

Natural gradient learning for spiking neurons

Elena Kreutzer
kreutzer@pyl.unibe.ch
Department of Physiology
Bern, Switzerland

Mihai A. Petrovici
petrovici@pyl.unibe.ch
Department of Physiology
Bern, Switzerland

Walter Senn
senn@pyl.unibe.ch
Department of Physiology
Bern, Switzerland

ABSTRACT

Due to their simplicity and success in machine learning, gradient-based learning rules represent a popular choice for synaptic plasticity models. While they have been linked to biological observations, it is often ignored that their predictions generally depend on a specific representation of the synaptic strength. In a neuron, the impact of a synapse can be described using the state of many different observables such as neurotransmitter release rates or membrane potential changes. Which one of these is chosen when deriving a learning rule can drastically change the predictions of the model. This is doubly unsatisfactory, both with respect to optimality and from a conceptual point of view. By following the gradient on the manifold of the neuron’s firing distributions instead of one that is relative to some arbitrary synaptic weight parametrization, natural gradient descent provides a solution to both these problems. While the computational advantages of natural gradient are well-studied in ANNs, its predictive power as a model for in-vivo synaptic plasticity has not yet been assessed. By formulating natural gradient learning in the context of spiking interactions, we demonstrate how it can improve the convergence speed of spiking networks. Furthermore, our approach provides a unified, normative framework for both homo- and heterosynaptic plasticity in structured neurons and predicts a number of related biological phenomena.

CCS CONCEPTS

• **Computing methodologies** → **Modeling and simulation; Neural networks; Learning paradigms.**

KEYWORDS

parametrization invariance, efficient learning, natural gradient descent, dendritic learning, heterosynaptic plasticity, homeostasis

ACM Reference Format:

Elena Kreutzer, Mihai A. Petrovici, and Walter Senn. 2020. Natural gradient learning for spiking neurons. In *NICE '20: Neuro-Inspired Computational Elements Workshop, March 17–20, 2020, Heidelberg, Germany*. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

NICE '20, March 17–20, 2020, Heidelberg, Germany

© 2020 Association for Computing Machinery.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$15.00

<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 NATURAL-GRADIENT PLASTICITY RULE

In an ongoing effort to provide a normative framework for synaptic plasticity, a multitude of top-down computational paradigms have been developed, which derive adaptation rules for synaptic weights as gradient descent on a particular objective function of the studied neural network. However, the exact physical quantity to which these synaptic weights correspond often remains unspecified. Depending on the particular choice of the modeler, the network behavior during learning can be markedly different.

As many aspects of network connectivity that arise during development – such as the exact contact point of afferents along the dendritic tree of a neuron – are largely random, the question of invariant learning also represents a practical problem for a brain that strives for optimal learning. What if neurons would seek to conserve the way they adapt their behavior regardless of, e.g., the specific positioning of synapses along their dendritic tree? What synaptic learning rule is able to fulfill this requirement?

First studied by Amari in the context of information geometry [2], natural gradient learning has the appealing property of alleviating such parametrization-related concerns by addressing the question of steepest descent at the level of neuronal outputs. By using the manifold of neuronal output distributions as a support for the cost function, parametrization invariance is achieved implicitly.

We consider Poisson neurons with firing rates that are nonlinear functions ϕ of the somatic membrane potential $V = \sum_i w_i x_i^\epsilon$, with synaptic weights w_i and inputs x_i^ϵ . Given a target spike train Y^* provided by a teacher neuron with distribution p^* , synapses should efficiently minimize the cost function $C = D_{\text{KL}}(p^*, p)$. By choosing the Fisher information G as a metric on the space of output distributions p , we can write the natural gradient as a linear transformation of the Euclidean gradient: $\nabla^N = G^{-1} \nabla^E$. From here, we can derive a synaptic plasticity rule of the form

$$\dot{\mathbf{w}} = \eta \gamma_s [Y^* - \phi(V)] \frac{\phi'(V)}{\phi(V)} \left(\frac{c_\epsilon \mathbf{x}^\epsilon}{\mathbf{r}} - \gamma_u + \gamma_w \mathbf{w} \right), \quad (1)$$

which recovers the error-correcting dynamics of other gradient learning frameworks [9, 13], but extends them in several important respects, as we discuss below. While the coefficients $\gamma_s, \gamma_u, \gamma_w$ are in general more complex functions of the student neuron’s dynamics, we have shown that, in many cases, they can be well approximated by simpler quantities that are locally accessible at every synapse.

