Optimal Population Coding of Dynamic Stimuli

Alex Kunze Susemihl

February 4, 2013

Contents

1 *Introduction* 5

2 *Filtering and Prediction with Point Process Observations* 11

3 *Mean-Squared-Error for Point Process Filtering* 13

4 *Optimal Control with Point Process Observations* 15

5 *Optimal Population Coding Revisited* 17

6 *Discussion* 19

7 *Bibliography* 21

1
Introduction

NEUROSCIENCE AS A WHOLE is concerned with the function of the nervous system. More precisely, it asks a very simple question: *What is the brain doing?*[1] The simplicity with which humans and animals perform in their environment makes it almost unnatural to ask how their brains enable these behaviors. It is often hard to explain to laymen the complexity involved in preparing even the simplest actions, such as saccades or walking, such is the ease with which these are normally performed. One can not realistically expect to answer that question in any general fashion, I will however, try to touch upon a number of points which shed light on a number of aspects of the nervous system and provide us with a *guiding principle* to understand what the brain is doing and why and possibly how.

[1] Or alternatively: *What is the nervous system doing?*

Neuroscience was born as a branch of biology, and although it is now often thought of as an interdisciplinary science in itself, its objects of study are still to a large extent biological systems. Theodosius Dobzhansky published an influential essay in 1973, entitled *Nothing in biology makes sense except in the light of evolution*[2], which defends exactly that point. Though it has been reviewed and revisited constantly since its proposal, the theory of evolution through natural selection remains the central pillar of biological sciences. As such, neuroscience must also view its objects of study through the lenses of evolution. More specifically, we can then ask ourselves *What evolutionary advantage would this brain bring to an individual?* instead of *Why is the brain this way?* That being said, there are caveats in the case of neuroscience. For one, the brain is capable of plasticity and adaptation unthinkable for other organs, and so we can not expect to understand the functionality of the brain in the same way which the shape of bird beaks can be understood as a function of their preferred fruits and seeds. Furthermore, the brain controls all of the motor and perceptual apparatus, having a multitude of uses and purposes, unlike simpler organs.

[2] Dobzhansky, T. (1973). Nothing in Biology Makes Sense Except in the Light of Evolution. *The American Biology Teacher*, 35(3):125–129

One particular aspect of the brain which has received increasing attention recently is its ability to deal with uncertainty. In a very productive line of research, a number of experiments have demonstrated that human and animal integrate uncertain information in a near-optimal way. The so-called *Bayesian Brain*[3], would explicitly represent the distribution over world states and perform inference in a manner consistent with Bayesian inference, obtaining optimal integration of sensory cues from different modalities, for example. It is still a matter of debate how these Bayesian computations would be implemented in the brain. One possibility is that the activity of neurons is sampling from a representation of the distribution of world states[4], which is frequently called the *sampling hypothesis*. Another is that the activity of the neurons itself represents the likelihood over world states[5], and the population as a whole codes for the distribution, hence the term *population coding*.

[3] Knill, D. C. and Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in neurosciences*, 27(12):712–9

[4] Berkes, P., Orbán, G., Lengyel, M., and Fiser, J. (2011). Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science (New York, N.Y.)*, 331(6013):83–7

[5] Ma, W. J., Beck, J. M., Latham, P. E., and Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature neuroscience*, 9(11):1432–8

Structure

THE MAIN GOAL OF THIS THESIS is to develop a conceptual framework for studying optimal population coding in a dynamic framework. Furthermore, I would like to establish a link between optimal dynamic encoders and the efficient coding hypothesis, first proposed by Horace Barlow[6]. I believe that the inclusion of time into the coding framework raises a number of questions, which have not been addressed in the scientific literature properly. In the remainder of this chapter, I will discuss the efficient coding hypothesis and its more recent developments, and I will touch upon its relationship to Shannon's information theory[7]. I will finish by discussing the issue of dynamic population coding, highlighting the issues which I believe are of importance in considering the temporal aspect of coding. I will make the case for a study of optimal filtering of partially observed stimuli as a model of stimulus inference based on spike trains. Following, in chapter 2 I will introduce the general theory of filtering of stochastic stimuli. After that, in chapter 3 I will discuss results regarding the Mean-Squared-Error (MSE) of optimal filters of point process data, presenting a number of new analytical results. In chapter 4, I will generalize the filtering framework to control problems, showing results for optimal control theory of point process-observed processes. In chapter 5 I will then provide the connection to neuroscience, by considering the optimal encoding strategy for a population of neurons coding for a stochastic stimulus. I will then finalize by discussing the impact of the work presented and suggesting future research directions.

