Theoretical Neuroscience II
Teacher: Prof. Jochen Braun
Assistant Teacher: Ehsan Kakaei
Name: Viviana Villafañe Barraza, Tasnuba Hossain Tazeen, Lee Po Shing
Date: 25th June, 2018

Exercise 8: Principal Component Analysis (PCA)

Introduction

In many fields, multiple variables under different conditions can be observed containing data sets having high-dimensions. In order to reduce the dimensionality of a large data set and potentially reveal a simpler structure hiding within the data set, a statistical tool called Principal component analysis (PCA) can be used. PCA uses an orthogonal transformation to convert these set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the next highest variance possible under the constraint that it is orthogonal to the preceding components.

In this assignment, the activity of a network of spiking neurons was recorded in the form of multi-unit activities. This network had dormant periods that were interrupted by transient periods with the neurons showing collective bursts. So the origin of these bursts was the focus and thus PCA was used to analyze the activity of this network.

Methods

In this exercise, a Matlab file called *MUA_b_t_g.mat* containing multi-unit recordings was provided. This file had the number of bursts (Nb in the code), the number of time points (Nt in the code), the number of neurons groups (Ng in the code), the vector of time points (ti in the code), and the 3d array MUA_b_t_g of size [Nb, Nt, Ng] which held the collective spike rate of each group of neurons at different time points and in different bursts (or trials).

For the average burst activity, firstly the mean of the activities (MUA_g_t in the code) was calculated and transformed in order to make the data as the size of [Ng, Nt]. This mean of the activity was plotted as a function of time. Then the mean from each row was subtracted ($\boldsymbol{X}$, X in the code) so that the mean value of each group across the time was all zero before the start of PCA, and the covariance matrix between different groups of neurons was found ($\boldsymbol{C_X}$, C_X in the code) and plotted using pseudo-colour plot. Next singular value decomposition was used to generate matrices $\boldsymbol{U}$, $\boldsymbol{S}$ and $\boldsymbol{V}$ (U, S and V in the code). The rationale behind this technique is to obtain a matrix $\boldsymbol{V}$ such that the covariance matrix for $\boldsymbol{Y} = \boldsymbol{V^T X}$ is a diagonal matrix. The diagonal values of $\boldsymbol{S}$, which the square values represented the variance captured by each principal component, was plotted. The formulae used to calculate the covariance matrix and the orthogonal transformation were:

Covariance Matrix

$$\boldsymbol{C_X} = \frac{\boldsymbol{XX^T}}{Nt - 1}$$

Singular value decomposition

$$\boldsymbol{X^T} = \boldsymbol{USV^T}$$

To transform the activity onto the orthonormal space formed by the principal components (***Y***, Y in the code), the observed activity was multiplied with the transpose of ***V*** and each principal component activity was plotted as a function of time. Once again, the new covariance matrix was determined ($\boldsymbol{C_Y}$, C_Y in the code) and plotted as pseudo-colour. Then the first three principal components were plotted (*Fig. 6*). After that, all the activity except the first three principal components was zeroed in order to "de-noise" the activity. This de-noised activity was now plotted against time.

For individual bursts, the activity of individual bursts were transformed into the orthonormal space of principal components, and the time-dependent activity of the first three principal components were plotted (*Fig. 7*), and then all the bursts were superimposed in the same plot (*Fig. 8*).

Results and Discussion

Considering a network of spiking neurons with a multi-unit activity that undergoes subsequent bursts of collective activity, the origin of the bursts was of interest. In *Fig. 1*, the activity of each group of neurons from 40 to 5 *ms* before the burst (the time for the burst happened was defined as 0 *ms*, so 40 *ms* before the burst means -40 *ms*) was shown. Each group of neurons had a different mean and a different activity pattern.

