

Neural data analysis

Notes

ATHUL VIJAYAN ED11B004

Contents

1	Introduction	3
1.1	Background	3
1.2	Experiment	3
1.3	Aim	3
2	Dataset	3
2.1	Restructuring the data	4
3	Analysis	4
3.1	Quantifying the neuron activity	4
3.1.1	Averaging	5
3.1.2	Autoregressive model	5
3.2	Estimating orientation selectivity of neuron	5
3.2.1	Selectivity Index measures - OSI, DSI	5
3.2.2	Circular variance of direction and orientation	5
3.2.3	Measure of reliability - confidence intervals	6
3.3	Classification of simple and complex cells	6
3.3.1	k-means clustering	8

1 Introduction

1.1 Background

1.2 Experiment

Virus expressing GCaMP6f was injected into the V1 of mice. Approximately 3 weeks post infection, mice were imaged under a 2-photon microscope while sinusoidal drifting gratings were presented on a computer screen placed 3 inches from the mouse (1 degree of visual space ~ 21.3 pixels on the screen). The stimulus was varied as:

1. Sinusoidal drifting gratings at 16 different directions (0:22.5:337.5). Spatial frequency was fixed at 0.03 cycles per degree.
2. Each direction was repeated 10 times (i.e. 10 trials per direction). Directions are presented in a randomized order.
3. Each direction was presented for 2s and was always preceded by a 4s gray screen. Therefore the total duration of the stimulus is 6s.
4. Calcium signals (GCaMP6f in awake mice) were acquired from awake mice at 20 frames per seconds. Thus, sampling rate is 20Hz.

1.3 Aim

- To map orientation tuning responses of excitatory pyramidal neurons in V1.
- To determine response reliability at preferred orientation.
- To determine signal and noise correlations between neuron as a function of orientation tuning.

2 Dataset

Notes about Data.mat

Data.mat contains 4 entries:

1. Data.rawF = raw fluorescence values. Matrix size = number of cells x number of frames.
2. Data.dFF = fluorescence normalized to baseline ($dFF = (F-F_0)/F_0$, where F_0 is the baseline fluorescence computed using a sliding window of 400 frames). Same size as above.
3. Data.Spks = inferred spike rate using the Vogelstein deconvolution algorithm. Same size as above.
4. Data.StimSeq = contains sequence of directions presented during that experiment. Vector size = 160 x 1.

Notes about Ori.mat

Ori.mat contains 20 entries. The most pertinent entries are:

1. Ori.OSI = orientation selectivity indices of the neurons
2. Ori.OrFit = double-wrapped Gaussian fits
3. Ori.OrFitQuality = goodness of Gaussian fits. (Higher the percentage value, the better the fit)

4. Ori.Width = tuning width in degrees
5. Ori.PrefOri = preferred orientation
6. Ori.SpKResponse = Contains neural responses for each cells sorted according to the different directions. Size: 1xNumber of cell Cell array. Each cell entry contains a 1x Number of Direction Cell array, which contains a Number of Frames x Trials matrix.

2.1 Restructuring the data

There are 19200 Ca readings for each mouse. The experiment is done using 16 different stimuli, Let us think of it as 16 classes. The angle of the drifting pattern changes between the classes. It is implicit that 0° and 180° will have same orientation, but the direction of drifting will be opposite.

First, the data is split into 16 parts corresponding to each class. There are 10 trials for each class. The data is split into 160 subsequences of $19200/160 = 120$ length. Now this is a time-series of 120 points for each trial, and there are ten sets of data for each class.

120 points in the data represents 6 seconds of brain activity. Out of the 6 seconds, first 4 seconds (80 samples for sampling frequency of 20Hz) correspond to no-stimuli (or Gray stimuli). This represents the background activity of the neuron. The last 40 samples show the response for the intended stimuli.

- cellData*i* contains all data for neuron *i*
- cellData*i* is a 160×121 matrix.
- Each row represents a single experiment. and corresponding gray duration.
- In a row first 120 entries show the time series data of 6 seconds (120 samples) long trial. First 4s (80 samples) gray time and last 2s (40 samples) stimuli time.
- Last column contain the angle of visual stimuli in radians. (class)

3 Analysis

Analysis is broken down into various problem statements. Various methods are considered for each of the problem statements.

3.1 Quantifying the neuron activity

The 120 point time series provide the activity of single cell in response to a particular stimulus in a trial. The activity of cell has to be studied. The first 80 samples are expected to have spikes due to only background activity, as the stimulus was gray during that time. Peaks in Ca concentration denotes the spikes in the neuron. From neuroanatomy, spike rate is a good quantify to measure neuron activity. In Calcium imaging, the amplitudes corresponds to Ca ion concentration. From the neuron models, more the Ca concentration, more is the spike rate. Following methods are used for quantifying spike rate.

