Toward Data Representation with Formal Spiking Neurons

Michael Gutmann
The University of Tokyo
gutmann@sat.t.u-tokyo.ac.jp

Kazuyuki Aihara
The University of Tokyo and Aihara
Complexity Modeling Project ERATO, JST
aihara@sat.t.u-tokyo.ac.jp

Abstract

Notable advances in the understanding of neural processing have been made when sensory systems were investigated from the viewpoint of adaptation to the statistical structure of its input space. For this purpose, mathematical methods for data representation were used. Here, we point out that emphasis on the input structure has happened at cost of the biological plausibility of the corresponding neuron models which process the natural stimuli. The signal transformation of the data representation methods does not correspond well to the signal transformations happening on the single cell level in neural systems. Hence, we propose here data representation by means of spiking neuron models. We formulate the data representation problem as an optimization problem and derive the fundamental quantities for an iterative learning scheme.
Keywords: Spiking neuron, encoding, decoding, learning, data representation.

1 Introduction

Science is about exploring structure and function of incompletely understood systems or phenomena. For the system “brain” or the phenomena “learning” and “vision”, great advances have been made since the debates in the early 20th century whether individual neurons are the basic elements of the nervous system or not (neuron doctrine). Since then, much emphasis has been on structure, i.e. on individual neurons or on how distinct classes of neurons are connected with each other. However, the functional aspect of these networks of neurons cannot be fully understood by its structure alone: How are the interconnected neurons marshaled to give rise to behavior? Why are the neurons as they are? Why are they connected the way they are?

These kind of questions were mostly addressed from the second half of the 20th century onwards. The brain was considered to be an information processing system. Principles and tools of signal processing and information theory were used to understand the function of some parts of the brain (e.g. the redundancy reduction hypothesis[1]). As information theory requires knowledge about the statistical structure of the information source, this approach triggered research into properties of the sensory environment, especially with respect to vision[2–4], and its link to neural processing[5,6]. For vision, in addition to the principles of information theory, other principles were used to explain its function in the form of "The early visual system might be optimized for ...", including energy expenditure[7] and minimal wiring among neurons[8].

The used data representation tools from signal processing and information theory, however, were methods that were not developed with sensory neuroscience in mind. When these methods are used in the neuroscientific framework, the following assumptions were implicitly made:

- Information is conveyed using a firing rate code.
- Neural processing is described by a linear filter.

Here, we might summarize them as linear rate-coding assumption. This assumption is of course well justified as a first approximation to reality, especially when learning from natural stimuli is involved. However, we feel it is time to re-consider it: The neural system is nonlinear and single spikes were found to be possible information carriers, at least in the early visual system of the fly[9].

In this paper, we propose data representation with a spiking neuron model. We formulate the problem in Section 2 as an optimization problem and give in Section 3.1 a detailed derivation of the key quantities for an iterative optimization rule. In Section 3.2 we give an interpretation, and a summary in Section 4 concludes the paper. We put here emphasis on the mathematical derivation. More detailed analysis of the resulting online rule and its properties along with application examples will be given elsewhere (in preparation: Learning encoding and decoding kernels for data representation with a spiking neuron).

2 Problem formulation

First, we specify the neuron model we are working with. Then, we put the aim "data representation with spiking neurons" into mathematical form. This is done by means of a cost functional, which needs to be minimized to accomplish data representation.

Neuron model The assumed neural model is closely related to the SMR_0 model[10] so that the equation for the membrane voltage u is

$$u(t) \quad = \quad \underbrace{\eta_0 \exp\left[-\frac{t-\hat{t}}{\tau}\right]}_{I_r(t)} + \underbrace{\int_0^{\min\{t,T_w\}} x(t-s)w(s)\mathrm{d}s}_{I(t)} + I_n(t), \qquad (1)$$

where $I_n(t)$ is a noise current, $\hat{t}$ is the last spike timing before time t, and w is the unknown encoding filter of length T_w, to be learned for the minimization of the aforementioned cost functional. The convolution of input x with encoding filter w produces the input current I. Spike timings $\{t^f; f = 1, \ldots\}$ are defined by $u(t^f) = \theta$, where $\theta > 0$ is a fixed threshold. The remaining constants are the recovery time constant τ of the recovery current I_r and the reset amount $\eta_0 < 0$.

