Calculating the gradient of a cosine similarity loss function

With v, w vectors, the cosine similarity is defined as

$$\frac{v \cdot w}{\sqrt{v \cdot v * w \cdot w}}$$

Now with a matrix transformation $v = Mx$, $w = My$, we have a cosine similarity of

$$\frac{My \cdot Mx}{\sqrt{Mx \cdot Mx * My \cdot My}}$$

We want to calculate the derivatives wrt M. First we note

$$\frac{\partial v_k}{\partial M_{ij}} = \delta_{ik} x_j$$

$$\frac{\partial w_k}{\partial M_{ij}} = \delta_{ik} y_j$$

$$\frac{\partial \left(a \cdot z\right)}{\partial M_{ij}} = \sum_k \frac{\partial a_k}{\partial M_{ij}} z_k + a_k \frac{\partial z_k}{\partial M_{ij}}$$

Applying this, we have

$$\begin{aligned}
\frac{\partial \left(\frac{v \cdot w}{\sqrt{v \cdot v * w \cdot w}}\right)}{\partial M_{ij}} &= \frac{\frac{\partial (v \cdot w)}{\partial M_{ij}}}{\sqrt{v \cdot v * w \cdot w}} - \frac{v \cdot w}{\left(v \cdot v * w \cdot w\right)^{\frac{3}{2}}} \frac{\partial \left(v \cdot v * w \cdot w\right)}{\partial M_{ij}} \\
&= \frac{\sum_k \delta_{ik} x_j w_k + v_k \delta_{ik} y_j}{\sqrt{v \cdot v * w \cdot w}} - \frac{v \cdot w}{\left(v \cdot v * w \cdot w\right)^{\frac{3}{2}}} \left(\sum_k \delta_{ik} x_j v_k * w \cdot w + v \cdot v * \sum_k \delta_{ik} y_j w_k\right) \\
&= \frac{x_j w_i + v_i y_j}{\sqrt{v \cdot v * w \cdot w}} - \frac{v \cdot w}{\left(v \cdot v * w \cdot w\right)^{\frac{3}{2}}} \left(x_j v_i * w \cdot w + v \cdot v * y_j w_i\right) \\
&= \frac{x_j w_i + v_i y_j}{\left(v \cdot v * w \cdot w\right)^{3/2}} \left(v \cdot v * w \cdot w\right) - \frac{v \cdot w}{\left(v \cdot v * w \cdot w\right)^{\frac{3}{2}}} \left(x_j v_i * w \cdot w + v \cdot v * y_j w_i\right) \\
&= \frac{\left(x_j w_i + v_i y_j\right) v \cdot v * w \cdot w - v \cdot w \left(x_j v_i * w \cdot w + v \cdot v * y_j w_i\right)}{\left(v \cdot v * w \cdot w\right)^{3/2}}
\end{aligned}$$

This last formula is the one that is coded