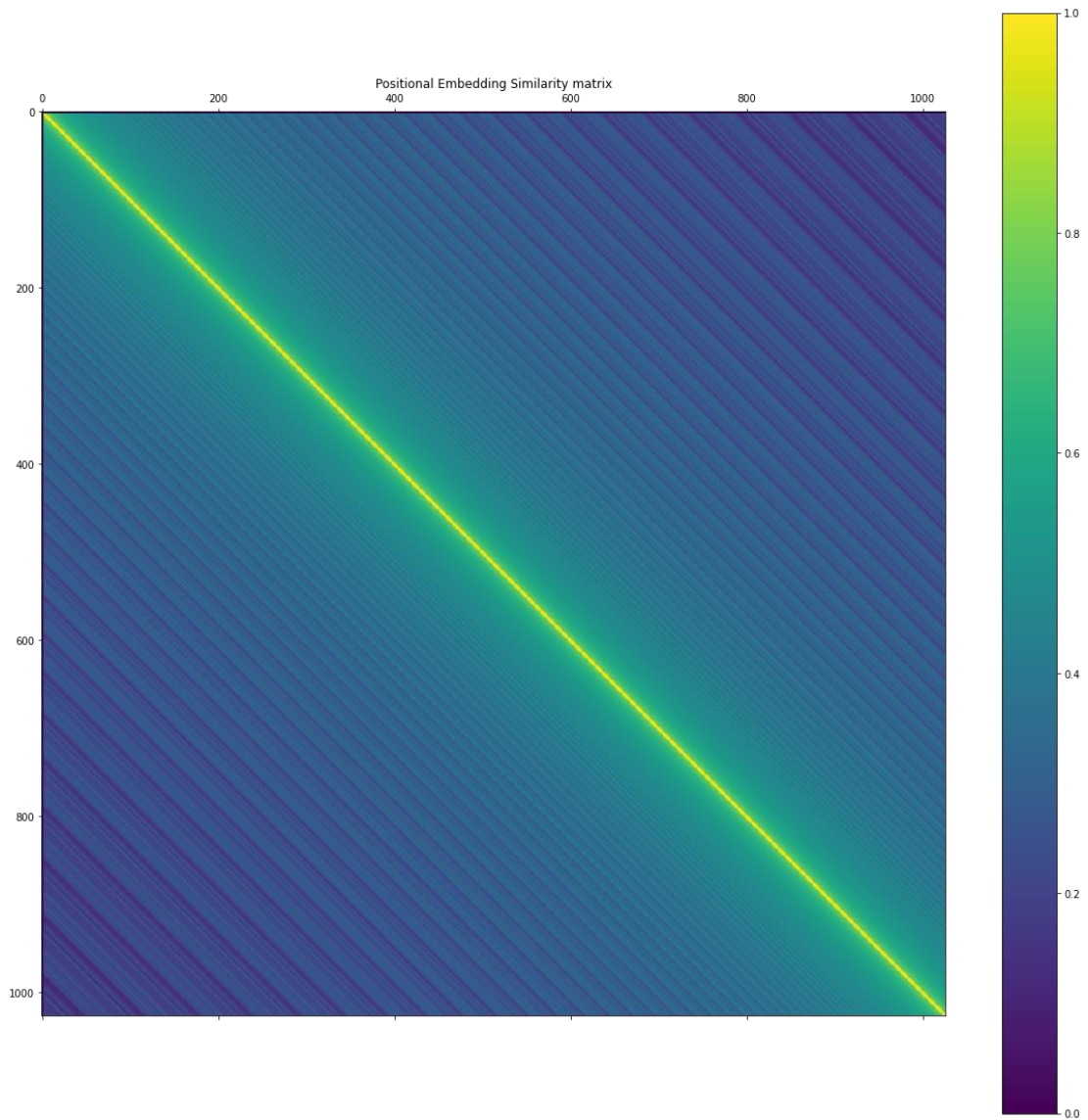● Problem 1

○ Visualize the similarity between different pairs of positional embedding and briefly explain the result.
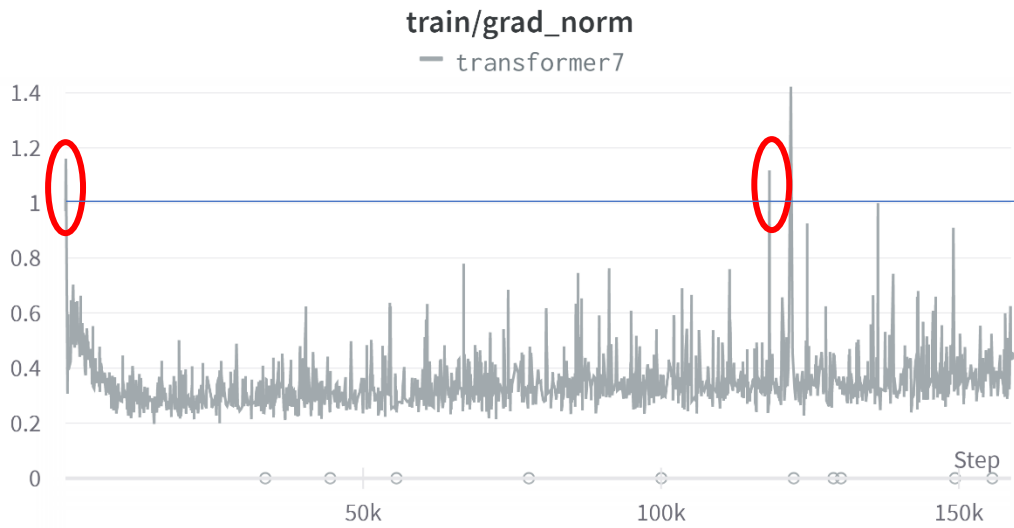

Positional Embedding Similarity matrix

由上圖可知，離自己越近，similarity 就越高，反之，離自己越遠，similarity 就越低，我想是因為語言的關係跟位置位置有關，位置相近的字，相關性就越高。

● Problem 2

○ Clip gradient norm and visualize the changes of gradient norm in different steps.
Circle two places with gradient explosion.

```
# clipping gradient norm helps alleviate gradient exploding
clip_norm=1,
```

```
gnorm = nn.utils.clip_grad_norm_(model.parameters(), config.clip_norm) # grad norm clipping prevents gradient exploding
```


train/grad_norm

clip_norm=1