Nonlinear Dendritic Coincidence Detection for Supervised Learning

Fabian Schubert

May 9, 2021

1 Introduction

In recent years, a growing body of research has addressed the functional implications of the distinct physiology and anatomy of cortical pyramidal neurons [27, 12, 22]. In particular, on the theoretical side, we saw a paradigm shift from treating neurons as point-like electrical structures towards embracing the entire dendritic structure [15, 21, 24]. This was mostly due to the fact that experimental work uncovered dynamical properties of these cells that simply could not be accounted for by point models [28, 11].

An important finding was that the apical dendritic tree of cortical pyramidal neurons can act as a separate nonlinear synaptic integration zone [27, 4]. Under certain conditions, a dendritic Ca^{2+} spike can be elicited that propagates towards the soma, causing rapid, bursting spiking activity. One of the cases in which dendritic spiking can occur was termed 'backpropagation-activated Ca^{2+} spike firing' ('BAC firing'): A single somatic spike can backpropagate towards the apical spike initiation zone, in turn significantly facilitating the initiation of a dendritic spike [29, 27, 14]. This reciprocal coupling is believed to act as a form of coincidence detection: If apical and basal synaptic input co-occurs, the neuron can respond with a rapid burst of spiking activity. The firing rate of these temporal bursts exceeds the firing rate that is maximally achievable under basal synaptic input alone, therefore representing a form of temporal coincidence detection between apical and basal input.

Naturally, these mechanisms also affect plasticity and thus learning within the cortex [26, 6]. While the interplay between basal and apical stimulation and its effect on synaptic efficacies is subject to ongoing research, there is some evidence that BAC-firing tends to shift plasticity towards long-term potentiation (LTP) [18]. Thus, coincidence between basal and apical input appears to also gate synaptic plasticity.

In a supervised learning scheme, where the top down input arriving at the apical compartment acts as the teaching signal, the most straight-forward learning rule for the basal synaptic weights would be derived from an appropriate loss function, such as a mean square error, based on the difference between basal and apical input, i.e. $I_\mathrm{p} - I_\mathrm{d}$. Theoretical work has investigated possible learning mechanisms that could utilize an intracellular error signal [30, 23, 7]. However, a clear experimental evidence for a physical quantity encoding such an error is—to our knowledge—yet to be found. On the other hand, Hebbian-type plasticity is extensively documented in experiments [8, 5, 20, 2]. Therefore, our work is based on the question whether the non-linear interactions between basal and apical synaptic input could, when combined with a Hebbian plasticity rule, allow a neuron to learn to reproduce an apical teaching signal in its proximal input.

In our work, we combine a phenomenological model that generates the output firing rate as a function of two streams of synaptic input (subsuming basal and apical input) with Hebbian, as well as BCM-like plasticity rules on basal synapses. We hypothesized that this combination of neural activation and plasticity rules would lead to an increased correlation between basal and apical input.

Furthermore, this temporal alignment could potentially facilitate apical inputs to act as top-down teaching signals, without the need for an explicit error-driven learning rule. Thus, we also test our model in a simple linear supervised classification task and compare it with the performance of a simple point neuron equipped with similar plasticity rules.

2 Model

2.1 Neuron Model

The neuron model used throughout this study is a discrete-time rate encoding model that uses two separate input variables, subsuming the total synaptic input current injected arriving at the basal (proximal) and apical (distal) dendritic structure of a pyramidal neuron, respectively. The model is a slightly simplified version of a phenomenological model proposed by Shai et al. [25]. Denoting the input currents by I_p (proximal) and I_d (distal), the model is written as

$$\begin{aligned} y\left(t\right) = {} & \alpha\sigma\left(I_\mathrm{p}(t) - \theta_{p0}\right)\left[1 - \sigma\left(I_\mathrm{d}(t) - \theta_d\right)\right] \\ & + \sigma\left(I_\mathrm{d}(t) - \theta_d\right)\sigma\left(I_\mathrm{p}(t) - \theta_{p1}\right) \end{aligned} \tag{1}$$

$$\sigma(x) \equiv \frac{1}{1+\exp(-4x)}\,. \tag{2}$$

Here, θ_{p0}, θ_{p1} and θ_d are threshold variables with respect to proximal and distal input. Overall, this equation describes two distinct regions of neural activation in the $(I_\mathrm{p}, I_\mathrm{d})$-space which differ in their maximal firing rates, which are set to 1 and α, where $0 < \alpha < 1$. A plot of (1) is shown in Fig. 1.

