ERC Report - June 6, 2010

Amir Hesam Salavati
E-mail hesam.salavati@epfl.ch

Supervisor: Prof. Amin Shokrollahi
E-mail amin.shokrollahi@epfl.ch
Algorithmics Laboratory (ALGO)
Ecole Polytechnique Federale de Lausanne (EPFL)

June 6, 2010

1 Introduction

In the past two weeks, I was mainly busy reading a couple of other papers on neuroscience as well as two books. In what follows, I briefly review these materials.

2 Neuron Models

In [4], Gerstner et al. have proposed mathematical models for neurons to investigate their spiking behavior. The book is very easy to read and simple to understand while the models are quite accurate. Specially the chapters on noise in neuronal system is very interesting from channel modeling point of view.

The other book [5] describes the theory behind neural computation and neural networks. It is particularly interesting for designing artificial neural networks which comes handy if one would like to design neural-based decoders. Nevertheless, it is not as easy to read as the previous one.

3 Robustness of Neural Code

A central neuronal sciences is if neural code is robust and if so, how robust it is. In neuroscience literature, robustness means immunity to noise in neural system. In this sense, robustness is also related to error correction mechanism. Nevertheless, it should be noted that noise immunity could be achieved by noise reduction methods. Noise reduction in neuronal systems via population coding is discussed in [1]. In this paper, authors have used *frame theory* to describe population coding in neural systems. They showed that when a group of neurons encode a stimulus, the result is more robust to noise. Frame theory is used here to assess robustness of population codes, i.e. how much noise reduction could be obtained if a population of neurons encode stimulus cooperatively.

Here is the model used in [1]: we have a message, s, which we would like to transmit over a number of channels (neurons). We use a set of vectors $\{\phi_i\}$, that **are not orthogonal**. We project s over the vectors to get coefficients, c_i, i.e. $c_i =< s, \phi_i >$, where $< . >$ means averaging. These coefficients are then transmitted over the channels. To reconstruct the message in the receivers, frame theory gives us a set of vectors $\{\widetilde{\phi_i}\}$ such that $s = \sum_i c_i \widetilde{\phi_i}$.

However, due to channel noise, the received coefficients are noisy. Hence, we get an estimate of $\widehat{s}$ of s according to the following equation:

$$\widehat{s} = \sum (c_i + n)\widetilde{\phi_i}, \tag{1}$$

where n is Gaussian noise with zero mean and variance σ^2.

We are now interested in the effects of population coding on noise, i.e. how much noise reduction could be achieved if neurons encode s collectively. It is shown that in case of cooperation among neurons, mean square error between the message s and its estimation $\widehat{s}$ is reduced by a factor depending on the characteristics of vectors $\{\phi_i\}$ [1].

Frame theory extends the notion of basis to a set of redundant vector (instead of just considering orthogonal ones). Therefore, frame theory is a general approach and could also be used to describe error correction codes [6]. The main question is though if this model realistic for neuronal systems. In other words, it is not clear if neurons behave according to the above model by projecting the stimulus over a set of vectors and transmit the projection coefficients. Nevertheless, the model mentioned in [1] is quite useful in assessing robustness of population codes in neuronal systems.

4 Information Theory and Neuronal Systems

The main task of information processing theory is to quantify how well signals represent information and how well systems process information by doing operations on those signals. In [2], information theoretic measures, Kullback-Leibler distance in particular, are used to evaluate the ability of signals to represent information. The ability of systems to process information with high fidelity is assessed here by measuring the information distance between input and output signals of such systems. It is suggested that instead of using entropy and mutual information, it is better to use Kullback-Leibler distance since computation of mutual information requires joint probability distribution between input and output, which is difficult to obtain.

Classical information theory developed by Shannon tackles technical part of communication systems. However, there are two more levels in such systems: semantic, and influential [2]. The proposed method in [2] deals with semantic part of the systems. It also addresses influential part by quantifying the effectiveness of the action performed as a result of interpreting the "meaning" of information (semantics). In the suggested framework, information processing results in actions. These actions, on the other hand, lead to signals that the information probe, the information-theoretic Kullback-Leibler distance, can evaluate. Therefore, in [2] Sinanov and Johnson extends classical information theory of Shannon to neuronal systems.

The model used in the paper is shown in figure 1. In this model, information source generates information shown by α. This and other extraneous information, indicated by X_0 and X_1, are encoded to give the sequence X (X_0 and X_1 are pieces of information that are not in information sink's interest). By nature, X is stochastic. This sequence is passed through a system with probability distribution $p(Y|X)$ to result in sequence Y. In other words, X is passed through a channel determined by $p(Y|X)$, where Y is the channel output. Finally, information sink "interprets" (decodes) Y by exhibiting an action Z.

To evaluate systems performance when a change from α_0 to α_1 has occurred in the relevant information[1], we measure the distance between the

[1] The relevance is very important in investigating neural coding. Because a neural system could be very bad at preserving all of input information in its output but its goal may be to just extract a few features out of input. Therefore, by considering relevant information, we observe that this system is preserving relevant input information with high fidelity.