2 BIOLOGICAL PREDICTIONS

Fig. 1 shows simulation results for a learning scenario using the natural-gradient plasticity rule derived above. By performing optimization directly on the space of output distributions, it always follows a locally straight path towards the local optimum (Fig. 1C,D), which is a hallmark of parametrization invariance. This also helps

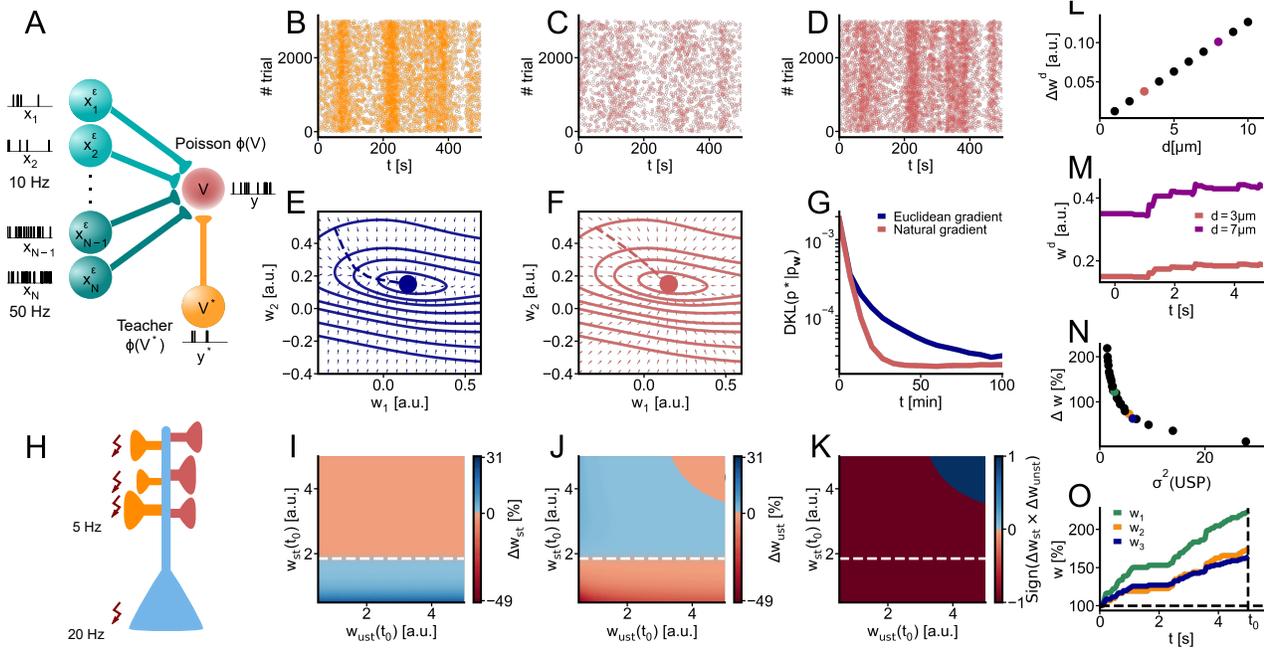


Figure 1: (A) Sketch of the learning task. A student neuron (red) learns to combine its afferents (teal) in a way that optimally reproduces the statistics of a teacher signal (yellow). (B-D) Spike trains of teacher (B), and student neuron before learning (C) and after learning with the natural gradient rule (D). (E,F) Cost function contour lines (solid), gradients (arrows) and trajectory (dashed) during learning for Euclidean (blue) and natural (red) gradient learning. (G) Convergence speed. (H) Sketch of simulated setup for evaluating homo- and heterosynaptic plasticity. Input and teacher signals were Poissonian, with mean rates of 5 and 20 Hz, respectively. Stimulated synapses shown in yellow, unstimulated ones in red. (I-K) Instantaneous weight changes (heatmaps) as a function of momentary stimulated and unstimulated synaptic weights (axes). (I) shows plasticity at stimulated synapses (homosynaptic), (J) at unstimulated ones (heterosynaptic) and (K) compares the sign of the two, showing that for most of the relevant phase space, homo- and heterosynaptic plasticity oppose each other. (L) Dendritic synapses exhibit larger weight changes the further they are from the soma (dendritic democracy). (M) Exemplary traces from (L). (N) Synaptic weight changes become larger, the more reliable (smaller variance) their input is. (O) Exemplary traces from (N).

learning converge significantly faster than with Euclidean gradients, as shown in Fig. 1E (with optimal learning rates for both plasticity rules).

Natural-gradient plasticity exhibits distance-dependent scaling, which can be observed when expressing the weight update directly in terms of EPSP amplitudes in the dendritic tree, rather than their somatic counterparts. Since synaptic changes are attenuated on their way to the soma, in order to evoke the same effect in terms of adapting the neuron’s firing behavior, changes at distal synapses must be increased compared to plasticity at proximal synapses with a similar impact on output firing, as also suggested by dendritic democracy experiments [5].

Just like in Euclidean gradient descent, our learning rule follows the postsynaptic error $[Y^* - \phi(V)]$, and its input-adjusted term $\frac{c_e x^e}{r}$ encourages modification of synapses that have recently been active. However, in contrast to classical gradient-based learning, this homosynaptic weight update is complemented by both a uniform as well as a weight-proportional heterosynaptic contribution. This is in line with experimental findings [3, 7], as well as with computational studies that list heterosynaptic plasticity as a necessary component of stable learning [15]. Similarly to experimental data [10, 14], our simulations show that under natural gradient learning, changes of unstimulated synapses often have an opposite sign compared to those of stimulated synapses, thereby acting as a homeostatic mechanism that keeps the neuron close to an optimal operating point.