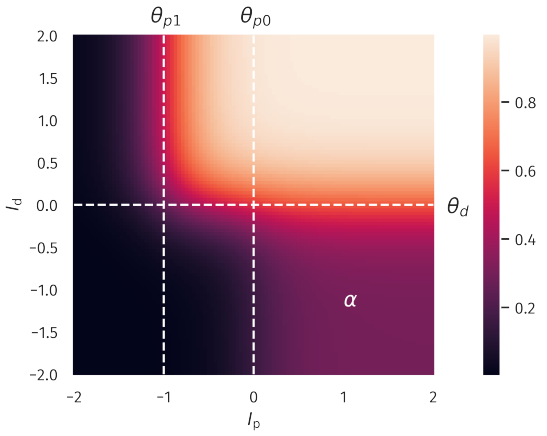

Figure 1: **Two-compartment rate model.** Firing rate as a function of proximal and distal input I_p and I_d, see (1). The thresholds θ_{p0}, θ_{p1} and θ_d define two regions of neural activity with a a maximal firing rate of 1 and α.

In all our numerical experiments, we compare this model with a simple point neuron model that was given by

$$y(t) = \sigma\left(I_\mathrm{p}(t) + I_\mathrm{d}(t) - \theta\right) \ . \tag{3}$$

The apical input I_d was generated 'as is', meaning, it is not dynamically calculated as a superposition of multiple presynaptic inputs, but given by

$$I_\mathrm{d}(t) = n_d(t)x_d(t) - b_d(t) \ , \tag{4}$$

where $n_d(t)$ is a scaling factor, $x_d(t)$ a pre-generated discrete time sequence and $b_d(t)$ a bias. Note that n_d and b_d are time dependent since they are subject to adaptation processes that are described in the next section.

Similarly, $I_\mathrm{p}(t)$ is given by

$$I_\mathrm{p}(t) = n_p(t)\sum_{i=1}^{N} x_{p,i}(t)w_i(t) - b_p(t) \ , \tag{5}$$

where N is the number of presynaptic afferents, $x_{p,i}(t)$ the corresponding sequences and $w_i(t)$ the synaptic efficacies. As for $I_\mathrm{d}(t)$, $n_p(t)$ and $b_p(t)$ is a time dependent scaling and bias.

2.2 Plasticity

We implemented a Hebbian plasticity rule for the basal synaptic weights given by the following update equation:

$$w_i(t+1) = w_i(t) + \mu_w\left(x_{p,i}(t) - \tilde{x}_{p,i}(t)\right)\left(y(t) - \tilde{y}\right) \tag{6}$$
$$\tilde{x}_{p,i}(t+1) = (1-\mu_{av})\tilde{x}_{p,i}(t) + \mu_{av}x_{p,i}(t+1) \tag{7}$$
$$\tilde{y}(t+1) = (1-\mu_{av})\tilde{y}(t) + \mu_{av}y(t+1) \tag{8}$$

Additionally, we use a synaptic normalization constraint

$$w_i(t) \to \frac{w_i(t)}{\|\mathbf{w}(t)\|} \tag{9}$$

in each time step, where $\|\mathbf{w}(t)\|$ denotes the Euclidean norm of the synaptic weight vector.

θ_{p0}	0	V_d^t	0.25
θ_{p1}	-1	μ_b	10^{-3}
θ_d	0	μ_n	10^{-4}
α	0.3	μ_{av}	$5\cdot10^{-3}$
μ_w	$5\cdot10^{-5}$	I_p^t	0
V_p^t	0.25	I_d^t	0

Table 1: Model parameters

For comparative reasons, the point neuron model is equipped with the same plasticity rule for the proximal weights as (6).

