{u}^T\boldsymbol{x} - \frac{\theta}{d}\Big)\boldsymbol{x}^T + \frac{2c}{d}\Big(\boldsymbol{u}^T\boldsymbol{x} - \frac{c||\boldsymbol{x}||^2\theta}{d}\Big)\boldsymbol{y}^T, \tag{3}$$

where

$$a = 1 + 2c\langle\boldsymbol{x}, \boldsymbol{y}\rangle + c||\boldsymbol{y}||^2, \tag{4}$$

$$b = 1 - c||\boldsymbol{x}||^2, \tag{5}$$

$$d = 1 + 2c\langle\boldsymbol{x}, \boldsymbol{y}\rangle + c^2||\boldsymbol{x}||^2||\boldsymbol{y}||^2, \tag{6}$$

$$\theta = a\boldsymbol{u}^T\boldsymbol{x} + b\boldsymbol{u}^T\boldsymbol{y}. \tag{7}$$

1.2. Exponential map at the origin

The exponential map at the origin is given by

$$\exp_0^c(\boldsymbol{v}) = \tanh(\sqrt{c}||\boldsymbol{v}||)\frac{\boldsymbol{v}}{\sqrt{c}||\boldsymbol{v}||}. \tag{8}$$

Its Jacobian, left-multiplied by the output gradient, can be written as

$$u^T J_{\boldsymbol{v}} \exp_0^c(\boldsymbol{v}) = \boldsymbol{u}^T\boldsymbol{v}\Big(\frac{1}{||\boldsymbol{v}||^2\cosh(\sqrt{c}||\boldsymbol{v}||)^2} - \frac{\tanh(\sqrt{c}||\boldsymbol{v}||)}{\sqrt{c}||\boldsymbol{v}||^3}\Big)\boldsymbol{v}^T + \frac{\tanh(\sqrt{c}||\boldsymbol{v}||)}{\sqrt{c}||\boldsymbol{v}||}\boldsymbol{u}^T. \tag{9}$$

1.3. Logarithmic map at the origin

The logarithmic map at the origin is given by

$$\log_0^c(\boldsymbol{y}) = \tanh^{-1}(\sqrt{c}||\boldsymbol{y}||)\frac{\boldsymbol{y}}{\sqrt{c}||\boldsymbol{y}||}. \tag{10}$$

Its Jacobian, left-multiplied by the output gradient, can be written as

$$\boldsymbol{u}^T J_{\boldsymbol{y}} \log_0^c(\boldsymbol{y}) = \boldsymbol{u}^T \boldsymbol{y}\left(\frac{1}{||\boldsymbol{y}||^2(1-c||\boldsymbol{y}||^2)} - \frac{\tanh^{-1}(\sqrt{c}||\boldsymbol{y}||)}{\sqrt{c}||\boldsymbol{y}||^3}\right)\boldsymbol{y}^T + \frac{\tanh^{-1}(\sqrt{c}||\boldsymbol{y}||)}{\sqrt{c}||\boldsymbol{y}||}\boldsymbol{u}^T. \tag{11}$$

1.4. Exponential map

The exponential map at $\boldsymbol{x}$ is defined as

$$\exp_{\boldsymbol{x}}^c(\boldsymbol{v}) = \boldsymbol{x} \oplus_c \left(\tanh\left(\frac{\sqrt{c}\lambda_{\boldsymbol{x}}^c||\boldsymbol{v}||}{2}\right)\frac{\boldsymbol{v}}{\sqrt{c}||\boldsymbol{v}||}\right), \tag{12}$$

which we can reformulate as

$$\exp_{\boldsymbol{x}}^c(\boldsymbol{v}) = \boldsymbol{x} \oplus_c \boldsymbol{z}_c(\boldsymbol{x}, \boldsymbol{v}), \tag{13}$$

where

$$\boldsymbol{z}_c(\boldsymbol{x}, \boldsymbol{v}) = \tanh\left(\frac{\sqrt{c}\lambda_{\boldsymbol{x}}^c||\boldsymbol{v}||}{2}\right)\frac{\boldsymbol{v}}{\sqrt{c}||\boldsymbol{v}||}. \tag{14}$$