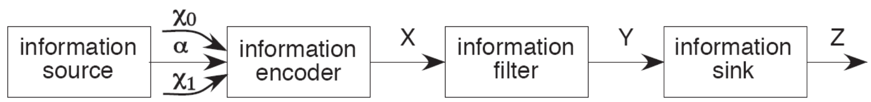

Figure 1: The model used in [2]

response of the system to this change, i.e. $d_x(\alpha_0, \alpha_1)$. Instead of depending on X directly, d_x depends on the distance between probability distributions $p(x, \alpha_0)$ and $p(x, \alpha_1)$. Different such measures are considered in the paper among which the Kullback-Leibler distance seems most appropriate. In particular, Kullback-Leibler distance is used to assess the fidelity of systems in processing information by defining the information transfer ration as below:

$$\gamma_{X,Y}(\alpha_0, \alpha_1) = \frac{d_Y \alpha_0, \alpha_1}{d_Y \alpha_0, \alpha_1} \tag{2}$$

where Y is the output and X is the input. d_X (d_Y) is the Kullback-Leibler distance in the input (output). The closer to one the γ is, the better the system would be.

For each α_0, γ could be equal to 1 only for some values of α_1. These values of α_0 and α_1 represent those features of information that are perfectly represented by the system.[2] If α_0 is the reference point, a plot of γ against α_0 and α_1 reveals those aspects that system represents well. These are the regions in the plot that γ is large. Ideally, we like γ be equal for all choices of α_0 and α_1.

Another issue which was addressed in this paper is the ability of signals in representing information parameter, α. The results show that in order for the signal to be able to encode information with high fidelity, small differences between α_0 and α_1 must results in large distances between them, i.e. $d_X(\alpha_0, \alpha_1)$ be large enough. Where d_X is the Kullback-Leibler distance.

What is really nice about the information processing theory proposed in [2] is that the authors have provided relationships between information transfer ratios for various combinations of systems as shown in figure 2.

[2]Hence, it could hint for features that are important for the system. It could depend on α_0, α_1 or both.

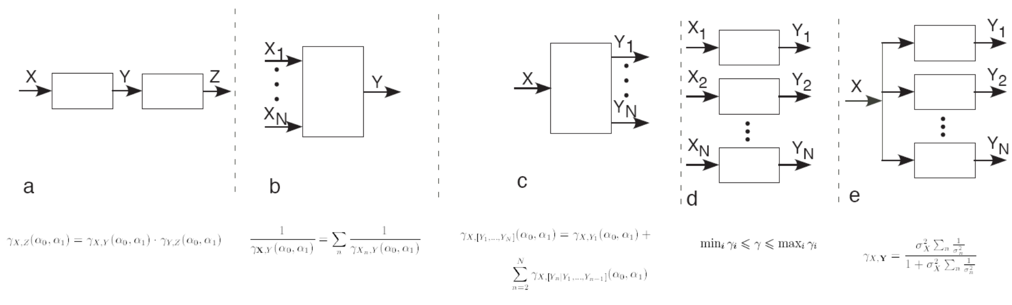

Figure 2: Different combinations of information processing systems [2]

Another interesting argument in [2] is that authors believe that methods based on rate distortion theory (the ones that assess the performance of the system based on a distortion measure between input and estimated output, such as MSE) are inappropriate for neuroscience as in neuroscience, when you gather statistical data to analyze part of a neural processing system, usually the distortion function is both unknown and irrelevant. It is unknown because when the decoded information is not evident, distortion can not be measured. Irrelevant because we do not know how to quantify the distortion between intended and derived meaning. Moreover, meanings translate into actions and actions usually lie in a different space than the input sequence. In these cases, the distortion measure is clearly unknown, unless one justifies the validity of its distortion function for a particular case.

Another work on information theory in neural science that I read was [3] in which authors use information maximization principle to find the optimal behavior of synaptic weights. More specifically, authors maximize mutual information between N input synapses of a neuron and its outputs subject to the constraint that post-synaptic average firing rate stay as close as possible to its typical value. This is basically the same as the power constraint in communication systems. Following this approach, authors obtain the optimal updating rule for synaptic weights which is very similar to the experimentally verified BienenstockCooperMunro rule [7]. In principle, BCM rule gives an update rule for synaptic weights based on the timing of pre and post synaptic spikes. The extension of the BCM rule to spiking neurons with refractoriness in [3] shows that synaptic changes should naturally depend on spike timing, spike frequency, and postsynaptic potential (PSP), which is in agreement with experimental results.

5 Conclusion and Future Works

Based on what I have read so far and my discussions with Prof. Gerstner and Prof. Moret, I am going to write my final report on applications of coding theory in biological systems. From what I have learned about neuronal systems, I think the main focus would be on such systems because they seem more interesting than molecular biological systems although I will continue working on such systems as well.

References

[1] C. J. Rozell, D. H. Johnson, "Analyzing the Robustness of Redundant Population Codes in Sensory and Feature Extraction Systems", Computational Neuroscience: Trends in Research, 2006, Vol. 69, No. 10, pp. 1215-1218

[2] S. Sinanov, D. H. Johnson, "Toward a Theory of Information Processing", Signal Processing, Vol. 87, No. 6, 2007, pp. 1326-1344.

[3] T. Toyoizumi, J. P. Pfister, K. Aihara, W. Gerstner, "Generalized BienenstockCooperMunro rule for Spiking Neurons That Maximizes Information Transmission", Proc. National Academy of Sciences, 2005 , pp. 5239-5244.

[4] W. Gerstener, W. Kistler, "Spiking Neuron Models", Cambridge University Press, 2002

[5] J. Hertz, A. Krogh, R. G. Palmer, "Introduction to the Theory of Neural Computation", Westview Press, 1991.

[6] G. Kutyniok, A. Pezeshki, R. Calderbank, T. Liu, "Robust Dimension Reduction, Fusion Frames, and Grassmannian Packings", Applied and Computational Harmonic Analysis, Vol. 26, No. 1, 2009, pp. 64-76.

[7] "Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex", Journal of Neuroscience, Vol 2, 1982, pp. 32-48.