Additionally, the scaling and bias variables are changing dynamically according to the following homeostatic plasticity rules:

$$b_p(t+1) = b_p(t) + \mu_b\left[I_\mathrm{p}(t) - I_\mathrm{p}^t\right] \tag{10}$$
$$b_d(t+1) = b_d(t) + \mu_b\left[I_\mathrm{d}(t) - I_\mathrm{d}^t\right] \tag{11}$$
$$n_p(t+1) = n_p(t) + \mu_n\left[V_p^t - \left(I_\mathrm{p}(t) - \tilde{I}_p(t)\right)^2\right] \tag{12}$$
$$n_d(t+1) = n_d(t) + \mu_n\left[V_d^t - \left(I_\mathrm{d}(t) - \tilde{I}_d(t)\right)^2\right] \tag{13}$$
$$\tilde{I}_p(t+1) = (1-\mu_{av})\tilde{I}_p(t) + \mu_{av}I_\mathrm{p}(t+1) \tag{14}$$
$$\tilde{I}_d(t+1) = (1-\mu_{av})\tilde{I}_d(t) + \mu_{av}I_\mathrm{d}(t+1) \tag{15}$$

Here, I_p^t, I_d^t, V_p^t and V_p^t define targets for the temporal means and variances of I_p and I_d. The dynamic variables $\tilde{I}_p$ and $\tilde{I}_d$ are simply low-pass filtered running averages of I_p and I_d.

A list of all parameter values is given in Table 1.

3 Results

3.1 Increased Alignment between Basal and Apical Inputs

As a first test, we wanted to quantify the neuron's ability to align its basal input to the apical teaching signal. To do so, we defined the pearson correlation coefficient $\rho[I_\mathrm{p}, I_\mathrm{d}]$ between the basal and apical input current as our measure of interest. We determined this temporal correlation coefficient after simulating all plasticity mechanisms under stationary random input sequences with certain statistical properties that shall be explained in the following. As a starting point, we chose all $x_{p,i}(t)$ to be randomly drawn from a uniform distribution, where $x_{p,i}(t) \in (0,1)$. For $I_\mathrm{d}(t)$ to be fully 'reconstructable' by the basal input, $x_d(t)$ had to be some linear combination of all $x_{p,i}(t)$. Therefore, we chose $x_d(t) = \sum_{i=1}^{N} a_i x_{p,i}(t)$, where $\mathbf{a}$ is a random vector with unit length. Since we used a Hebbian learning scheme, we expected that the direction and magnitude of the principal components of the basal input would also significantly affect the outcome of the experiment: A large

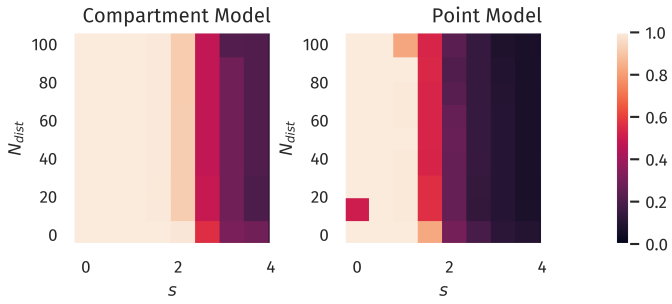

Figure 2: **Alignment between Basal and Apical Input.** Color encodes the Pearson correlation $\rho[I_p, I_d]$ for different number of orthogonal distraction directions N_{dist} and the corresponding scaling faction s.

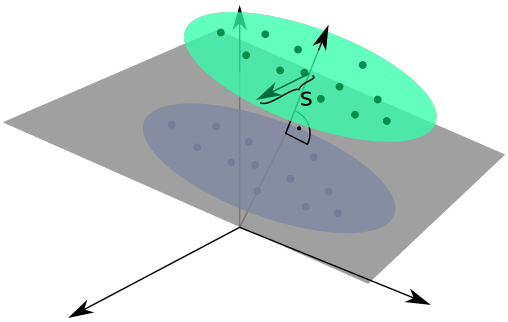

Figure 3: **Input Space for a Simple Linear Classification Task.** Two clusters of presynaptic basal activity were generated from multivariate Gaussian distributions. Here, s denotes the standard deviation orthogonal to the normal vector of the classification hyperplane.

variance in the basal input orthogonal to the 'reconstruction vector' $\mathbf{a}$ should act as a distraction for the plasticity and reduce the resulting alignment between I_p and I_d. Therefore, we applied a transformation to the input sequences $x_{p,i}(t)$ that were parameterized by two quantities, a scaling factor s and the dimension N_{dist} of a subspace of the basal input space. A set of N_{dist} orthonormal basis vectors was randomly generated, which were also orthogonal to $\mathbf{a}$. Withing this N_{dist}-dimensional subspace, the input sequences $x_{p,i}(t)$ were then rescaled by the factor s.