Using various methods, a time sequence F_t is calculated. F_t represents the spiking activity of neuron. A relatively large F_t represents a spike and small corresponds to no-spike. Number of spikes in an interval is what we seek. In the 120 sample data, F_t of size 120 is found using various methods. Finally, spike rate is calculated by subtracting the background activity from the response.

$$spikeRate = \frac{\sum_{n=81}^{120} F_n}{40} - \frac{\sum_{n=1}^{80} F_n}{80}$$

Various methods for finding F_t is the first problem statement. Following approaches are made to solve this.

3.1.1 Averaging

The Calcium signal amplitudes averaged and the average is believed to be a proxy for the spike rate. From the smoothed data, just taking the mean of last 40 samples (time during which stimuli was presented) provides the estimated spike rate for that trial.

3.1.2 Autoregressive model

3.2 Estimating orientation selectivity of neuron

Neurons in the striate cortex of primary visual cortex in the brain are of three types. Cells which respond to orientation of a contrasting visual stimuli and which are not sensitive to orientation. Among the orientation selective cells, some cells are also sensitive to motion of the stimuli in a particular direction in addition to orientation. These cells are called complex cells. Cells which are only sensitive to orientation, but not to motion are called simple cells. These cells are pyramidal in structure and have a receptive field in shape of a rectangular slit. This section deals with methods to quantify the selective behavior of these cells from experimental data.

3.2.1 Selectivity Index measures - OSI, DSI

OI and DI are used as the normalized measures of “peak to trough” orientation selectivity and direction selectivity. These quantities are defined as:

$$OSI = \frac{(R_{pref_ori} - R_{orth})}{(R_{pref_ori} + R_{orth})} \quad (1)$$

$$DSI = \frac{(R_{pref} - R_{null})}{(R_{pref} + R_{null})} \quad (2)$$

The response at the preferred orientation R_{pref_ori} and the response at the preferred direction R_{pref} can be determined in different ways. In some measures, these are taken to be the best response to one of the stimulus orientations or directions that was explicitly measured; that is, if we measure responses at stimulus directions $\theta_1, \dots, \theta_n$, then we choose the response at the best θ_i . In other measures, we perform a fit to the tuning curve, and choose the maximum value of the fit as R_{pref_ori} or R_{pref} .

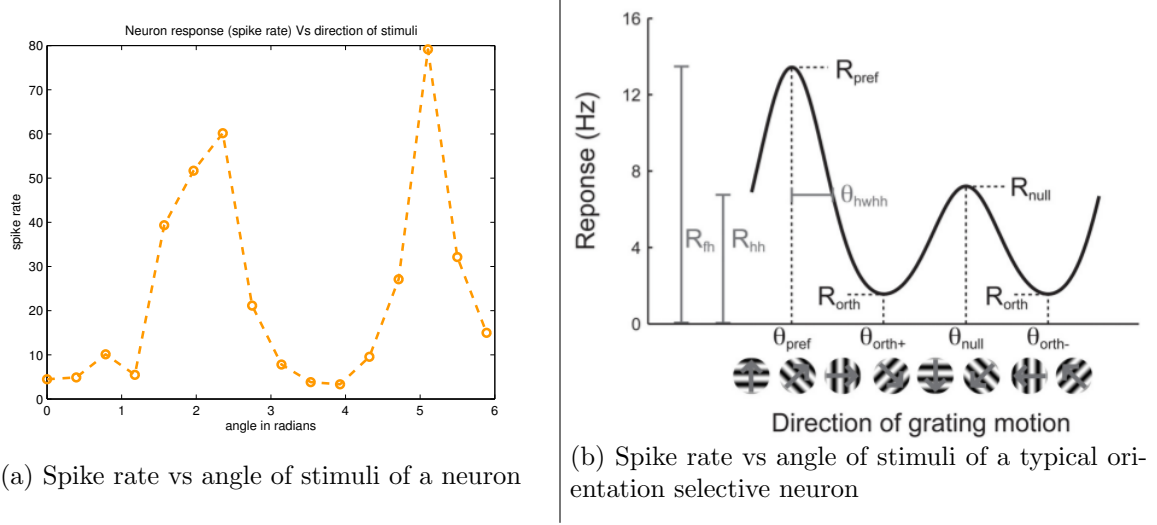
This approach is more sensitive to noise. One of the plot of spike rate vs angle of stimuli of a neuron is shown in Figure 1.

