Cathy Zhuang

BIOL 450 Dr. Murugan

8 April 2022

Assignment 9

Part 1

Question 1

The code does not do well in pulling out individual neurons, but it is also not entirely inaccurate. For example, one can see two calcium spikes in graph 1 on the right at around 60 and 120 units, and on graph 3 on the right at around 220 units, that correspond with a few points in the actual images.

The code performs better on simulated data instead of real data because real data has much more noise and other factors that may not occur during simulated data. For example, we may have to correct for x-y-z movement in real data (such as piecewise rigid correction), and we may not necessarily have to do that for simulated data. There may also be noise introduced for one-photon imaging since neural components exhibit strong spatial overlap. Additionally, not all noise may follow a normal distribution (for example, MRI noise follows a Rayleigh distribution) making it more difficult to distinguish between noise. Most times, one must run de-noising before running nnmf—otherwise, the factorization may have to reduce the resulting W and H matrices into a rank that is smaller than what is expected, resulting in an output where neurons are not isolated as well.

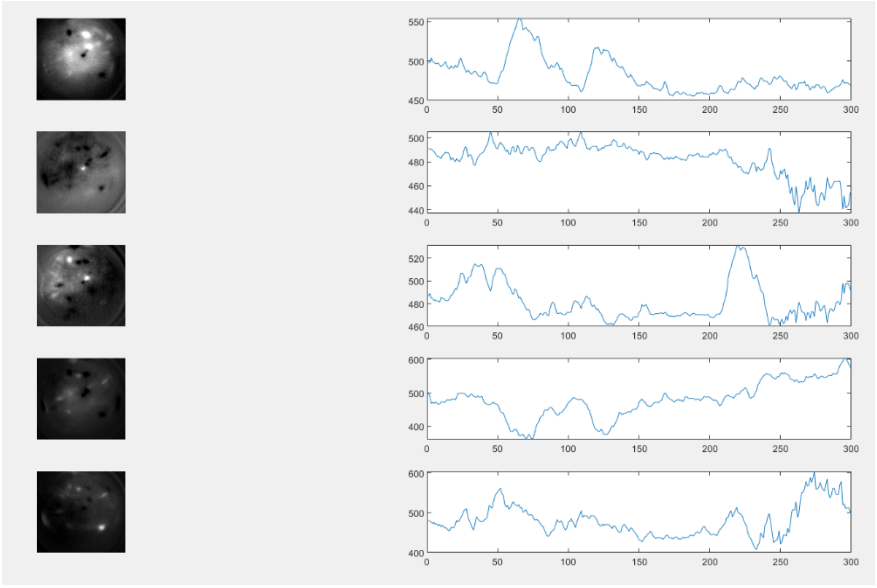


Figure 1. Program output after running nnmf.

Warning: Algorithm converged to a solution of rank 9 rather than 20 as specified.

> In nnmf (line 227)

Elapsed time is 617.177130 seconds

Question 2

Since the factorization uses an iterative algorithm where the room mean square residuals might have local minima, we will have different W and H results for repeated factorizations. Thus, if we were to run 1000 repetitions instead of 100, we will have tried many more combinations of W and H and are thus less likely to converge to a solution that is not optimal, that is a lower rank than *k* (a less optimal solution) for a W matrix of *n x k* and an H matrix of *k x m*. More repetitions may thus reflect the original matrices W and H better than less repetitions. Thus, we will be able to extract the individual neurons better with more replicates.

Part 2: Place Cells

Question 1

The animal spends most of its time at both ends of the track. This is very typical mouse behavior; they spend times at the ends to avoid predators, as spending time out in open space makes them vulnerable to predators.

Question 2

The overall shape of these sample place fields follows a normal/gaussian distribution; some are skewed. A typical place field seems to be around 1-2 meters wide. The maximum firing rate is not the same for all 12 place fields, meaning that some neurons fire more than others in their place field.

Question 3

Since we are using a winner-take-all strategy, the errors that occur are not terrible in that they are not completely inaccurate. However, there is bias when the animal is right between two neuron's receptive fields, and the small number of cells we have sampled may mean that we get errors. In other words, we need to measure more cells for a less erroneous result; as it stands, we have under-sampled, and we may not have a sample that is representative of the trajectory that the mouse actually takes. If we had more cells and could decode the position of the animal perfectly well, our model would fit better.

Question 4

It seems like there is more noise with smaller time windows. With larger time windows, the error is minimized and there is less noise. This advantage is because larger time windows average more points, so we are less likely to converge to something that is just noise. The disadvantage with smaller time windows is that one point that is just noise can skew the data, making the model not necessarily representative of the actual trajectory. However, too large of a time window may also pose some problems, as averaging across too much may also generalize the actual trajectory and result in something that is not representative.

Each trial will have an optimal tau, and it seems like out of the given requested taus of 0.025 and 1.0 seconds, each model fits the trajectory the best in time windows of 1.00 seconds rather than 0.025 seconds. In the plots, the models with 1.0 second time bins tend to follow the trajectory, whereas the models with 0.025 second time bins also have points that are not on the trajectory. It seems like 0.025 seconds is too small of a time window and introduces too much noise, whereas 1.0 seconds is more optimal.

