Learning efficient backprojections across cortical hierarchies in real time

Kevin Max[1], Laura Kriener, Garibaldi Pineda García, Thomas Nowotny, Walter Senn, Mihai A. Petrovici

1. Presenting author. Department of Physiology, University of Bern, Switzerland

Summary. We present Phaseless Alignment Learning (PAL), a biologically plausible approach for learning efficient feedback weights in layered cortical hierarchies. Our dynamical system facilitates the simultaneous learning of all weights with always-on plasticity, and exclusively utilizes information locally available at the synapses. Our method is entirely phase-free, avoiding the need for forward and backward passes or phased learning, and enables efficient error propagation across multi-layer cortical hierarchies, while maintaining biologically plausible signal transport and learning.

Modulation of neural activity occurs through learning, i.e., the long-term adaptation of synaptic weights. However, it remains unresolved how weights are adapted across the cortex to effectively solve a given task. A key question is how to assign credit to synapses that are situated deep within a hierarchical network. In deep learning, backpropagation (BP) is the current state-of-the-art for solving this issue, and may potentially serve as an inspiration for neuroscience.

The application of BP to cortical processing is however non-trivial, due to several biologically implausible requirements it entails. For example, it requires information to be buffered for use at different stages of processing. Additionally, error propagation occurs through weights that must be mirrored at synapses in different layers, resulting in the weight transport problem. Furthermore, artificial neural networks (ANNs) operate in separate forward and backward phases, with inference and learning alternating strictly.

We introduce Phaseless Alignment Learning (PAL) [1], a biologically plausible technique for learning effective top-down weights across layers in cortical hierarchies. We propose that cortical networks can learn useful backward weights by utilizing a ubiquitous resource of the brain: noise. Despite being usually treated as a disruptive factor, noise can be leveraged by the feedback pathway as an additional carrier of information for synaptic plasticity.

PAL describes a fully dynamic system that effectively addresses all of the aforementioned problems: it models the dynamics of biophysical substrates, and all computations are carried out using information **locally available** at the synapses; learning occurs in a **completely phase-less** manner; **plasticity is always-on for** all synapses, both forward and backward, at all times. Our approach is consistent with biological observations and facilitates efficient learning without the need for wake-sleep phases or other forms of phased plasticity found in many other models of cortical learning.

PAL can be applied to a broad range of models and represents an improvement over previously known biologically plausible methods of credit assignment. For instance, when compared to feedback alignment (FA), PAL can solve complex tasks with fewer neurons and more effectively learn useful latent representations. We illustrate this by conducting experiments on various classification tasks using a cortical dendrite microcircuit model [2], which leverages the complexity of neuronal morphology and is capable of prospective coding [3]. For an excerpt of results, see Fig. 1 (b-h).

We argue that PAL can be realized both in biological and, more generally, physical components. Specifically, it capitalizes on the inherent noise present in physical systems and leverages simple filtering techniques to distinguish between signal and noise where necessary. A realization of PAL (or a variant) in physical form, whether in the cortex or on neuromorphic systems, constitutes an elegant solution to the weight transport problem, while enabling efficient learning with purely local computations.

References

[1] Kevin Max, Laura Kriener, Garibaldi Pineda García, Thomas Nowotny, Walter Senn, and Mihai A Petrovici. Learning efficient backprojections across cortical hierarchies in real time. *arXiv preprint arXiv:2212.10249*, 2022.

[2] João Sacramento et al. Dendritic cortical microcircuits approximate the backpropagation algorithm. *Advances in neural information processing systems*, 31, 2018.

[3] Paul Haider et al. Latent equilibrium: A unified learning theory for arbitrarily fast computation with arbitrarily slow neurons. *Advances in Neural Information Processing Systems*, 34:17839–17851, 2021.

