$$\begin{bmatrix} \vdots \\ y_{d_q} \end{bmatrix}$$

We can rotate each contiguous pair of elements in $q_m$ so that we will have $d_q/2$ rotation matrices,

Henceforth, the general formulation of RoPE is

$$e^{im\theta_1}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

$$\begin{bmatrix} \cos m\theta_1 & -\sin m\theta_1 & 0 & \cdots & 0 \\ \sin m\theta_1 & \cos m\theta_1 & \vdots & & \vdots \\ 0 & 0 & \ddots & & \\ 0 & 0 & & & \\ \vdots & \vdots & \cdots & \cos m\theta_{d_q/2} & -\sin m\theta_{d_q/2} \\ 0 & \cdots & 0 & \sin m\theta_{d_q/2} & \cos m\theta_{d_q/2} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{d_q-1} \\ y_{d_q} \end{bmatrix}$$

where $\Theta = \{ \theta_i = 10000^{-2(i-1)/d},\ i \in [1, 2 \ldots d_q/2] \}$

We assign a $\theta_i$ per pair of $(y_i, y_{i+1})$

A more useful and computationally efficient version would be :

$$\begin{bmatrix} \cos m\theta_1 & -\sin m\theta_1 \\ \sin m\theta_1 & \cos m\theta_1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

$$= \begin{bmatrix} y_1 \cos m\theta_1 - y_2 \sin m\theta_1 \\ y_2 \sin m\theta_1 + y_2 \cos m\theta_1 \end{bmatrix}$$

$$= \begin{bmatrix} \cos m\theta_1 \\ \cos m\theta_1 \end{bmatrix} \underset{\text{Hadamard}}{\otimes} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + \begin{bmatrix} \sin m\theta_1 \\ \sin m\theta_1 \end{bmatrix} \otimes \begin{bmatrix} -y_1 \\ y_2 \end{bmatrix}$$

General case:

$$\begin{bmatrix} \cos m\theta_1 \\ \cos m\theta_1 \\ \vdots \\ \cos m\theta_{d/2} \end{bmatrix} \otimes \begin{bmatrix} y_1 \\ \vdots \\ y_d \end{bmatrix} + \begin{bmatrix} \sin m\theta_1 \\ \sin m\theta_1 \\ \vdots \\ \sin m\theta_{d/2} \end{bmatrix} \otimes \begin{bmatrix} -y_1 \\ y_2 \\ \vdots \\ -y_d \\ y_d \end{bmatrix}$$

This is much easier to compute and code!

# RoPE for ViT

The previous RoPE was for token embeddings for NLP. How do we extrapolate that