Proceedings of Machine Learning Research vol 125:1–5, 2020 33rd Annual Conference on Learning Theory

ODE-Inspired Analysis for the Biological Version of Oja's Rule in Solving Streaming PCA

Chi-Ning Chou CHININGCHOU@G.HARVARD.EDU
Harvard University

Mien Brabeeba Wang BRABEEBA@MIT.EDU
Massachusetts Institute of Technology

Editors: Jacob Abernethy and Shivani Agarwal

[1]Abstract

Oja's rule [Oja, Journal of mathematical biology 1982] is a well-known biologically-plausible algorithm using a Hebbian-type synaptic update rule to solve streaming principal component analysis (PCA). Computational neuroscientists have known that this biological version of Oja's rule converges to the top eigenvector of the covariance matrix of the input in the limit. However, prior to this work, it was open to prove any convergence rate guarantee.

In this work, we give the first convergence rate analysis for the biological version of Oja's rule in solving streaming PCA. Moreover, our convergence rate matches the information theoretical lower bound up to logarithmic factors and outperforms the state-of-the-art upper bound for streaming PCA. Furthermore, we develop a novel framework inspired by ordinary differential equations (ODE) to analyze general stochastic dynamics. The framework abandons the traditional *step-by-step* analysis and instead analyzes a stochastic dynamic in *one-shot* by giving a closed-form solution to the entire dynamic. The one-shot framework allows us to apply stopping time and martingale techniques to have a flexible and precise control on the dynamic. We believe that this general framework is powerful and should lead to effective yet simple analysis for a large class of problems with stochastic dynamics.

Keywords: Theoretical neuroscience, streaming PCA, stochastic process, stopping time, dynamical system

1. Introduction

Brains processes high dimensional visual inputs constantly. In our eyes, 100 millions photoreceptors in the retina receive gigabytes of information per second Wandell (1995). In addition, the retina is a highly convergent pathway: 100 million photoreceptors converges the visual information onto one million retina ganglion cells in optical nerves Ganguli and Sompolinsky (2012). Therefore, it is important to understand a neural implementation of the dimensionality reduction in the retina. Furthermore, many works in theoretical neuroscience Atick and Redlich (1990) demonstrated from the efficient coding principle that the retina might implement Principal Component Analysis (PCA). Specifically, they showed that under natural image statistics, PCA-like solution recovers the center-surround receptive fields in the retina. However, their work only proposed PCA as an potential solution to the pathway and did not provide a dynamic to explain the learning process of PCA.

---

[1] Extended abstract. Full version appears on the arXiv as [`https://arxiv.org/abs/1911.02363`, v2]

On the other hand, in the seminal work of Oja (1982), he proposed a mathematical model for the biological neural network that solves streaming PCA with several biologically-plausible properties: the network not only updates its synaptic weights locally but also normalizes the strength of synapses. This rule, now known as the biological version of Oja's rule (*biological Oja's rule*[2] in abbreviation), has been the subject of extensive theoretical Oja (1982) and experimental Swinehart and Abbott (2006) studies aimed at understanding its performance. Despite its popularity, the theoretical understanding of the biological Oja's rule cannot account for the speed of the convergence because the state-of-the-art theoretical analysis only provides a guarantee on convergence in the limit via Kushner-Clark theory Duflo (2013).

In biology, the retina can change its receptive field to adapt to environments with different illumination Shapley and Enroth-Cugell (1984), contrast Shapley and Enroth-Cugell (1984), spatial frequency , orientation and temporal correlation Hosoya et al. (2005) in the time scale of seconds despite the high dimensional input. This suggests that a plausible dynamic for explaining the retina-optical nerve pathway should have little or no dependency on the dimension, *i.e.,* the number of neurons, which in this case is on the order of 100 million. Meanwhile, researchers have observed that the biological Oja's rule (and its variants) has fast convergence rates Swinehart and Abbott (2006) in simulations. Thus, to demonstrate the feasibility of fast retina adaptation in theory, it is important to give an analysis to show that the biological Oja's rule solves streaming PCA in a biologically-realistic time scale. This is nevertheless a challenging task and has remained elusive for almost 40 years Oja (1982).