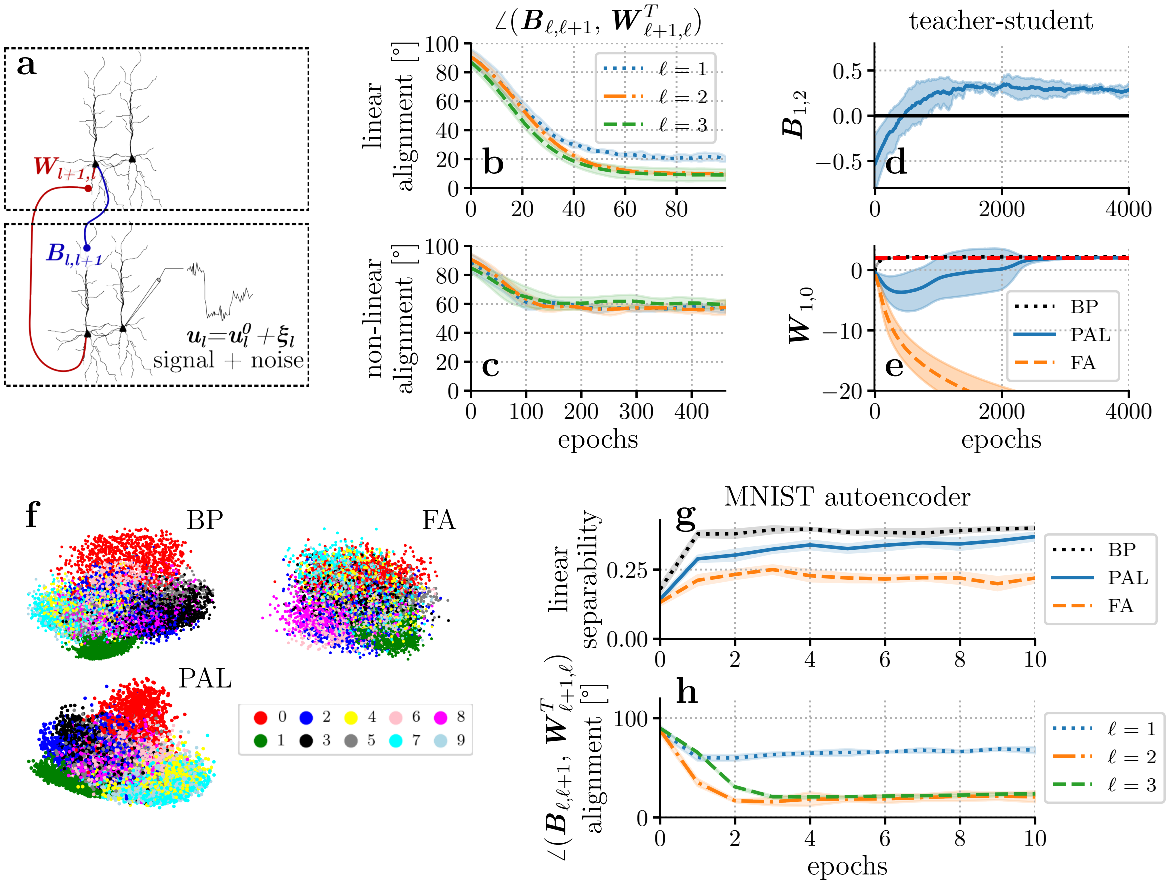

Figure 1: **PAL aligns weight updates with backpropagation in hierarchical cortical networks.** **(a)** Cortical pyramidal cells as functional units of sensory processing and credit assignment. Bottom-up ($\boldsymbol{W}_{\ell+1,\ell}$) and top-down ($\boldsymbol{B}_{\ell,\ell+1}$) projections preferentially target different dendrites. Due to stochastic dynamics of individual neurons, noise is added to the signal. **(b)** We train the backward projections in a deep, dendritic microcircuit network of multi-compartment neurons with layer sizes [5-20-10-20-5] using our method PAL. All backward weights $\boldsymbol{B}_{\ell,\ell+1}$ are learnt simultaneously, while forward weights are fixed. Forward weights are initialised s.t. neurons are activated in their linear regime. **(c)** Same as **b**, but with weights initialised in non-linear regime. **(d)** In a simple teacher-student task with a neuron chain [1-1-1] of dendritic microcircuits, PAL is able to flip the sign of backwards weights, which is crucial for successful reproduction of the teaching signal. **(e)** PAL solves teacher-student task, where feedback alignment fails. The teaching signal (red dashed) requires positive forward weights, whereas all student networks are initialised with negative $\boldsymbol{W}_{1,0}$. Note that PAL only learns the correct forward weights once the backwards weights have flipped sign (at epoch ~ 500). **(f-h)** PAL learns useful latent representations on the MNIST autoencoder task, whereas FA leads to poor feature separation. We train a network [784-200-2-200-784] using leaky-integrator neurons on the MNIST autoencoder task: **(f)** Shown are the activations after training in the two-neuron layer for all samples in the test set; colors encode the corresponding label. BP and PAL show improved feature separation compared to FA. **(g)** Linear separability of latent activation. **(h)** Alignment angle of top-down weights to all layers for networks trained with PAL. PAL is able adapt top-down weights while forward weights are learnt at the same time.

All curves show mean and standard deviation over 5 seeds.