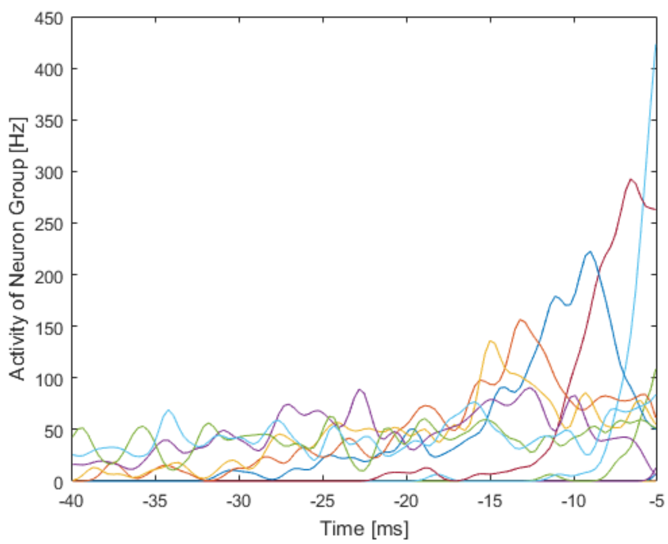

Figure 1 Activity of each group of neurons -40 to -5 ms. Each group of neurons showed a differential pattern of activity along the time.

The next step was to determine the covariance matrix, in which it would be possible to determine the groups of neurons sharing a similar activity pattern. Those with the highest variance were the ones that were more active. For *Fig. 2*, the groups of neurons with the highest co-activity were the ones seen in a warmer color from the rest of the graph, i.e. group numbers 13th, 14th and 15th.

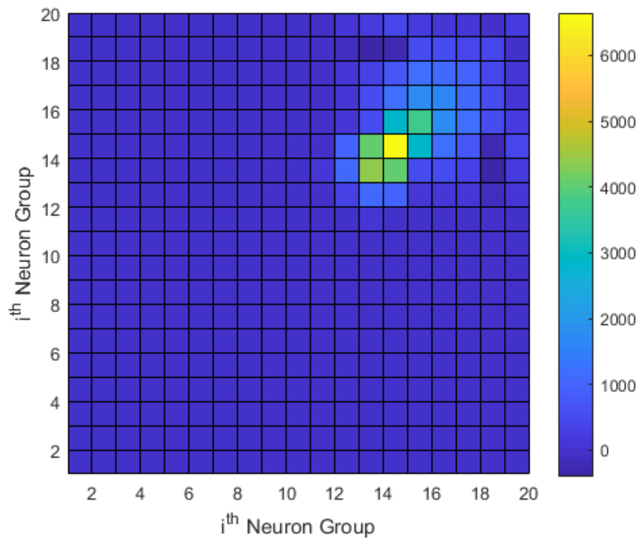

Fig. 2 Covariance matrix of network of spiking neurons. The degree of synchronization of the activity between different groups of neurons was shown. Group number 13th, 14th and 15th showed the highest co-variance among the different pairs as illustrated with warmer colors.

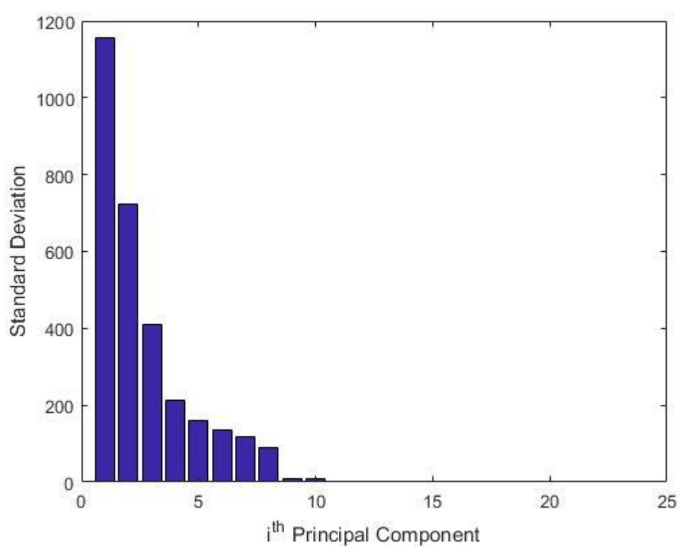

Fig. 3 Standard deviation captured by principal components of a network of spiking neurons. The first three groups of neurons presented a significant amount of 94.8% of the total variance.

Despite having a graphical discovery of the more synchronized groups, it is not enough to understand whether their activities provided the most variance of the system. Therefore PCA was adopted to extract factors that provided this piece of information. Using singular value decomposition, the singular value matrix was plotted in the form of bar chart (*Fig. 3*). It can be seen that the first three PC had already covered significant amount of variance of the system. Further calculation showed that they represented 94.8% of the total variance.