In this paper, we provide the first convergence rate guarantee for the biological Oja's rule in solving streaming PCA. Furthermore, the convergence rate matches the information-theoretic lower bound for streaming PCA up to logarithmic factors. In terms of the techniques, we develop an ODE-inspired framework to analyze stochastic dynamics. We believe this general framework of using tools and insights from ODE and SDE in analyzing stochastic dynamics is elegant and powerful. Also, as a byproduct, our convergence rate guarantee for biological Oja's rule outperforms the state-of-the-art upper bound for streaming PCA (using a non-biologically plausible variant of Oja's rule) Allen-Zhu and Li (2017).

Biological Oja's rule. Oja (1982) proposed a streaming PCA algorithm using n input neurons and one output neuron. The firing rates of the input neurons at time t are denoted by a vector $\mathbf{x}_t \in \mathbb{R}^n$ and the firing rate of the output neuron is denoted by a scalar $y_t \in \mathbb{R}$. The synaptic weights at time t from the input neurons to the output neuron are denoted by a vector $\mathbf{w}_t \in \mathbb{R}^n$. Note that the weight vector will be the output and ideally it will converge to the top eigenvector $\mathbf{v}_1$.

The input stream $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T$ arrives in the form of firing rates of the input neurons. The firing rate of the output neuron is simply the inner product of the synaptic weight vector and the firing rate vector of the input neurons, *i.e.,* $y_t = \mathbf{x}_t^\top \mathbf{w}_{t-1}$. Now, from the biological Oja's rule, the dynamic of the synaptic weight vector is described by the following equation.

Definition 1 (Biological Oja's rule) *For any initial vector* $\mathbf{w}_0 \in \mathbb{R}^n$ *such that* $\|\mathbf{w}_0\|_2 = 1$*, the dynamic of the biological Oja's rule is the following. For any* $t \in \mathbb{N}$*, define*

$$\mathbf{w}_t = \mathbf{w}_{t-1} + \eta_t y_t \left(\mathbf{x}_t - y_t \mathbf{w}_{t-1}\right) \tag{1.1}$$

---

[2] Also known as *Oja's rule* in the literature. However, many works in the machine learning community use the name "Oja's rule" for *non-biologically-plausible* variants of the original Oja's rule. Thus, in this paper we emphasize the term "*biological*" to distinguish the two.

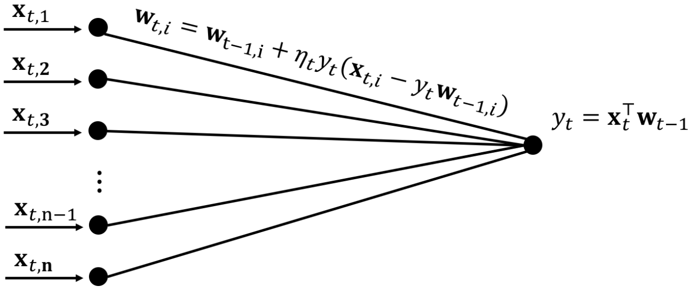

Figure 1: Biological Oja's rule is a local synaptic learning rule

where $y_t = \mathbf{x}_t^\top \mathbf{w}_{t-1}$ *and* $\mathbf{x}_t$ *is the input at time* t.

Notice that biological Oja's rule is local, i.e. the synaptic weight updates only depends on the activities of the two ends neurons and itself (Figure 1). The locality is an important criteria for the biological plausibility of a synaptic learning rule Hebb (1949).

Our results There are two common convergence notions in the streaming PCA literature. The *global convergence* requires the algorithm/dynamic to start from a random initial vector while the *local convergence* allows the algorithm/dynamic to start from an initial vector that is highly correlated to the top eigenvector of the covariance matrix. Now, we are ready to state our main theorem as follows.

