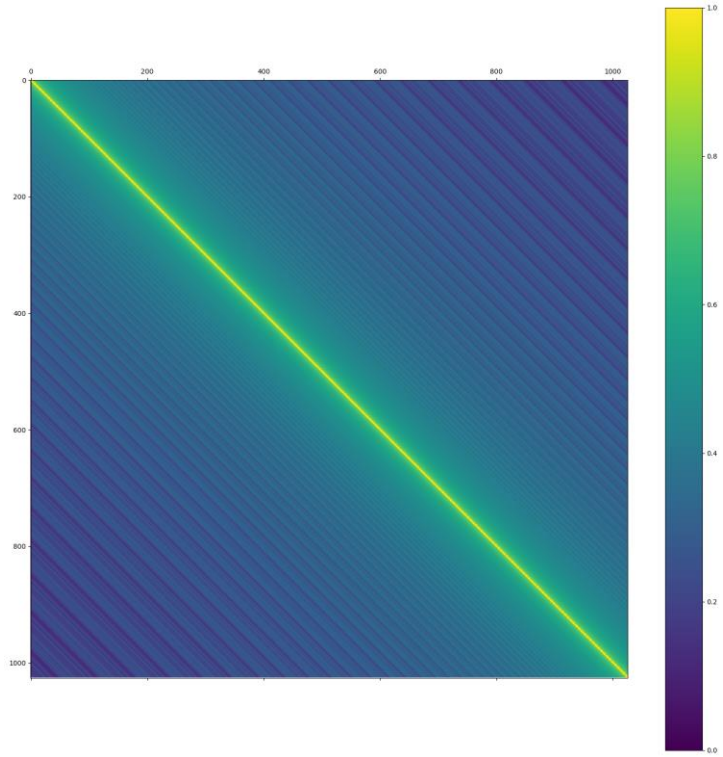
1. ◯Visualize the similarity between different pairs of positional embedding and briefly explain the result.
◯Additionally , attach the code that you used for visualization.



In [46]:
```
pos_embed = model.decoder.embed_positions.weights.cpu().detach()
print(pos_embed.shape)


M = np.zeros((1026, 1026))
for i in range(1026):
    for j in range(1026):
        M[i][j] = torch.nn.functional.cosine_similarity(pos_embed[i],pos_embed[j],dim=-1)

fig, ax = plt.subplots(figsize=(20,20))
cax = ax.matshow(M, interpolation='nearest')
fig.colorbar(cax)
plt.show()
```

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