After both $x_{p,i}(t)$ and $x_d(t)$ were generated, a simulation was run using all previously described plasticity mechanisms until the dynamic variables reached a stationary state. After this learning phase, another set of input sequences was generated using the previously described protocol and $\rho[I_\mathrm{p}, I_\mathrm{d}]$ was calculated. Note that plasticity was turned off in this phase. This entire procedure allowed us to calculate $\rho[I_\mathrm{p}, I_\mathrm{d}]$ as a function of the distraction parameters s and N_{dist}. The results for both neuron models is shown in Fig. 2. Here, the total number of basal inputs is $N = 10$. One can observe a decorrelation transition for both models. However, the compartment model supports a significantly stronger distraction in terms of the scaling factor s as compared to the point model. This is a first confirmation of our hypothesis that nonlinear interactions between basal and apical input could improve learning guided by top-down signals.

3.2 Supervised Learning in a Linear Classification Task

Next, we investigated if the observed differences would also improve the performance in an actual supervised learning task. For this purpose, we constructed presynaptic basal input as illustrated in Fig. 3. Written in vector form, each sample from the basal input is generated from the following expression:

$$\mathbf{x}_p(t) = \mathbf{b} + \mathbf{a}\left(c(t) + \sigma_a \zeta_a(t)\right) + s \cdot \sum_{i=1}^{N_{dist}} \zeta_{dist,i}(t)\mathbf{v}_{dist,i} \ . \tag{16}$$

Here, $\mathbf{b}$ is a random vector drawn uniformly from $(0,1)^N$, $\mathbf{a}$ is random unit vector as introduced in Section 3.1, $c(t)$ is a binary variable drawn from $\{-0.5, 0.5\}$ with equal probability and $\zeta_a(t)$ and the $\zeta_{dist,i}(t)$ are independent random Gaussian variables with zero mean and unit variance. Hence, σ_a simply denotes the standard deviation of each Gaussian cluster along the direction of the normal vector $\mathbf{a}$ and was set to $\sigma_a = 0.25$. Finally, the set of $\mathbf{v}_{dist,i}$ forms a randomly generated orthogonal basis of N_{dist} unit vectors which—just as in Section 3.1—are also orthogonal to $\mathbf{a}$. The free parameter s parameterized the standard deviation along this subspace orthogonal to $\mathbf{a}$. As indicated by the time dependence, the Gaussian and binary random variables are drawn for each time step. The vectors $\mathbf{b}$, $\mathbf{a}$, and $\mathbf{v}_{dist,i}$ are generated once before the beginning of a simulation run.

For the classification task, we set up two output neurons receiving the same basal presynaptic input, while the top-down input encoded the correct linear classification in a one-hot scheme, that is

$$x_{d,0}(t) = 1 - \Theta\left(\left(\mathbf{x}_p(t) - \mathbf{b}\right)^T \mathbf{a}\right) \tag{17}$$

$$x_{d,1}(t) = \Theta\left(\left(\mathbf{x}_p(t) - \mathbf{b}\right)^T \mathbf{a}\right) \ , \tag{18}$$

where $\Theta(x)$ is the Heaviside step function.

As in the previous experiment, we ran a full simulation until all dynamic variables reached a stationary state. After this, a test run without plasticity and with the apical input turned off was used to evaluate the classification performance. For each sample, the index of the neuron with the highest activity was used as the predicted class. Accuracy was then calculated as the fraction of correctly classified samples.

The resulting accuracy as a function of N_{dist} and s is shown in Fig. 4. Albeit differences are small, the compartment model does show a better overall accuracy for the tested parameter range. It should be noted, though, that the advantage of the compartment model becomes even less prominent when looking at the actual correlation between proximal and distal input as a measure of successful

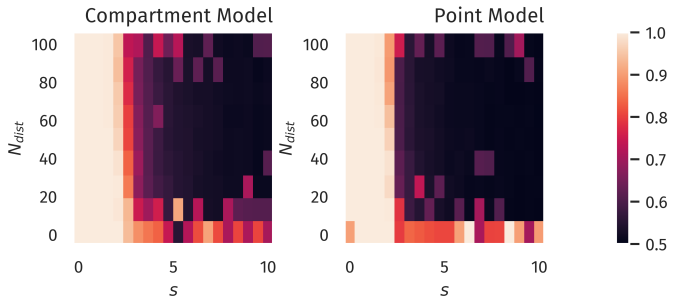

Figure 4: **Binary Classification Accuracy.** Fraction of correctly classified patterns as illustrated in Fig. 3, see Section 3.2.¸