Now we can backpropagate through this operation in two steps. First, the Jacobians of $\boldsymbol{z}_c$, left-multiplied by the output gradient, can be written as

$$\boldsymbol{u}^T J_{\boldsymbol{x}} \boldsymbol{z}_c(\boldsymbol{x}, \boldsymbol{v}) = \frac{2c\boldsymbol{u}^T\boldsymbol{v}}{\cosh(\frac{\sqrt{c}||\boldsymbol{v}||}{1-c||\boldsymbol{x}||^2})^2(1-c||\boldsymbol{x}||^2)^2}\boldsymbol{x}^T, \tag{15}$$

$$\boldsymbol{u}^T J_{\boldsymbol{v}} \boldsymbol{z}_c(\boldsymbol{x}, \boldsymbol{v}) = \boldsymbol{u}^T \boldsymbol{v}\left(\frac{1}{||\boldsymbol{v}||^2\cosh(\frac{\sqrt{c}||\boldsymbol{v}||}{1-c||\boldsymbol{x}||^2})^2(1-c||\boldsymbol{x}||^2)^2} - \frac{\tanh(\frac{\sqrt{c}||\boldsymbol{v}||}{1-c||\boldsymbol{x}||^2})}{\sqrt{c}||\boldsymbol{v}||^3}\right)\boldsymbol{v}^T + \frac{\tanh(\frac{\sqrt{c}||\boldsymbol{v}||}{1-c||\boldsymbol{x}||^2})}{\sqrt{c}||\boldsymbol{v}||}\boldsymbol{u}^T. \tag{16}$$

Next, for backpropagating through the Möbius addition, we can use the expressions given in equations (2, 3).

1.5. Logarithmic map

The logarithmic map at $\boldsymbol{x}$ is defined as

$$\log_{\boldsymbol{x}}^c(\boldsymbol{y}) = \frac{2}{\sqrt{c}\lambda_{\boldsymbol{x}}^c}\tanh^{-1}\left(\sqrt{c}|| - \boldsymbol{x} \oplus_c \boldsymbol{y}||\right)\frac{-\boldsymbol{x} \oplus_c \boldsymbol{y}}{|| - \boldsymbol{x} \oplus_c \boldsymbol{y}||}, \tag{17}$$

which we can reformulate as

$$\log_{\boldsymbol{x}}^c(\boldsymbol{y}) = f_c(\boldsymbol{x}, \boldsymbol{z}_c(\boldsymbol{x}, \boldsymbol{y})) = \frac{2}{\sqrt{c}\lambda_{\boldsymbol{x}}^c}\tanh^{-1}\left(\sqrt{c}||\boldsymbol{z}_c(\boldsymbol{x}, \boldsymbol{y})||\right)\frac{\boldsymbol{z}_c(\boldsymbol{x}, \boldsymbol{y})}{||\boldsymbol{z}_c(\boldsymbol{x}, \boldsymbol{y})||}, \tag{18}$$

where

$$\boldsymbol{z}_c(\boldsymbol{x}, \boldsymbol{y}) = -\boldsymbol{x} \oplus_c \boldsymbol{y}. \tag{19}$$

Again, we can backpropagate through this operation in two steps. First, we backpropagate through $\boldsymbol{z}_c(\boldsymbol{x}, \boldsymbol{y})$ using equations (2, 3). Then, the Jacobians of $f_c(\boldsymbol{x}, \boldsymbol{z})$, left-multiplied by the output gradient, can be written as

$$\boldsymbol{u}^T J_{\boldsymbol{x}} f_c(\boldsymbol{x}, \boldsymbol{z}) = -\tanh^{-1}(\sqrt{c}||\boldsymbol{z}||)\frac{2c\boldsymbol{u}^T\boldsymbol{z}}{\sqrt{c}||\boldsymbol{z}||}\boldsymbol{x}^T, \tag{20}$$