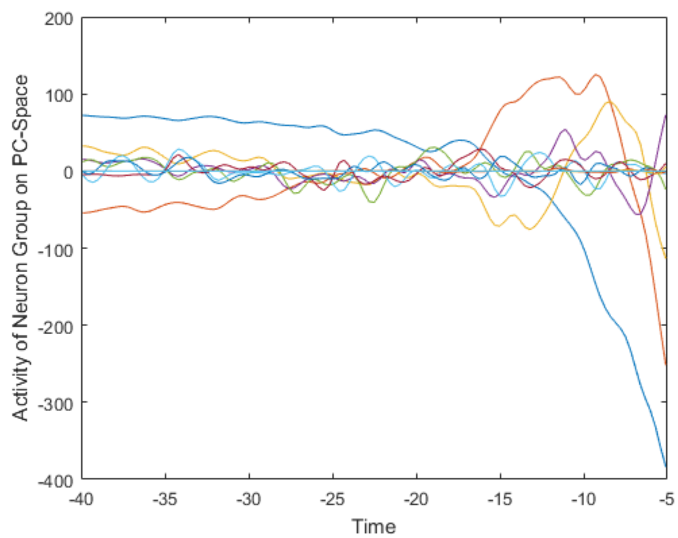

Fig. 4 Transformation of activity observed into the orthonormal space formed by the principal components. The activities of the groups of neurons were represented into a linear summation of different principal components.

After determining the significant principal components, the observed activity of the network was transformed onto the orthonormal space formed by the principal components (*Fig. 4*). To confirm the singular value decomposition is successful, the covariance matrix of the transformed matrix was identified (*Fig. 5*). Since the activity of the network had been transformed, there was a diagonal matrix left in which the first three principal components showed to be the most significant, and therefore they had the most variance about the network. As a result, these three principal components were focused and the activities along the time were plotted in *Fig. 6*.

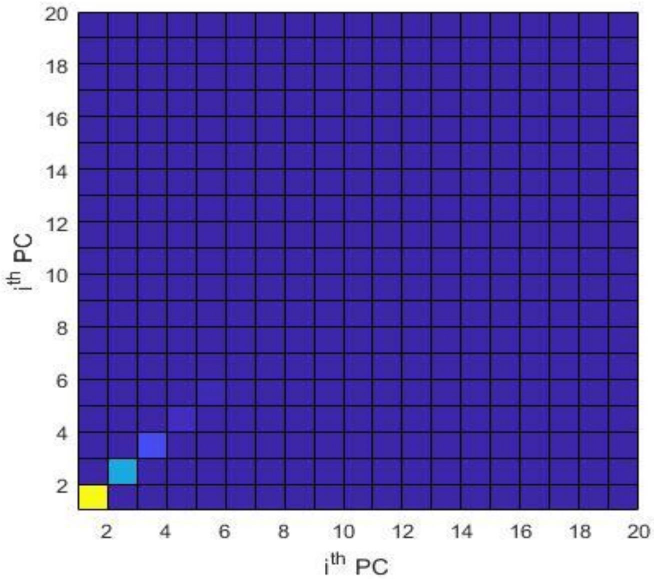

Fig.5 Covariance Matrix of the activity of spiking neurons after being transformed into the orthonormal space formed by the principal components. After the transformation, the first three PCs were identified to be the most significant. Notably, this is a diagonal matrix, which is the expected result of singular value decomposition.

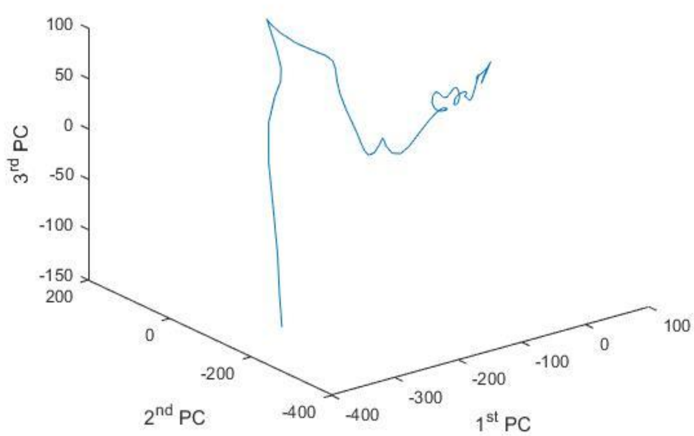

Fig. 6 Time-dependent activity of the average burst projected on the first three principal components.