Reconstruction From the obtained spike timings $\{t^f\}$, we aim at linearly reconstructing the stimulus x via

$$\hat{x}(t) = \sum_{f: t-T_p<t^f<t+T_d} h(t-t^f), \tag{2}$$

where h is the decoding filter, also to be learned for the minimization of the cost functional, and T_d the estimation time delay. For the spikes generated prior to t, only those within a time window of length T_p before t are considered for the reconstruction.

From Equation (2) we see that the arguments for h are in the range $[-T_d \quad T_p]$. For a good decoder $h(-T_d) = h(T_p) = 0$ should hold. The role of $h(s)$ is different for $s > 0$ and $s < 0$. For $s > 0$, the input at t is predicted from a spike event at $t^f < t$. On the other hand, for $s < 0$, the input is reconstructed from a later spike event at $t^f > t$.

Cost functional Both the encoding filter w and the decoding filter h are unknown. They are determined in order to minimize the cost functional

$$J(w,h) = \underbrace{\frac{1}{2T}\int_0^T (\hat{x}(t)-x(t))^2 \mathrm{d}t}_{\text{Reconstruction error}} + \underbrace{\frac{\alpha}{2}\int_0^{T_w} w(t)^2 \mathrm{d}t.}_{\text{Energy cost}} \tag{3}$$

T is a fixed time horizon, and α weights the energy cost. This optimization problem implements the above formulated aim of data representation with spiking neurons in mathematical form. It is quadratic in h and hence a standard problem, but on the other hand not trivial in w.

3 The functional derivative $\delta J/\delta w$

Gradient based methods are often used for optimization. We obtain here the expression for the functional derivative $\delta J/\delta w$ for the optimization with respect to w when the decoder h is held fixed.

3.1 Derivation

Let us start by perturbating $w(s)$ to $w(s) + \delta w(s)$ where

$$\delta w(s) = \epsilon \varphi(s), \tag{4}$$

for a small constant $\epsilon > 0$ and a sufficiently smooth, but else arbitrary function $\varphi(s)$. The perturbation δw causes a perturbation δt^f in the spike timings which in turn causes a perturbation $\delta\hat{x}(t)$ of the reconstruction $\hat{x}$. The resulting perturbation δJ of the cost functional J is

$$\begin{aligned} \delta J \quad = \quad & \frac{1}{T}\int_0^T (\hat{x}(t)-x(t))\delta\hat{x}(t)\mathrm{d}t + \alpha\int_0^{T_w} w(t)\delta w(t)\mathrm{d}t \\ & + \frac{1}{2T}\int_0^T (\delta\hat{x}(t))^2\mathrm{d}t + \frac{\alpha}{2}\int_0^{T_w} (\delta w(t))^2 \mathrm{d}t. \end{aligned} \tag{5}$$

The perturbation $\delta\hat{x}(t)$ We use the chain rule to obtain

$$\delta\hat{x}(t) = \sum_f \frac{\partial\hat{x}(t)}{\partial t^f}\delta t^f. \tag{6}$$

For a fixed spike index f, Equation (2) leads to

$$\frac{\partial\hat{x}(t)}{\partial t^f} = \begin{cases} -\dot{h}(t-t^f) & t^f - T_d < t < t^f + T_p \\ 0 & \text{else} \end{cases} \tag{7}$$

In order to evaluate Equation (6) and thus Equation (5), we must know the perturbation δt^f.

The perturbation δt^f The spike timing $t^f > T_w$ is defined by $u(t^f) = \theta$, i.e.

$$\theta \;=\; \eta_0 \exp\left[-\frac{t^f - t^{f-1}}{\tau}\right] + \int_0^{T_w} x(t^f - s)w(s)\mathrm{d}s + I_n(t^f). \tag{8}$$

After perturbation of $w(s)$, me make for δt^f the following ansatz

$$\delta t^f = \epsilon a_f + o(\epsilon^2), \tag{9}$$

where a_f needs to be determined. The implicit equation for a_f is given by

$$\begin{aligned}\theta \;=\; & \eta_0 \exp\left[-\frac{t^f - t^{f-1}}{\tau}\right]\exp\left[-\frac{\epsilon(a_f - a_{f-1}) + o(\epsilon^2)}{\tau}\right] \\ & + \int_0^{T_w} x(t^f + \epsilon a_f + o(\epsilon^2) - s)(w(s) + \epsilon\varphi(s))\mathrm{d}s \\ & + I_n(t^f + \epsilon a_f + o(\epsilon^2)).\end{aligned} \tag{10}$$

