Lagrangian neurodynamics for real-time error-backpropagation across cortical areas

Dominik Dold[a,b], Akos F. Kungl[a,b], João Sacramento[e], Mihai A. Petrovici[a,b]
Kaspar Schindler[d], Jonathan Binas[c], Yoshua Bengio[c], Walter Senn[a]

[a] *Department of Physiology, University of Bern.*
[b] *Kirchhoff-Institute for Physics, Heidelberg University.*
[c] *MILA, University of Montreal.*
[d] *Department of Neurology, Inselspital Bern.*
[e] *Institute of Neuroinformatics, UZH / ETH Zurich.*

The hierarchical structure of the cortex raises the question how plasticity in the brain is able to shape such a structure in the first place. The distant cousins of biological neurons, deep abstract neural networks, are commonly trained with the backpropagation-of-errors algorithm (backprop), which solves the credit assignment problem for deep neural networks and is behind many of the recent achievements of deep learning. Despite its effectiveness in abstract neural networks, it remains unclear whether backprop might represent a viable implementation of cortical plasticity. Here, we present a new theoretical framework that uses a least-action principle to derive a biologically plausible implementation of backprop.

In our model, neuronal dynamics are derived as Euler-Lagrange equations of a scalar function (the Lagrangian). The resulting dynamics can be interpreted as those of multi-compartment neurons with apical and basal dendrites, coupled with a Hodgkin-Huxley-like activation mechanism allowing neurons to phase-advance their somatic input and hence undo temporal delays introduced by somatic and dendritic low-pass filtering. We suggest that a neuron's apical potential encodes a local prediction error arising from the difference between top-down feedback from higher cortical areas and bottom-up predictions represented by activity in its home layer. This computation is enabled by a stereotypical cortical microcircuit, projecting from pyramidal neurons to interneurons back to the pyramidal neurons' apical compartments. When a subset of output neurons is slightly nudged towards a target behavior that cannot be explained away by bottom-up predictions, an error signal is induced that propagates back throughout the network via feedback connections. By defining synaptic dynamics as gradient descent on the Lagrangian, we obtain a biologically plausible plasticity rule that acts on the forward projections of pyramidal neurons in order to reduce this error.

The presented model incorporates several features of biological neurons that cooperate towards approximating a time-continuous version of backprop, where plasticity acts at all times to reduce local prediction errors, thereby minimizing a global output error or cost function. Finally, the model is not only restricted to supervised learning, but can also be applied to unsupervised and reinforcement learning schemes, as demonstrated in simulations.

Lagrangian dynamics

We propose a model based on an energy function composed of layerwise prediction errors[1] and a cost function defined over a subset of neurons that act as output neurons, e.g., neurons in the last layer of a hierarchical network

$$E = \frac{1}{2}\sum_{k}^{N} \underbrace{\|u_k - W_k\bar{r}_{k-1}\|^2}_{\text{prediction error}} + \underbrace{\beta C}_{\text{cost}}, \qquad (1)$$

where u_k are the membrane potentials of the k^{th} layer, W_k weights projecting to neurons in the k^{th} layer and $\bar{r}_{k-1} = \varphi(u_{k-1})$ the steady-state activation function of neurons in the previous layer. β is a scalar weighting of the costs. The cost function is given by the Euclidean norm between observed and target behaviour $C = \frac{1}{2}\|u_N - y_N\|^2$. By applying a change of variables $u = \tilde{u} - \tau\dot{\tilde{u}}$, we can define the Lagrangian L as $L = -E(\tilde{u}, \dot{\tilde{u}}, W)$. We assume that neural dynamics minimizes an energy integral (or "action"), i.e., $\delta\int L\mathrm{d}t = 0$. The equation of motion solving this constraint is given by the Euler Lagrange equations with respect to $\tilde{u}$, i.e., $\frac{\partial L}{\partial \tilde{u}} = \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{\tilde{u}}}$, leading to

$$\tau\dot{u}_k = -u_k + W_k r_{k-1} + e_k\,, \qquad (2)$$

$$r_{k-1} = \bar{r}_{k-1} + \tau\dot{\bar{r}}_{k-1}\,, \quad e_k = \bar{e}_k + \tau\dot{\bar{e}}_k\,, \qquad (3)$$