Furthermore, our learning rule includes a synapse-specific scaling of the homosynaptic contribution by the input rate, similar to the rescaling observed in a Bayesian plasticity framework [1]. A further prediction revealed by our simulated stimulation protocol is that the effective learning rate of our rule is inversely related to the USP variance. This is also reminiscent of other gradient-based approaches in machine learning [6].

3 CONCLUSIONS

Our learning rule provides a consistent, normative theory of synaptic plasticity that combines fast learning with multiple experimentally testable predictions, including a unified framework for both homo- and heterosynaptic changes. The bridge to biology also hints at applications on artificial neuromorphic systems, many of which offer increasing flexibility with respect to spike-based learning [12]. Compared to recently demonstrated gradient-based learning paradigms [4, 8, 11], our framework may improve the robustness and speed of convergence, ultimately reducing the required time- and energy-to-solution.

ACKNOWLEDGMENTS

We gratefully acknowledge funding from the European Union under grant agreements 604102, 720270, 785907 (HBP) and the Manfred Stärk Foundation.

REFERENCES

- [1] Laurence Aitchison and Peter E Latham. 2014. Bayesian synaptic plasticity makes predictions about plasticity experiments in vivo. *arXiv preprint arXiv:1410.1029* (2014).
- [2] Shun-Ichi Amari. 1998. Natural gradient works efficiently in learning. *Neural computation* 10, 2 (1998), 251–276.
- [3] Marina Chistiakova and Maxim Volgushev. 2009. Heterosynaptic plasticity in the neocortex. *Experimental brain research* 199, 3–4 (2009), 377.
- [4] Steven K Esser, Paul A Merolla, John V Arthur, Andrew S Cassidy, Rathinakumar Appuswamy, Alexander Andreopoulos, David J Berg, Jeffrey L McKinstry, Timothy Melano, Davis R Barch, et al. 2016. Convolutional networks for fast energy-efficient neuromorphic computing. *Proc. Nat. Acad. Sci. USA* 113, 41 (2016), 11441–11446.
- [5] Michael Häusser. 2001. Synaptic function: dendritic democracy. *Current Biology* 11, 1 (2001), R10–R12.
- [6] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
- [7] Gary S Lynch, Thomas Dunwiddie, and Valentin Gribkoff. 1977. Heterosynaptic depression: a postsynaptic correlate of long-term potentiation. *Nature* 266, 5604 (1977), 737.
- [8] Mihai A Petrovici, Sebastian Schmitt, Johann Klähn, David Stöckel, Anna Schroeder, Guillaume Bellec, Johannes Bill, Oliver Breitwieser, Ilja Bytschok, Andreas Grübl, et al. 2017. Pattern representation and recognition with accelerated analog neuromorphic systems. In *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 1–4.
- [9] Jean-Pascal Pfister, Taro Toyozumi, David Barber, and Wulfram Gerstner. 2006. Optimal spike-timing-dependent plasticity for precise action potential firing in supervised learning. *Neural computation* 18, 6 (2006), 1318–1348.
- [10] Sébastien Royer and Denis Paré. 2003. Conservation of total synaptic weight through balanced synaptic depression and potentiation. *Nature* 422, 6931 (2003), 518.
- [11] Sebastian Schmitt, Johann Klähn, Guillaume Bellec, Andreas Grübl, Maurice Guettler, Andreas Hartel, Stephan Hartmann, Dan Husmann, Kai Husmann, Sebastian Jeltsch, et al. 2017. Neuromorphic hardware in the loop: Training a deep spiking network on the brainscales wafer-scale system. In *2017 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2227–2234.
- [12] Chetan Singh Thakur Thakur, Jamal Molin, Gert Cauwenberghs, Giacomo Indiveri, Kundan Kumar, Ning Qiao, Johannes Schemmel, Runchun Mark Wang, Elisabetta Chicca, Jennifer Olson Hasler, et al. 2018. Large-scale neuromorphic spiking array processors: A quest to mimic the brain. *Frontiers in neuroscience* 12 (2018), 891.
- [13] Robert Urbanczik and Walter Senn. 2014. Learning by the dendritic prediction of somatic spiking. *Neuron* 81, 3 (2014), 521–528.
- [14] Robert Wöhr, Dorothea Von Haebler, and Uwe Heinemann. 2007. Low-frequency stimulation of the direct cortical input to area CA1 induces homosynaptic LTD and heterosynaptic LTP in the rat hippocampal–entorhinal cortex slice preparation. *European Journal of Neuroscience* 25, 1 (2007), 251–258.
- [15] Friedemann Zenke, Everton J Agnes, and Wulfram Gerstner. 2015. Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks. *Nature communications* 6 (2015), 6922.