The constant $\epsilon > 0$ can be made arbitrarily small, so that via Taylor series and Equation (8) we obtain

$$\begin{aligned}0 \;=\; & \epsilon\Bigg\{a_f\left(\frac{-\eta_0}{\tau}\exp\left[-\frac{t^f - t^{f-1}}{\tau}\right] + \dot{I}_n(t^f) + \int_0^{T_w}\dot{x}(t^f - s)w(s)\mathrm{d}s\right) \\ & + \frac{a_{f-1}\eta_0}{\tau}\exp\left[-\frac{t^f - t^{f-1}}{\tau}\right] + \int_0^{T_w} x(t^f - s)\varphi(s)\mathrm{d}s\Bigg\} + o(\epsilon^2).\end{aligned} \tag{11}$$

From Equation (11), we see that a_f must satisfy

$$a_f = -\frac{\int_0^{T_w} x(t^f - s)\varphi(s)\mathrm{d}s}{\dot{u}(t^f)} + \gamma(t^f, t^{f-1})a_{f-1} \tag{12}$$

where

$$\gamma(t^f, t^{f-1}) \;=\; \frac{-\eta_0}{\tau\dot{u}(t^f)}\exp\left[-\frac{t^f - t^{f-1}}{\tau}\right], \tag{13}$$

$$\begin{aligned}\dot{u}(t^f) \;=\; & \frac{-\eta_0}{\tau}\exp\left[-\frac{t^f - t^{f-1}}{\tau}\right] + \dot{I}_n(t^f) \\ & + \int_0^{T_w}\dot{x}(t^f - s)w(s)\mathrm{d}s.\end{aligned} \tag{14}$$

Finally, we see from Equation (12) that a_f has the form

$$a_f = \int_0^{T_w} y_f(s)\varphi(s)\mathrm{d}s, \tag{15}$$

so that we obtain the update rule for y_f

$$y_f(s) = \frac{-x(t^f - s)}{\dot{u}(t^f)} + \gamma(t^f, t^{f-1})y_{f-1}(s), \tag{16}$$

and δt^f is given by

$$\delta t^f = \epsilon \int_0^{T_w} y_f(s)\varphi(s)\mathrm{d}s + o(\epsilon^2). \tag{17}$$

Calculation of the functional derivative $\delta J/\delta w$ The derived Equation (17) allows with Equation (7) to evaluate Equation (6), and hence to obtain δJ via Equation (5).

Equation (6) for $\delta\hat{x}(t)$ introduces a sum over the spike timings t^f into the integral

$$M = \int_0^{T} (\hat{x}(t) - x(t))\delta\hat{x}(t)\mathrm{d}t \tag{18}$$

of Equation (5). Since there are only finitely many spikes during the time interval $[0 \quad T]$, the sum can only have finitely many terms. We interchange thus the summation and the integration to obtain with Equation (7)

$$M = -\sum_f \underbrace{\int_{t^f - Td}^{t^f + T_p} (\hat{x}(t) - x(t))\dot{h}(t - t^f)\mathrm{d}t}_{\bar{e}(t^f)} \cdot \int_0^{T_w} \epsilon y_f(s)\varphi(s)\mathrm{d}s + o(\epsilon^2) \tag{19}$$

The quadratic terms in Equation (5) yield terms in the order of ϵ^2 so that we obtain with the previous equation for M

$$\delta J = \epsilon\left(-\frac{1}{T}\sum_f \bar{e}(t^f)\int_0^{T_w} y_f(s)\varphi(s)\mathrm{d}s + \alpha\int_0^{T_w} w(s)\varphi(s)\mathrm{d}s\right) + o(\epsilon^2) \tag{20}$$

Taking the principal linear part[11], we obtain the final expression for the functional derivative of J with respect to w

$$\frac{\delta J}{\delta w(s)} = -\frac{1}{T}\sum_f \bar{e}(t^f)y_f(s) + \alpha w(s) \tag{21}$$

3.2 Interpretation

For the interpretation of Equation (21), we first transform $\bar{e}(t^f)$ by change of variables and partial integration, using $h(-T_d) = h(T_p) = 0$, into

$$\bar{e}(t^f) = -\int_{-T_d}^{T_p} \dot{e}(t^f + s)h(s)\mathrm{d}s. \tag{22}$$

The rate of the error $e = \hat{x} - x$ is averaged with a weighting given by the reconstruction filter h. Equation (2) shows that the spike timing t^f contributes via $h(s)$ to the reconstruction at $t^f + s$. Hence, the weighting is such that $\bar{e}(t^f)$ indicates the reconstruction error caused by spike timing t^f on the time interval $[t^f - Td \quad t^f + T_p]$.