MLE decodes better than winner-take-all in this case because when we have an under-sampled dataset, we can account for the under-sampling by generating a likelihood on what *all* the neurons are doing instead of just one neuron. In other words, decoding depends not on just one neuron now, but also what the population is doing as a whole. This is because we are looking at n number of spikes observed in neuron 1 at x1 position, and so on and so forth. Thus, MLE decodes better than winner-take-all by considering all the neurons.

Question 5

For Trial 1, the period of time that MLE works best is a tau of 1.0 seconds. The superimposed model almost exactly fits the trajectory of the mouse. For Trial 2, the period of time that MLE works best is also a tau of 1.0 seconds; again, the superimposed model almost exactly fits the trajectory of the mouse. For Trial 3, the superimposed model does not fit as well to the trajectory at 1.0 seconds compared to Trials 1 and 2; however, the model at 1.0 second time bins fits much better than the model at 0.025 seconds.

Additionally, it seems like a time bin of 1.5 seconds may fit better. Overall, the larger time bins of 1 second recover the spatial trajectory much better than the smaller time bins of 0.025 seconds. This is because time bins that are too small will introduce noise, while time bins that are larger will average across more points, resulting in less emphasis on noise.

For Trial 1, the recovery works best during 6242-6248 s and 6252-6256 s. It does not do well in between those times. For Trial 2, the recovery works best from 6528 – 6534 s. Finally, for Trial 3, the recovery works best from 6640-6650 s, and from 6655 – 6660 s.

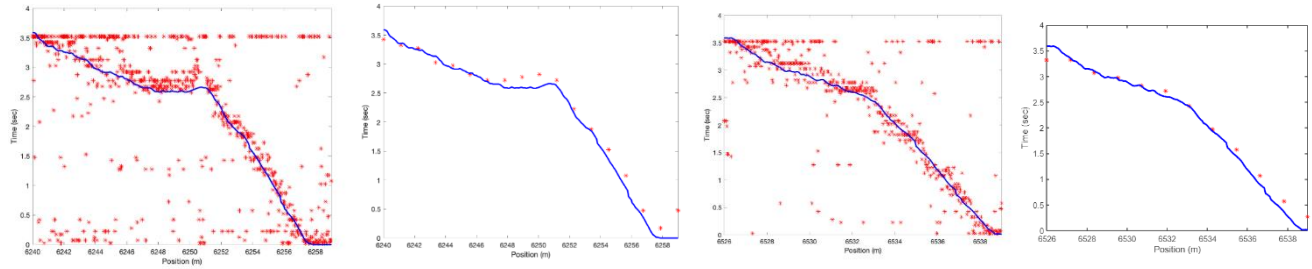
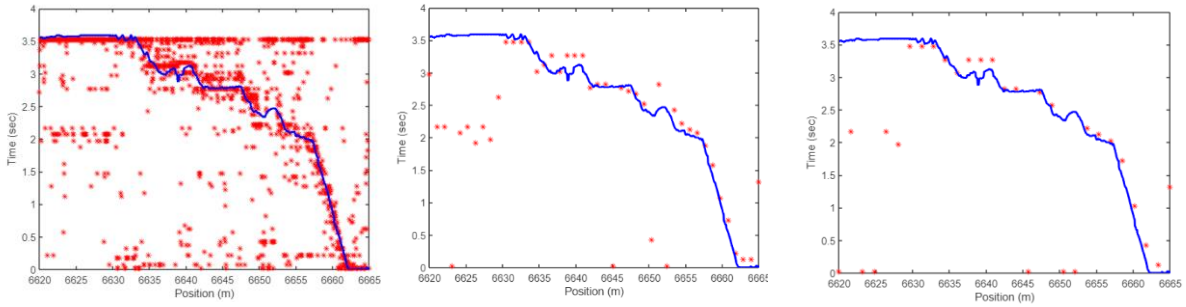


Figure 2 (top). From left to right: Trial 1 with tau = 0.025 seconds. Trial 1 with tau = 1 second. Trial 2 with tau = 0.025 seconds. Trial 2 with tau = 1 second. I believe the axes labels are flipped.

Figure 3 (bottom). From left to right: Trial 3 with tau = 0.025 seconds. Trial 3 with tau = 1 second. Trial 3 with tau = 1.5 seconds. I believe the axes labels are flipped.



Question 6

One can observe the animal moving at high velocities when there is a quick drop in position over a short period of time (i.e. a large negative slope). One can observe the animal moving at low velocities when there is a slow drop in position over a long period of time (i.e. a slope closer to horizontal). The trajectory reflects an animal changing movement direction when there is a local maximum or minimum—when the trajectory goes from increasing to decreasing, or from decreasing to increasing.

It seems like the trajectory recovery does not model the actual trajectory well when the animal is not moving or moving at slow velocities. We can see that when the slope of the trajectory is horizontal (which indicates stationary or slow velocity) in the middle section of Trial 1 and the beginning of Trial 3,

the red dots superimposed do not line up with the actual trajectory, meaning that the estimate did not correspond to the actual trajectory.

At high velocities, trajectory recovery generally fits the actual trajectory of the mouse; in all trials, when there is a steep slope that indicates a high velocity, the red asterisks generally follow the blue line, meaning that the model fits the trajectory.

The model does not do well when the mouse changes direction. Directional changes can be seen as minima and maxima, and they are most evident in Trial 1 and Trial 3. In the event of a maxima, the superimposed asterisks generally do not fall on the blue line, indicating that trajectory recovery was not as accurate.