$$\bar{e}_k = \bar{r}'_k \cdot W_{k+1}^{\mathrm{T}}(u_{k+1} - W_{k+1}\bar{r}_k)\,, \qquad (4)$$

$$\bar{e}_N = \beta(y_N - u_N)\,. \qquad (5)$$

Synaptic dynamics are derived as gradient descent on the energy function, i.e., plasticity reduces prediction errors:

$$\dot{W}_k \propto -\nabla_{W_k}E = (u_k - W_k\bar{r}_{k-1})\bar{r}_{k-1}^{\mathrm{T}}\,. \qquad (6)$$

Biophysical interpretation

The resulting neuron dynamics can be interpreted as containing somatic (u_k), basal ($W_k\bar{r}_{k-1}$) and apical (e_k) compartments. Prediction errors $\bar{e}_k$ are encoded in the apical dendrite and are formed by comparing top-down feedback ($W_{k+1}^{\mathrm{T}}u_{k+1}$) and bottom-up prediction mediated via lateral interneurons ($W_{k+1}^{\mathrm{T}}u_k^I$ with $u_k^I = W_{k+1}\bar{r}_k$), see Fig. 1A. As discussed in [1], the weights of the interneuron circuit must not be identical to the forward weights but can be learned. In this framework, neurons are both carriers of feedforward input as well as error signals. A crucial difference to ordinary rate models is the appearance of "look-ahead" rates $r_k(t) \approx \bar{r}_k(t+\tau)$, undoing temporal delays by low-pass filtering. We identify this as a Hodgkin-Huxley-like activation mechanism, setting $r \approx I_{\mathrm{Na}}$ which can be shown to behave like the look-ahead rate under certain conditions. This allows the neuron to encode, at every time step, the correct error signal with respect to its current state, enabling plasticity to reduce the cost at all times.

Error backpropagation

Synaptic plasticity is driven by the comparison between basal and somatic potentials. By low-pass filtering Eq. (2) and using Eq. (6), we recover the backprop formulas $\dot{W}_k \propto \bar{e}_k\bar{r}_{k-1}^{\mathrm{T}}$ and $\bar{e}_k = \bar{r}'_k \cdot W_{k+1}^{\mathrm{T}}\bar{e}_{k+1}$. To train the network, output neurons are slightly nudged towards their target $y_N(t)$, reducing the cost function. However, this leads to non-zero prediction errors between layers, driving plasticity to reduce these errors to zero again. For small β, it can be shown that this interplay between nudging and reducing layerwise errors can be used to train the network. We demonstrate the learning capabilities of the model for supervised, unsupervised and reinforcement learning examples (see Fig. 1B-D).

Related work

Recently, the possibility of biologically plausible backprop obtained a huge boost with the discovery of feedback alignment [2]. In [1, 3], it was further shown how cortical microcircuits can be used to approximate error backpropagation. Additionally, Equilibrium Propagation [4] introduced a connection between energy-based models and error backpropagation. In [5], a connection between predictive coding and error backpropagation has been established. See also [6] for a recent review of the aforementioned models. The presented model com-

[1] For simplicity, we restrict the description to layered networks, but the model generalizes to arbitrary connectivities.