Theorem 2 (Global and local convergence) *With the setting in the dynamic in Definition 1, let* $\boldsymbol{gap} := \lambda_1 - \lambda_2 > 0$. *For any* $\epsilon, \delta \in (0,1)$, *we have the following results.*

- *(Local Convergence) Suppose* $\frac{\langle \mathbf{w}_0, \mathbf{v}_1\rangle^2}{\|\mathbf{w}_0\|_2^2} = \Omega(1)$. *For any* $n \in \mathbb{N}$, $\delta, \epsilon \in (0,1)$), *let*

$$\eta = \tilde{\Theta}\left(\frac{\epsilon \boldsymbol{gap}}{\lambda_1}\right), \; T = \Theta\left(\frac{\lambda_1}{\epsilon \boldsymbol{gap}^2} \cdot \log^2\left(\frac{1}{\epsilon}, \frac{1}{\delta}\right)\right).$$

Then, we have

$$\Pr\left[\frac{\langle \mathbf{w}_T, \mathbf{v}_1\rangle^2}{\|\mathbf{w}_T\|_2^2} < 1 - \epsilon\right] < \delta\,.$$

- *(Global Convergence) Suppose* $\mathbf{w}_0$ *is uniformly sampled from the unit sphere of* $\mathbb{R}^n$. *For any* $n \in \mathbb{N}$, $\delta, \epsilon \in (0,1)$, *let*

$$\eta = \tilde{\Theta}\left(\frac{(\epsilon \wedge \delta^2)\boldsymbol{gap}}{\lambda_1}\right), \; T = \Theta\left(\frac{\lambda_1}{(\epsilon \wedge \delta^2)\boldsymbol{gap}^2} \cdot \log^3\left(\frac{1}{\epsilon}, \frac{1}{\delta}, \frac{1}{\boldsymbol{gap}}, n\right)\right).$$

Then, we have

$$\Pr\left[\frac{\langle \mathbf{w}_T, \mathbf{v}_1\rangle^2}{\|\mathbf{w}_T\|_2^2} < 1 - \epsilon\right] < \delta\,.$$

The notation $a \wedge b$ *stands for* $\min\{a, b\}$ *and* $\tilde{\Theta}$ *hides the poly-logarithmic factors with respect to* $\epsilon^{-1}, \delta^{-1}, \boldsymbol{gap}^{-1}, n$.

General Framework One of the main difficulty to analyze the dynamic of biological Oja's rule is the nonlinearity of the update rule. This makes the traditional *step-by-step* expectation approach unusable because the improvement factor at each step can depend on the process itself and at worst case, the dynamic can show no improvement or even deteriorate. Taking expectation loses precise controls of the values of the process. This makes naive martingale analysis difficult to work. We use an ODE-inspired *one-shot* framework to solve this problem. Our high-level strategy is to first consider the *continuous* version of the Oja's rule where the learning rate η is set to be infinitesimal. In the continuous setting, the dynamic can be fully understood by tools from the theory of ordinary differential equations (ODE) or stochastic differential equations (SDE). The continuous analysis helps us not only to understand how to analyze the dynamic via different linearization but also to write down a closed form solution of the dynamic where the noise is adapted. Since we can write down a closed form solution for the dynamic, by bounding the noise, we can guarantee the improvement in *one-shot*. Next, by using stopping time technique, we are able to gain precise control of the process itself and therefore give a tight bound on the stopped concentration of the noise. Finally, by exploiting the structure of the dynamic, we are able to pull out the stopping time without introducing additional failure probability to recover a tight concentration on the original noise and therefore guarantee the improvement. This framework is conceptually elegant and powerful. We refer readers to the full version of this paper and a subsequent work Chou et al. (2020) focusing on the framework for more details.