3.2.2 Circular variance of direction and orientation

As discussed in the paper, Circular variance is a more robust quantifier of orientation and direction selectivity. A scalar value representing the neuron response is extracted from the 6s long time series for each of the orientations and directions (There are 8 orientations and 16 directions). As we have 10 trials for each class, we will have 160 such values for each neuron. The values are first expressed as vectors in the orientation space. The vector sum of the values provide the ‘preferred orientation’; Magnitude should indicate how much orientation selective the neuron is.

Let $R(\theta)$ be the response of neuron to angle θ and we denote the ‘preferred direction’ as θ_{pref} .

Figure 1: Comparing response of a neuron Vs angle of stimuli with the expected response



Normalized vector sum in orientation and direction space is found as.

$$L_{ori} = \frac{\sum_k R(\theta_k) \exp(2\theta_k)}{\sum_k R(\theta_k)} \quad (3)$$

$$L_{dir} = \frac{\sum_k R(\theta_k) \exp(\theta_k)}{\sum_k R(\theta_k)} \quad (4)$$

- Neurons with orientation selectivity are expected to have high L_{ori} .
- Simple cells are orientation selective, but less direction selective. They are expected to have large L_{ori} and small L_{dir} .
- Complex cells are orientation selective as well as direction selective. They are expected to have both large L_{ori} and large L_{dir} .

Circular orientation variance and direction variance is given as

$$CirVar = 1 - |L_{ori}| \quad (5)$$

$$DirCirVar = 1 - |L_{dir}| \quad (6)$$

Following shows plots of orientation and direction selectivity of various kinds of neurons.

1. Orientation selective simple cell

Refer Figure 2

2. Direction selective complex cell

Refer Figure 3

3. Orientation and direction Un-selective cells

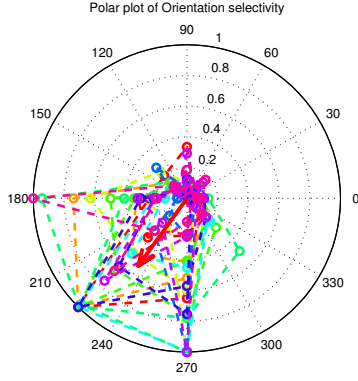
Refer Figure 4

3.2.3 Measure of reliability - confidence intervals

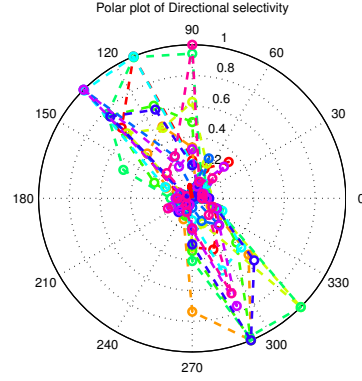
3.3 Classification of simple and complex cells

The properties that distinguish the tree types of cells are discussed before. This problem statement deals with using machine learning tools for making a classifier for these three classes

Figure 2: Orientation selective simple cell

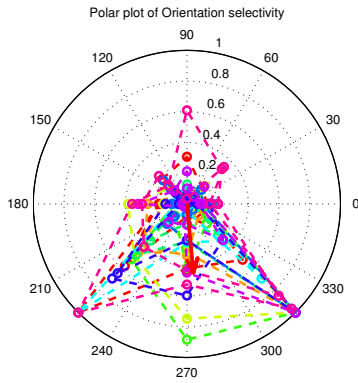


(a) Orientation selectivity plot of an orientation selective simple cell. Each curve shows a trial. Note that the cell is sensitive to an optimal orientation. The red arrow represents the direction and magnitude of orientation selectivity

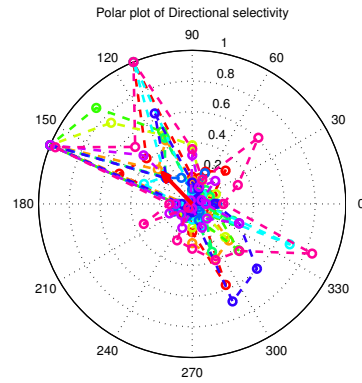


(b) Directional selectivity plot of an orientation selective simple cell. Each curve shows a trial. Note that the cell is not selective to direction, but is sensitive to orientation. The red arrow represents the direction and magnitude of direction selectivity