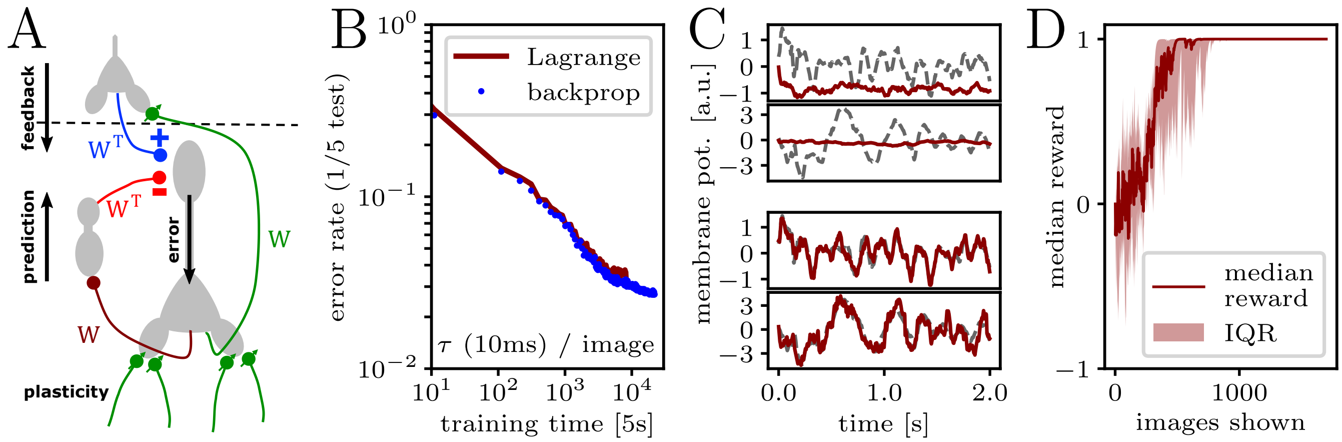

Figure 1: **(A)** Error coding scheme with compartmental model. A bottom-up prediction mediated via interneurons tries to explain away top-down feedback coming from higher cortical areas. If not all feedback can be explained away, this leads to a non-zero prediction error in the apical compartment, driving plasticity in the forward connections. Here, microcircuit, feedback and forward weights are coupled. However, these can also be learned independently, without requiring any weight transport. **(B)** Learning MNIST with a layered network (784-500-10). During training, every training image is only shown for a short amount of time (10ms) such that the network never reaches a stationary state. **(C)** Unsupervised learning of a time-continuous human intracortical EEG signal (56 electrodes, modelled by 56+40 fully recurrent neurons) before and after training. During test runs, the network only sees 46 of 56 inputs and reproduces the remainder (only 2 shown). **(D)** Classification of three images with reinforcement learning (reward is $+1/-1$). A winner-take-all like connectivity among the output neurons provides the necessary nudging when learning is based on scalar reward signals.

bines the previous approaches and extends them to allow real-time learning with backpropagated errors, where plasticity does not depend on a separation of training and free phases or dynamical time scales.

Summary

We present a formalized approach to deriving biologically plausible neurosynaptic dynamics implementing error backpropagation. Different from previous models, the derived dynamics allow a real-time backpropagation of errors, where none of the dynamics have to be stationary in order for plasticity to be able to reduce an output cost. Finally, we would like to stress two key points of the model's biological implementation: **(i)** Learning is driven by a local plasticity rule (*the dendritic prediction of somatic activity*) and **(ii)** by minimizing *local* prediction errors, each neuron contributes to reducing a *global* cost.

Funding

This work has received funding from the European Union under grant agreements 720270, 785907 (HBP) and the Manfred Stärk Foundation. Calculations were performed on UBELIX (University of Bern HPC) and bwHPC (state BaWü HPC, supported by the DFG through grant no INST 39/963-1 FUGG).

References

[1] Sacramento, J., et al. (2018). Dendritic cortical microcircuits approximate the backpropagation algorithm. arXiv preprint arXiv:1810.11393.

[2] Lillicrap, T. P., et al. (2016). Random synaptic feedback weights support error backpropagation for deep learning. Nature communications, 7, 13276.

[3] Guerguiev, J., et al. (2017). Towards deep learning with segregated dendrites. eLife, 6.

[4] Scellier, B., & Bengio, Y. (2017). Equilibrium propagation: Bridging the gap between energy-based models and backpropagation. Frontiers in computational neuroscience, 11, 24.

[5] Whittington, J., & Bogacz, R. (2017). An approximation of the error backpropagation algorithm in a predictive coding network with local Hebbian synaptic plasticity. Neural computation 29.5 : 1229-1262.

[6] Whittington, J., & Bogacz, R. (2019). Theories of Error Back-Propagation in the Brain. Trends in cognitive sciences.