Biological Perspectives. Biological Oja's rule is one of the earliest local learning rules that incorporate both *Hebbian* and *homeostatic plasticity* Oja (1982), two major activity-dependent synaptic modification mechanisms Abbott and Nelson (2000). Both mechanisms work together to form memory and drive learning behaviors in the brain. Hebbian plasticity is a synapse-specific correlation-based plasticity mechanism that strengthens the connection when the input has a high correlation with the weights while weakening the connection when the input has a poor correlation Kelso et al. (1986); Dudek and Bear (1992). However, this type of mechanism alone can often make networks unstable since the highly correlated input will keep strengthening synapses unboundedly Abbott and Nelson (2000). Homeostatic plasticity, in contrast, stabilizes the network by keeping the activities of the neurons relatively constant through calcium sensors Turrigiano (2008). Synaptic scaling is a specific kind of homeostatic plasticity where the strength of the incoming synapses is normalized while still encoding the information from Hebbian learning in their relative strength after normalization Turrigiano (2008). Oja's rule is one example of this. Concretely, Oja's rule can be expressed as the following

$$w_t = w_{t-1} + \eta_t(x_t y_t - y_t^2 w_{t-1})\,.$$

One can see that $x_t y_t$ term corresponds to the Hebbian plasticity while $y_t^2 w_{t-1}$ term corresponds to the homeostatic plasticity. One can then show the synaptic scaling property where $\|w_t\| \approx 1$ for all t. In this sense, this work demonstrates that a biologically plausible mechanism can solves streaming PCA in a biologically-realistic timescale to explain fast retina adaptation.

Acknowledgments

CNC is supported by by NSF awards CCF 1565264 and CNS 1618026. MBW is supported by NSF Awards CCF-1810758, CCF-0939370, CCF-1461559 and Akamai Presidential Fellowship. We thank Nancy Lynch for valuable comments on the presentation of this paper, thank Kai-Min

Chung for helpful discussions, and thank Rohit Agrawal, Boaz Barak, Chun-Hsiang Chan, Yanlin Chen, and Santhoshini Velusamy for comments on the draft of this paper. We thank Nancy Lynch again for organizing a brain algorithm reading group at MIT in Spring 2019 and all the participants for the inspiring conversation.

References

Larry F. Abbott and Sacha B. Nelson. Synaptic plasticity: taming the beast. *Nature Neuroscience*, 3:1178–1183, 2000.

Zeyuan Allen-Zhu and Yuanzhi Li. First efficient convergence for streaming k-pca: a global, gap-free, and near-optimal rate. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 487–492. IEEE, 2017.

Joseph J. Atick and A. Norman Redlich. Towards a theory of early visual processing. *Neural Computation*, 2:308–320, 1990.

Chi-Ning Chou, Mien Brabeeba Wang, and Tiancheng Yu. A general framework for analyzing stochastic dynamics in learning algorithms. *arXiv:2006.06171 [math.OC]*, 2020.

Serena M. Dudek and Mark F. Bear. Homosynaptic long-term depression in area CA1 of hippocampus and effects of N-methyl-D-aspartate receptor blockade. *Proceedings of the National Academy of Sciences of the United States of America*, 89:4363–4367, 1992.

Marie Duflo. *Random iterative models*, volume 34. Springer Science & Business Media, 2013.

Surya Ganguli and Haim Sompolinsky. Compressed sensing, sparsity and neural data. *Annual review of neuroscience*, 35(1):463–483, 2012.

Donald O. Hebb. *The Organization of Behavior*. Wiley, New York, 1949.

Toshihiko Hosoya, Stephen A. Baccus, and Markus Meister. Dynamic predictive coding by the retina. *Nature*, 436:71–77, 2005.

Stephan R. Kelso, Alan H. Ganong, and Thomas H. Brown. Hebbian synapses in hippocampus. *Proceedings of the National Academy of Sciences of the United States of America*, 83:5326–5330, 1986.

Erkki Oja. Simplified neuron model as a principal component analyzer. *Journal of mathematical biology*, 15(3):267–273, 1982.

Robert Shapley and Christina Enroth-Cugell. Visual adaptation and retinal gain controls. *Progress in Retinal Research*, 3:263–346, 1984.

Christian D. Swinehart and Larry F. Abbott. Dimensional reduction for reward-based learning. *Network: Computation in Neural Systems*, 17(3):235–252, 2006.

Gina G. Turrigiano. The self-tuning neuron: Synaptic scaling of excitatory synapses. *Cell*, 135(3): 422–435, 2008.

Brian A. Wandell. *Foundations of vision*. Sunderland, MA: Sinauer, 1995.