[6] Barlow, H. (1961). Possible principles underlying the transformation of sensory messages. *Sensory communication*, pages 217–234

[7] Shannon, C. E. (1948). The mathematical theory of communication. 1963. *Bell System Technical Journal*, 27:379–423 and 623–656

Contribution

THE MAIN CONTRIBUTION OF THIS THESIS is in providing a conceptual toolbox to study optimal coding problems in a dynamic environment. I propose that the study of the average performance of an optimal Bayesian filter reconstructing the relevant stimulus provides a good measure of the quality of a dynamic code. Using this framework, I derive analytical results for the fast population code for dense populations of Gaussian neurons proposed by Quentin Huys[8]. These are to my best knowledge the first results of this kind obtained for temporal coding of dynamic stimuli.

[8] Huys, Q. J. M., Zemel, R. S., Natarajan, R., and Dayan, P. (2007). Fast population coding. *Neural Computation*, 19(2):404–441

1.1 *Efficient Coding Hypothesis*

THE INFORMATION associated with a random event is defined as the logarithm of its inverse probability. This is at first nonintuitive, but with this definition, Shannon [9] has stated a number of interesting results regarding the coding and transmission of messages. We can further define the entropy of a distribution over a set of events as the average information conveyed by these events. So if we have a random variable X taking values $x \in \mathcal{A}$ and a probability distribution $P : \mathcal{A} \to [0, 1]$, we will have

[9] Shannon, C. E. (1948). The mathematical theory of communication. 1963. *Bell System Technical Journal*, 27:379–423 and 623–656

$$H(X) = \sum_x P_X(x) \log \left(\frac{1}{P_X(x)} \right).$$

Furthermore, let us define the conditional entropy of two random variables as

$$H(Y|X) = \sum_x P_X(x) \sum_y P_Y(y|x) \log \left(\frac{1}{P_Y(y|x)} \right),$$

i.e. the conditional entropy is the average entropy of Y given X, averaged over X. Let us also define the mutual information between Y and X as

$$I(X;Y) = H(Y) - H(Y|X).$$

Shannon regarded a noisy communication channel as a set of two random variables, one representing the message to be transmitted (X) and another representing the message received (Y). The noise in the channel would then be given by the conditional distribution of received messages given the transmitted messages ($P_Y(Y|X)$). The capacity of this channel is then given by

$$C = \max_{P_X} I(X;Y).$$

The rate of a given code is given by the number of bits needed to represent X divided by the number of bits needed to represent Y, so if to send a one-bit message x we must transmit a three-bit codeword y, our code would have a rate of $1/3$. The noisy-channel coding theorem[10] then states

[10] MacKay, D. J. C. (2003). *Information theory, inference and learning algorithms*. Cambridge university press

Theorem 1. *For every discrete memoryless channel with capacity C, for any $\epsilon > 0$, any rate $R < C$, and for large enough N, there exists a code of length N and rate $\leq R$ and a decoding algorithm such that the maximal probability of block error is ϵ.*

Before Shannon's work, it was generally believed that to achieve a vanishingly small error one would need a code with vanishingly small rate. The theorem shows, however, that one can achieve any rate below the channel capacity asymptotically.

Shannon's work had profound impacts throughout science. Horace Barlow proposed to use the redundancy of a code as a measure of its inefficiency. The redundancy of a code is given by

$$\mathcal{R} = 1 - \frac{I(X;Y)}{C},$$

and it quantifies how *efficiently* a given code encodes codewords x into messages y. Note that in the case of a noiseless channel, this reduces to

$$\mathcal{R} = 1 - \frac{H(X)}{C} = 1 - \frac{H(Y)}{C}.$$

The limit given by theorem 1 then gives us the perfect redundancy-free channel. The *efficient coding hypothesis*, first proposed by Barlow[11] states that sensory relays in the nervous system recode the messages to reduce the redundancy in them. Assuming that the transmission is noiseless we can then decompose the redundancy into two terms[12]