Figure 3: Direction selective complex cell

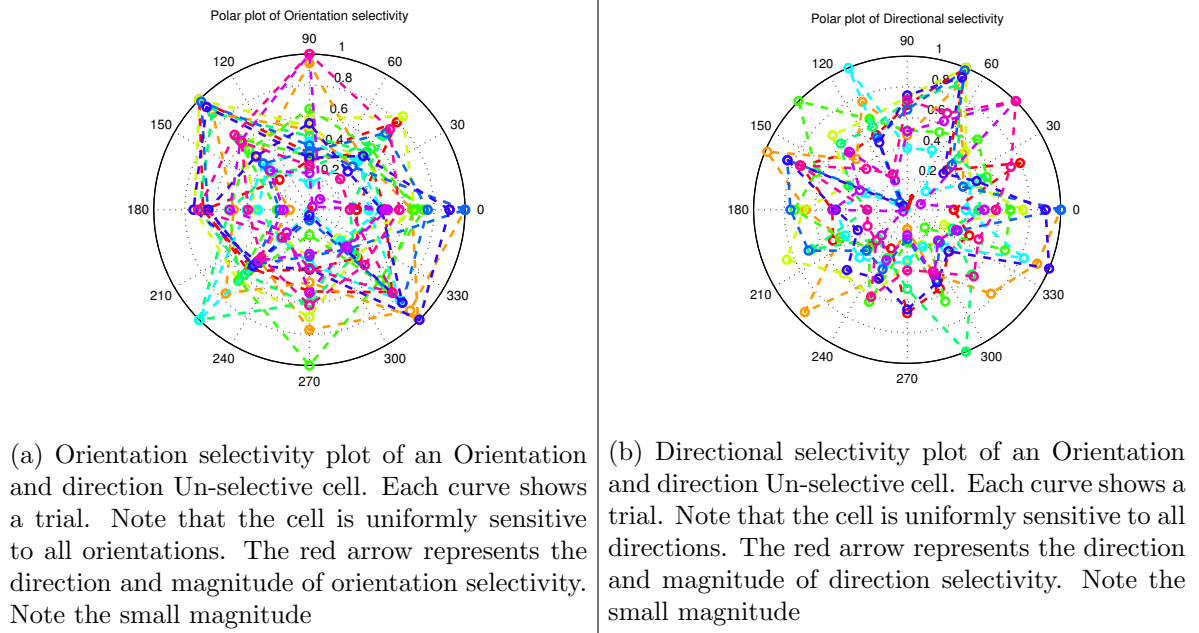


(a) Orientation selectivity plot of an direction selective complex cell. Each curve shows a trial. Note that the cell is sensitive to an optimal orientation. The red arrow represents the direction and magnitude of orientation selectivity



(b) Directional selectivity plot of an direction selective complex cell. Each curve shows a trial. Note that the cell is selective to direction as well as orientation. The red arrow represents the direction and magnitude of direction selectivity

Figure 4: Orientation and direction Un-selective cells



of cells.

The main challenge here is that we don not have the ground truth data for the data, which force us into category of unsupervised learning methods. This also makes it hard to evaluate the performance of the classification algorithm. We make an effort to classify cells based on some features extracted from the time series data. The feature selection process and the intuition behind it are discussed below.

- OSI : This is a measure of orientation selectivity, as an individual feature this may not be a strong classifier as the measure is not very robust, but along with other features, this could contribute towards classification.
- DSI : This is a measure of directional selectivity, as an individual feature this may not be a strong classifier as the measure is not very robust, but along with other features, this could contribute towards classification.
- L_{ori} : This is more robust that OSI, and is expected to be a better feature.
- L_{dir} : This is more robust that DSI, and is expected to be a better feature.

Following unsupervised learning techniques are use for the classification.

3.3.1 k-means clustering

We hope that three classes of cells will form three clusters in the feature space. We run the algorithm with 3 clusters and features selected above.

As we said, analyzing efficiency of classification is hard without ground truth. Here, what we use to analyze is this:

- Find the global mean
- Find the sum of squared distance of each point in feature space to global mean. denoted by *total_ss*.

- Find the sum of squared distance of each centroids to the global mean multiplied by number points in the cluster. denoted by *between_ss*.
- If there is no well clustering, the centroids will be close to the global mean, and *between_ss* will be small compared to *total_ss*.
- If $clusterEfficiency = between_ss/total_ss$ is high, the clustering is good and forms distinct clusters.

Table 1 shows the clustering efficiency in each mice.

Subject	Cluster efficiency (%)
Mouse A	78.8155
Mouse B	72.3963
Mouse C	72.293
Mouse D	78.8882
Mouse E	76.4573

Table 1: Cluster efficiency of various subjects