Via $y_f(s)$ and Equation (16), there is a recursion inherent in Equation (21). We solve this recursion to allow for better interpretation of the functional derivative $\delta J/\delta w(s)$. Grouping the terms with the common factor $x(t^k - s)$ together, we obtain

$$\frac{\delta J}{\delta w(s)} = -\frac{1}{T} \sum_f x(t^f - s) \frac{\tilde{e}(t^f)}{\dot{u}(t^f)} + \alpha w(s), \tag{23}$$

with

$$\tilde{e}(t^f) = \bar{e}(t^f) + \sum_{p \geq f+1} \gamma(t^{f+1}, t^f) \ldots \gamma(t^p, t^{p-1}) \bar{e}(t^p). \tag{24}$$

The expression $\tilde{e}(t^f)$ can be interpreted as the total reconstruction error caused by the spike timing t^f. Equation (23) shows then that the functional derivative $\delta J/\delta w(s)$ is the difference between αw and the correlation between the total reconstruction error caused by spike timing t^f and the normalized input $x(t^f - s)/\dot{u}(t^f)$ at time s before the spike.

4 Summary

In the ongoing search for understanding of early sensory systems notable advances have been made through data representation methods applied to natural stimuli. We pointed out that the corresponding neuron models, which process the natural stimuli, tend however to be abstract and might be a limiting factor for further advances. This is especially so when the aim is to connect to experimental results at the single-cell level. Here, we made a small step to biologically more plausible models and considered data representation by means of a spiking neuron. Requiring linear reconstructability of the input from the spike train, we derived in detail the essential quantities for an iterative learning rule, and discussed their meaning.

Acknowledgments

This research is partially supported by Grant-in-Aid for Scientific Research on Priority Areas - System study on higher-order brain functions - from the Ministry of Education, Culture, Sports, Science and Technology of Japan (17022012). MG is further supported by MEXT grant No. 040680.

References

1. H.B. Barlow. Possible principles underlying the transformations of sensory messages. In W. Rosenblith, editor, *Sensory Comm.*, pages 217–234. MIT Press, 1961.

2. D. Dong and J. Atick. Statistics of natural time-varying images. *Network: Comp. in Neural Systems*, 6(3):345–358, August 1995.

3. R.A. Frazor and W.S. Geisler. Local luminance and contrast in natural images. *Vision Research*, 46(10):1585–1598, May 2006.

4. D. Ruderman. The statistics of natural images. *Network: Comp. in Neural Systems*, 5(4):517–548, November 1994.

5. A. Hyvärinen, P. O. Hoyer, J. Hurri, and M. Gutmann. Statistical models of images and early vision. In *Proc. of the Int. Symp. on Adaptive Knowledge Representation and Reasoning*, 2005.

6. E.P. Simoncelli and B. A. Olshausen. Natural image statistics and neural representation. *Ann. Rev. Neuroscie.*, 24(1):1193–1216, 2001.

7. B. Vincent, R. Baddeley, T. Troscianko, and I.D. Gilchrist. Is the early visual system optimised to be energy efficient? *Network: Comp. in Neural Systems*, 16(2 - 3):175–190, June 2005.

8. D.J. Graham, D.M. Chandler, and D.J. Field. Can the theory of "whitening" explain the center-surround properties of retinal ganglion cell receptive fields? *Vision Research*, 46(18):2901–2913, September 2006.

9. F. Rieke, D. Warland, R. de R. van Steveninck, and W. Bialek. *Spikes: Exploring the neural code*. MIT Press, 1997.

10. W. Gerstner and W. K Kistler. *Spiking Neuron Models*. Cambridge University Press, 2002.

11. I.M Gelfand and S.V. Fomin. *Calculus of Variations*. Dover, 2000.