1 Exercise Distances

Consider 50 Ca atoms starting at A102 of the main protease of SARS-COV-2 given in complex with a peptidelike inhibitor (PDB id: 6LU7). Construct the 51 × 51 Cayley-Menger matrix B.

1. Compute rank(B); explain why the obtained value is correct.

2. Perturb entries of B by 5 % (maintaining symmetry, positive entries, 0's, 1's), then explain the new valueof rank(B). Compute Gram matrix G, apply SVD: G = $U\Sigma U^T$. Let S be the diagonal matrix containingthe 3 largest singular values of G. Get the 3D coordinates as $\sqrt{S}U^T$, and report the c-RMSD against the original structure.

2 Solution

At first i import from the PDB file the CA atoms starting from the A102 position and end 50 positions after using the `get_ca_atoms()` function.

I resulted in the matrix with the following values:

-33.343	26.18	60.877
-31.916	24.031	63.672
-33.501	20.58	63.455
.	.	.
.	.	.
.	.	.
.	.	.
-24.54	18.274	58.391
-27.49	19.848	56.642
-30.535	19.417	54.444

Table 1:

The Caley-Menger matrix contains the distances between atoms. In order to make this matrix, i had to append a row and a column to the distance matrix M form ca_atoms matrix. The first row and the first column of the Caley-Menger matrix is filled with 1's except the diagonal value, which is always zero. The CM matrix is symmetric since the M matrix is symmetric. We know that $M_{ij} = \frac{1}{2} dist(p_i, p_j)^2$ (where p_i are points from the ca_atoms matrix).

0	1	.	.	1
1				
1				
.		M		
.				
1				

Table 2: Cayley-Menger matrix form

I applied the `CM_mtrx()` to create the CM matrix 1.

