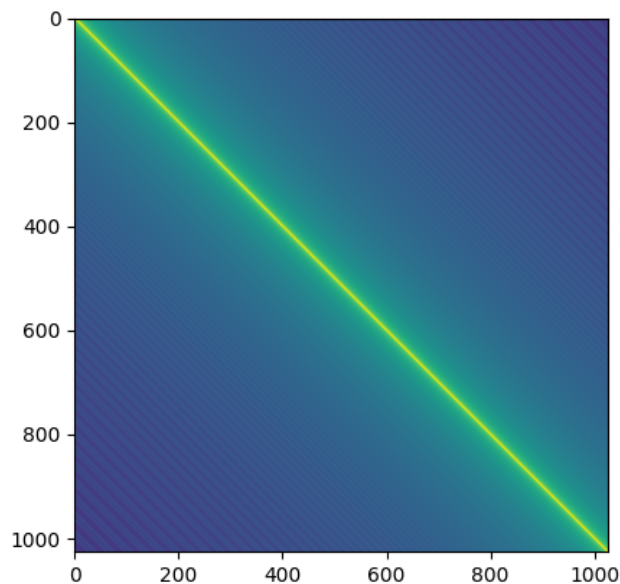


# HW5 Gradescope

B08901051 電機四 王濰紳

- Problem 1
  - Visualize the similarity between different pairs of positional embedding and briefly explain the result.



This is a figure visualizing the similarities between 1026 positional embeddings. The greener the color is, the more similar the two pairs are, the diagonal is the most green because the positional embeddings compared are themselves. As mentioned in the lecture, the positional embedding closer to each other should be more similar which is also shown as the result of above graph. The further way the positional embeddings are, the bluer the color is, indicating that the compared pair of position embeddings are more different to each other.

- Additionally, attach the code that you used for visualization.

```
1 try_load_checkpoint(model, name="avg_last_5_checkpoint.pt")
2 pos_emb = model.decoder.embed_positions.weights.cpu().detach()
3 final_matrix = np.zeros((pos_emb.shape[0], pos_emb.shape[0]), dtype=float)
4
5 print(final_matrix.shape)
6
7 for pos1 in range(pos_emb.shape[0]):
8     for pos2 in range(pos_emb.shape[0]):
9         final_matrix[pos1][pos2] = F.cosine_similarity(x1=pos_emb[pos1], x2=pos_emb[pos2], dim=0)
10
11 plt.imshow(final_matrix)
12 plt.savefig("gradescope1.png")
```

- Problem 2
  - Clip gradient norm and visualize the changes of gradient norm in different steps. Circle two places with gradient explosion.