$$\boldsymbol{u}^T J_{\boldsymbol{z}} f_c(\boldsymbol{x}, \boldsymbol{z}) = \boldsymbol{u}^T \boldsymbol{z}\left(\frac{1-c||\boldsymbol{x}||^2}{(1-c||\boldsymbol{z}||^2)||\boldsymbol{z}||^2} - \tanh^{-1}(\sqrt{c}||\boldsymbol{z}||)\frac{1-c||\boldsymbol{x}||^2}{\sqrt{c}||\boldsymbol{z}||^3}\right)\boldsymbol{z}^T + \tanh^{-1}(\sqrt{c}||\boldsymbol{z}||)\frac{1-c||\boldsymbol{x}||^2}{\sqrt{c}||\boldsymbol{z}||}\boldsymbol{u}^T. \tag{21}$$

1.6. Conformal factor

The conformal factor is given as

$$\lambda_{\boldsymbol{x}}^c = \frac{2}{1 - c||\boldsymbol{x}||^2}. \tag{22}$$

Its Jacobian, multiplied by the output gradient (which is a scalar here), can be written as

$$\boldsymbol{u} J_{\boldsymbol{x}} \lambda_{\boldsymbol{x}}^c = \frac{4c\boldsymbol{u}}{(1 - c||\boldsymbol{x}||^2)^2} \boldsymbol{x}^T. \tag{23}$$

1.7. Projection onto the Poincaré ball

An operation that is often applied in hyperbolic geometry, but rarely mentioned, is projection onto the Poincaré ball. This operation can be used to ensure numerical stability. It is defined as

$$\text{Proj}_c(\boldsymbol{x}) = \boldsymbol{x} \mathbb{1}_{\{c||\boldsymbol{x}||^2 < 1\}}(\boldsymbol{x}) + \frac{\boldsymbol{x}}{\sqrt{c}||\boldsymbol{x}||} \mathbb{1}_{\{c||\boldsymbol{x}||^2 > 1\}}(\boldsymbol{x}), \tag{24}$$

where $\mathbb{1}_A(\boldsymbol{x})$ is the indicator function, which is 1 if $\boldsymbol{x} \in A$ and 0 if $\boldsymbol{x} \notin A$. The Jacobian of this projection operation, left-multiplied by the output gradient, can be computed as

$$\boldsymbol{u}^T J_{\boldsymbol{x}} \text{Proj}_c(\boldsymbol{x}) = \left(\mathbb{1}_{\{c||\boldsymbol{x}||^2 < 1\}}(\boldsymbol{x}) + \frac{1}{\sqrt{c}||\boldsymbol{x}||} \mathbb{1}_{\{c||\boldsymbol{x}||^2 > 1\}}(\boldsymbol{x}) \right) \boldsymbol{u}^T - \left(\frac{\boldsymbol{u}^T \boldsymbol{x}}{\sqrt{c}||\boldsymbol{x}||^3} \mathbb{1}_{\{c||\boldsymbol{x}||^2 > 1\}}(\boldsymbol{x}) \right) \boldsymbol{x}^T. \tag{25}$$
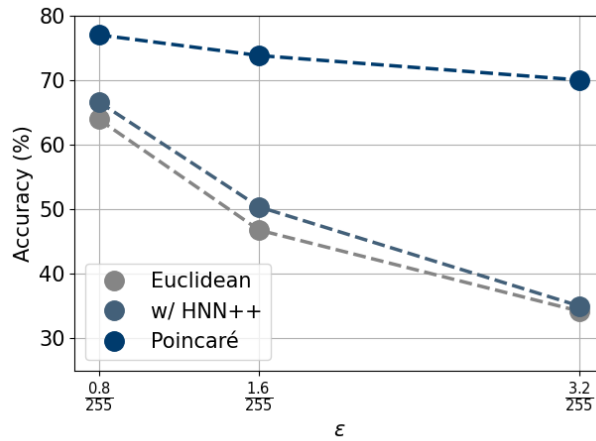
2. Adversarial attack with running statistics

Figure 1: Adversarial attack results with running statistics for Euclidean models. The running statistics make the Euclidean models perform slightly better, but significantly more susceptible to adversarial attacks.