[11] Barlow, H. (1961). Possible principles underlying the transformation of sensory messages. *Sensory communication*, pages 217–234

[12] Atick, J. J. (1992). Could information theory provide an ecological theory of sensory processing? *Network: Computation in neural systems*, 3(2):213–251

$$\mathcal{R} = \frac{1}{C}\left(C - \sum_i H(X_i)\right) + \frac{1}{C}\left(\sum_i H(X_i) - H(X)\right).$$

The first term accounts for redundancy arising from unequal frequency of use of different symbols and the second accounts for redundancy arising from correlations between the symbols X_i. This can provide different approaches towards optimizing the coding mechanisms in the nervous system. In one popular example, Simon Laughlin related the distribution of contrasts in the natural environment of the blowfly to the tuning function of the large monopolar cells (LMC's) in the blowfly's visual system[13].

[13] Laughlin, S. (1981). A simple coding procedure enhances a neuron's information capacity. *Z. Naturforsch*

Since we are considering only one neuron, the analysis is somewhat simplified. Let us write the activity of the neuron as a function of the contrast c as $o = g(c)$. We will then have that the redundancy of the firing is given by

$$\mathcal{R} = \frac{1}{C}\left(C - H(O)\right),$$

which is maximized when all output levels are equally probable. Further, we can say that

$$P(o)do = P(c)dc, \text{ and therefore } o(c) = o_{max}\int_{-1}^{c} hP(c')dc',$$

as is shown in figure 1.1.

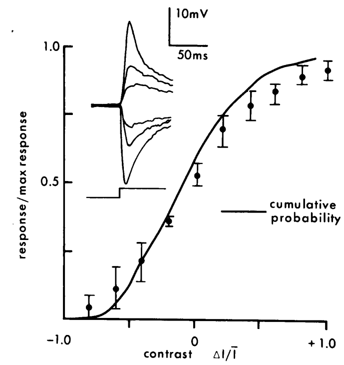

Figure 1.1: The response function of the blowfly LMC closely resembles the cumulative distribution of visual contrasts in its natural environment. Figure taken from Laughlin, 1981

1.2 *Dynamic Population Coding*

2
Filtering and Prediction with Point Process Observations

2.1 Fast Population Coding

2.2 General Filtering for Point Process Observations

2.3 Offline Inference

2.4 Prediction

3
Mean-Squared-Error for Point Process Filtering

4
Optimal Control with Point Process Observations

4.1 Stochastic Optimal Control

4.2 Linear-Quadratic-Gaussian Control

4.3 The Point Process Controller

5
Optimal Population Coding Revisited

5.1 Performance Measures

5.2 Equilbrium Analysis

5.3 Mean-Field Approach

5.4 Analytic Approaches

5.5 Performance Measures

6
Discussion

6.1 References

7
Bibliography

Atick, J. J. (1992). Could information theory provide an ecological theory of sensory processing? *Network: Computation in neural systems*, 3(2):213–251.

Barlow, H. (1961). Possible principles underlying the transformation of sensory messages. *Sensory communication*, pages 217–234.

Berkes, P., Orbán, G., Lengyel, M., and Fiser, J. (2011). Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science (New York, N.Y.)*, 331(6013):83–7.

Dobzhansky, T. (1973). Nothing in Biology Makes Sense Except in the Light of Evolution. *The American Biology Teacher*, 35(3):125–129.

Huys, Q. J. M., Zemel, R. S., Natarajan, R., and Dayan, P. (2007). Fast population coding. *Neural Computation*, 19(2):404–441.

Knill, D. C. and Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in neurosciences*, 27(12):712–9.

Laughlin, S. (1981). A simple coding procedure enhances a neuron's information capacity. *Z. Naturforsch*.

Ma, W. J., Beck, J. M., Latham, P. E., and Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature neuroscience*, 9(11):1432–8.

MacKay, D. J. C. (2003). *Information theory, inference and learning algorithms*. Cambridge university press.

Shannon, C. E. (1948). The mathematical theory of communication. 1963. *Bell System Technical Journal*, 27:379–423 and 623–656.