Theoretical Neuroscience II
Teacher: Prof. Jochen Braun
Assistant Teacher: Ehsan Kakaei
Name: Viviana Villafañe Barraza, Tasnuba Hossain Tazeen, Lee Po Shing
Date: 25th June, 2018

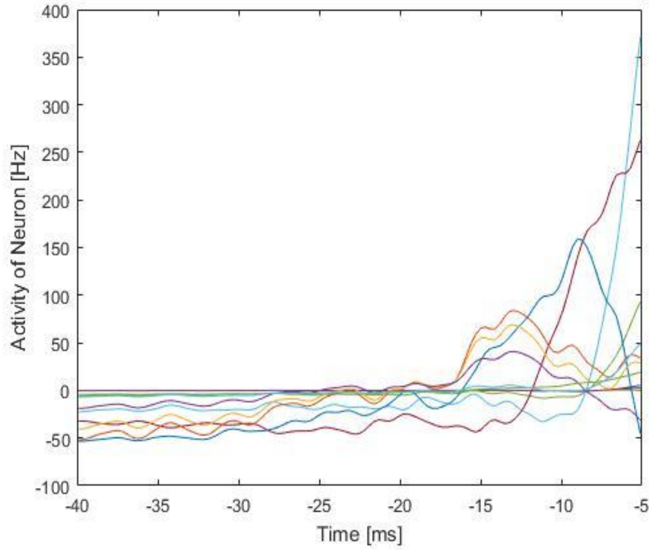

Fig. 7 'De-noised' activity of each group of neurons as a function of time. Using the truncated values of ***Y***, the matrix was back transformed onto the space formed by the activities of the groups of neurons. As one can see, some of the activities was retained, including the orange, yellow and purple peaks (16^{th} – 18^{th} group correspondingly) at time = -15 *ms*, and the blue, red and cyan peaks (13^{th} – 15^{th} group correspondingly) from time = -10 to -5 *ms*.

Once the significant principal components were determined, the truncated matrix of activities on the space formed by the principal components was plotted back into the original space of neuron groups (*Fig. 7*). This was 'de-noised' activity since the neurons that were less necessary (noisy) for the network to generate a burst would not contributed in the significant PCs. Compared with *Fig. 1*, some of the activities retained while the others diminished. Notably, 16^{th} – 18^{th} groups showed synchronized peaks at time = 15 *ms*. Subsequently, the activity seemed to have passed from 15^{th} group to 14^{th} then to 13^{th}. These results also consistent with the covariance matrix observed with these pairs having higher covariance. This may indicate that the synchronization activity of 16^{th} – 18^{th} may be the origin of the collective bursting activity.

Since the above result was from the average burst data, similar patterns as *Fig. 6* should also be observed in the individual burst data. Therefore the time-dependent individual bursts were plotted on the space formed by the first three principal components (*Fig. 8*).

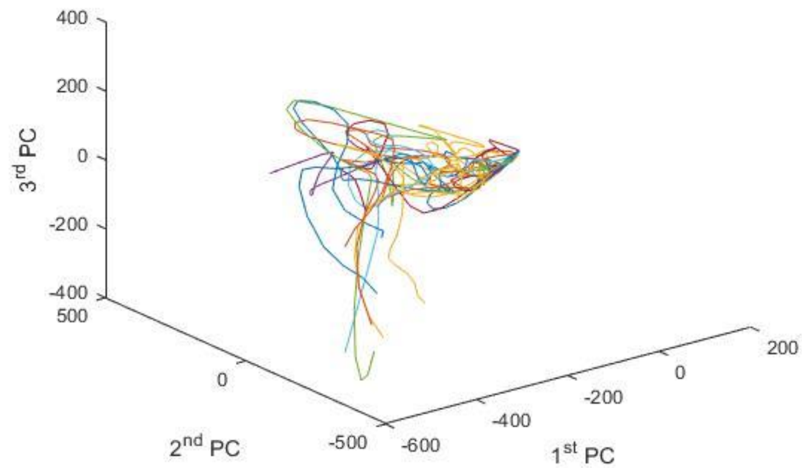

Fig. 8 Time-dependent activity of individual bursts projected on the first three principal components. The time progression in the individual bursts on this space showed similar development as the average burst as in *Fig. 6*.