Theoretical Neuroscience II
Exercise 8
Principal Component Analysis (PCA)

Due date: Thursday, 9 July 2015

June 11, 2015

1 Motivation

Principal component analysis is an extremely useful tool for analyzing high-dimensional data. It identifies correlated sets of variables, so called 'principal components' and thereby reduces the dimensionality of the data.

In this exercise, you will analyze multi-unit activity from a network of spiking neurons. In such networks, quiescent periods with comparatively little activity are interrupted spontaneously by transient periods of intense collective activity ('bursts'). Your task is to investage the origin of these bursts. How does excitation spread through the network? Is there a stereotypical propagation path? Or does each burst start in a unique way?

2 Principal components analysis

Given n observations of m variables, each with zero mean, collected in a matrix $\boldsymbol{X}$

$$\boldsymbol{X} = \begin{pmatrix} x_{11} & x_{12} \dots & x_{1n} & \\ \vdots & \ddots & \ddots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{pmatrix} \in \mathbb{R}^{m \times n}, \qquad \sum_j x_{ij} = 0, \ \forall i$$

and the covariance matrix

$$\boldsymbol{C_X} = \frac{1}{n-1} \boldsymbol{X}\, \boldsymbol{X}^T \in \mathbb{R}^{m \times m}$$

we seek an orthonormal transformation $\boldsymbol{P} \in \mathbb{R}^{m \times m}$ such that the transformed observations have diagonal covariance:

$$\boldsymbol{Y} = \boldsymbol{P}\, \boldsymbol{X} \in \mathbb{R}^{m \times n}, \qquad \boldsymbol{C_Y} = \frac{1}{n-1} \boldsymbol{Y}\, \boldsymbol{Y}^T = \begin{pmatrix} y_{11} & 0 & \dots & 0 \\ 0 & y_{22} & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & y_{mm} \end{pmatrix}$$

From the 'skinny' singular value decomposition $\boldsymbol{X}^T = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T$, we obtain the left singular vectors $\boldsymbol{U} \in \mathbb{R}^{n \times n}$, the singular values $\boldsymbol{S} \in \mathbb{R}^{n \times m}$, and the right singular vectors $\boldsymbol{V} \in \mathbb{R}^{m \times m}$.

The desired orthonormal transformation is

$$\boldsymbol{P} = \boldsymbol{V}^T \in \mathbb{R}^{m\times m}$$

with the **rows** of $\boldsymbol{P}$ being the **principal components**. The variance captured by each component is revealed by diagonal matrix $\boldsymbol{S}$. Specifically, element s_{ii} (the i^{th} diagonal element) represents the variance captured by the i^{th} principal component.

3 Average burst

Multi-unit-activity is provided in a Matlab-file **MUA_b_t_g**. After loading this file, you have the number of bursts **Nb**, the number of time points **Nt**, the number of neurons groups **Ng**, the vector of time points **ti**, and the 3d array **MUA_b_t_g** of size [**Nb, Nt, Ng**], which holds the collective spike rate of each group of neurons for different time points and for different bursts.

Begin by averaging the activity over bursts, to obtain a 2d array **MUA_g_t** of size [**Ng, Nt**]! Note that the variables (neuron groups) do now range over *rows*, whereas the observations (time points) now range over columns, in agreement with the matrix $\boldsymbol{X}$ above.

Plot the activity of each group of neurons as a function of time!

Subtract the mean from each row and save the subtracted means for later!

Compute the covariance matrix and plot with Matlab function **pcolor**!

Perform the singular value decomposition (don't forget to transpose the input!), using the Matlab command **[U, S, V] = svd(X')**! Plot the diagonal values of **S** with Matlab function **bar**! How many principal components capture significant variance?

4 Transformed activity

Transform the observed activity into the orthonormal space of principal components, beginning with the average over bursts!

Plot the activity of each principal component as a function of time!

Compute the covariance matrix of the transformed activity and plot with Matlab function **pcolor**! The result should be a diagonal matrix!

Plot the time-dependent activity of the first three principal components, using Matlab function **plot3**!

Zero the activity of all but the first three principal components and project back into the original space of neuron groups! Plot this 'denoised' activity of each group of neurons as a function of time!

5 Individual bursts

Transform the activity of individual bursts into the orthonormal space of principal components!

Plot the time-dependent activity of the first three principal components! Superimpose all bursts in the same plot, by repeatedly using Matlab function**plot3**!