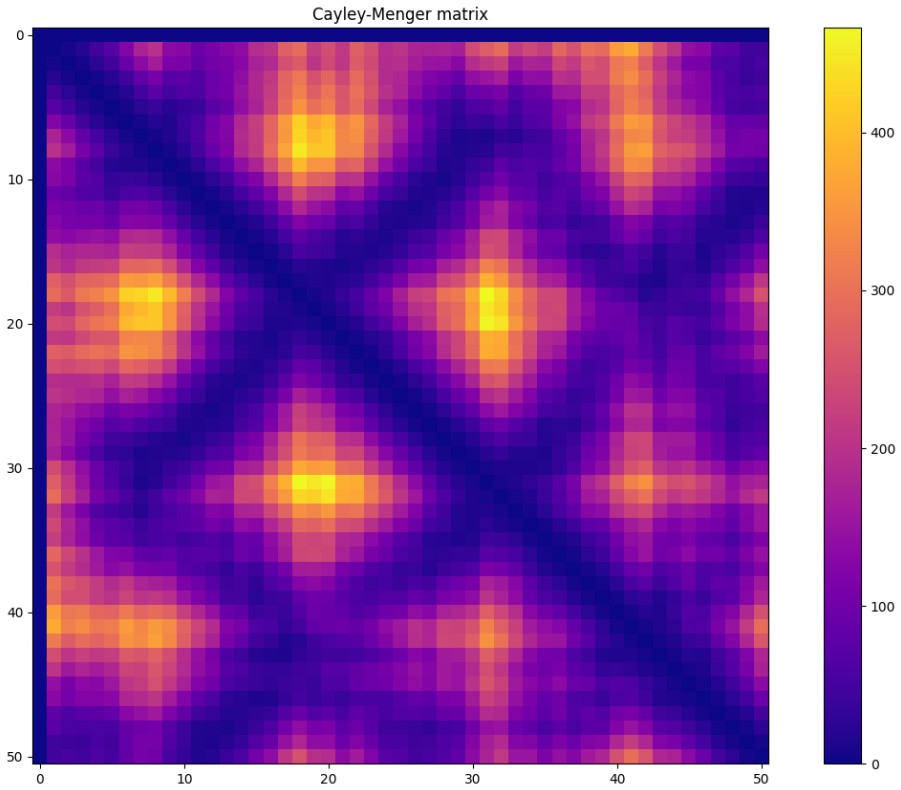

Figure 1: Cayley-Menger matrix Heatmap

2.1 Question 1

Using the `rank=np.linalg.matrix_rank(cm_matrix, tol = 0.01)` command I observed that the rank of the matrix is 5, Although if I change the tolerance and make it zero the rank becomes 51, thus is equal with the rows of the matrix, so there is an independence among all the rows. The upper score of the rank of the CM matrix is equal to the dimensions of the matrix M+1 (in our case this is 51).

2.2 Question 2

Using the `perturb()` function I added or removed a random percentage 5% or smaller. It was expected that the disturbed matrix would have a higher rank (50 with the same tolerance) than the original one as the noise made the independence of the rows more intense.

After that using the cm_matrix at the function `Gram_mtrx()`, in which I apply the formula $\frac{d_{i0}^2 - d_{ij}^2 + d_{j0}^2}{2} = G_{ij}$ for the values from the cm_matrix because they are distances. It is important to note that the Gram matrix is a nxn matrix, where n is the number of the atoms that we examined.

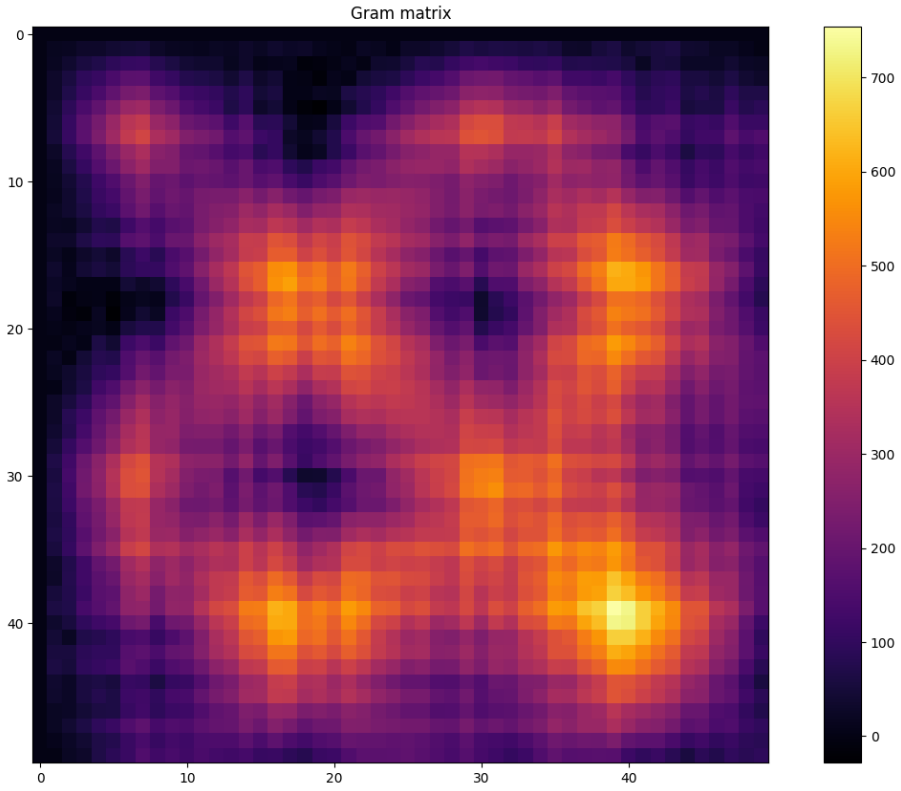

Figure 2: Gram Matrix Heatmap

Eventually using the `S_arr()` function i compute the diagonal S matrix with the biggest values of the diagonal of the SVD and also the U^T matrix, in order to find the coordinates matrix which is $coordinates = (\sqrt{S}U^T)^T$. I center the coordinates and find the cRMSD value = 8.22990846115472, between the initial matrix and the coordinates one.