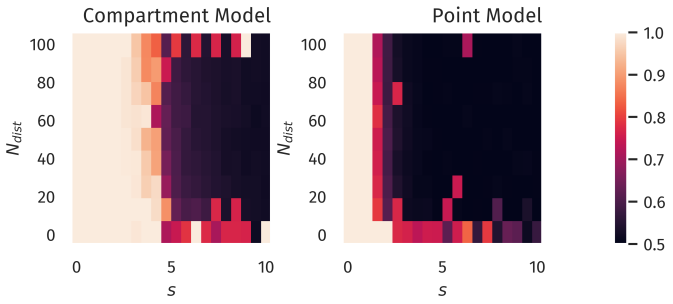

Figure 5: **Binary Classification Accuracy, BCM Rule.** Fraction of correctly classified patterns as illustrated in Fig. 3, after training with a BCM-like learning rule.¸

learning (as done in the previous section). Fig. 8 in the supplementary section shows that the compartment model allows for a marginally larger scaling among the distraction dimensions.

3.3 Non-Hebbian Learning Rules

Instead of Hebbian learning, we also considered a BCM-like learning rule for the basal weights [3, 13]. The form of the BCM-rule we consider here reads

$$\Delta w_i \propto y\,(y - \theta_M)\,x_i - \epsilon w_i\ , \tag{19}$$

where θ_M is a threshold defining a transition from LTP to LTD and ϵ is an optional decay term on the weights. In the variant introduced by Law and Cooper [16], the sliding threshold is simply the temporal average of the squared neural activity, $\theta_M = \langle y^2 \rangle$. In practice, this would be calculated as a running average, thereby preventing the weights from growing indefinitely.

However, for our compartment model, we chose to explicitly set the threshold to be the mean value between the high- and low-activity regime in our compartment model, i.e. $\theta_M = (1 + \alpha)/2$. By doing so, LTP is preferably induced if both basal and apical input are present at the same time. Furthermore, instead of the weight decay term, we chose to keep the weight normalization as introduced in (9). Obviously, for the point model, the reasoning behind our choice of θ_M did not apply. Still, to provide some level of comparability, we also ran simulations with a point model where the sliding threshold was calculated as a running average of y^2. Furthermore, we did not use weight normalization in this case, but chose to use a small weight decay term with $\epsilon = 0.1$. The results are shown in Fig. 5 (classification task) and Fig. 5 (Basal-Apical alignment). While the accuracy of the classification for the point model is at most comparable to Hebbian learning, the BCM-like rule for the compartment model significantly increases the accuracy for the tested parameter range (compare Fig. 4). Still, as for the Hebbian learning rule, this result should be taken with a grain of salt, as Fig. 9 in the supplementary section indicates that a rather large part of the region ($s > 5$) showing good classification performance for the compartment model in Fig. 5 does hardly exhibit any

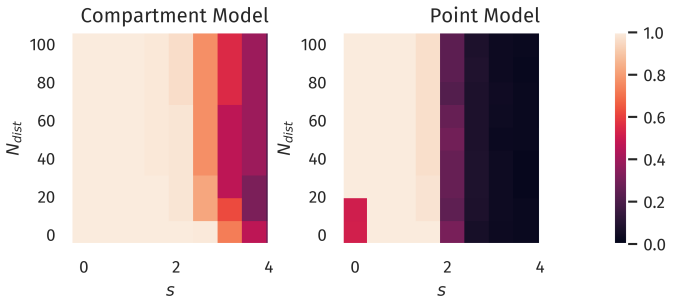

Figure 6: **Alignment between Basal and Apical Input, BCM Rule.** Color encodes the Pearson correlation $\rho[I_p, I_d]$ for different number of orthogonal distraction directions N_{dist} and the corresponding scaling faction s after training with a BCM-like rule. Compare to Fig. 2.

actual correlation between proximal and distal input, making it likely that the accuracy of the classification could be easily impaired in this regime, e.g. by injecting small amounts of noise.

The compartment model also significantly benefits from the BCM rule in terms of basal-apical alignment as tested in Sect. 3.1, while only marginal improvements can be observed for the point model (compare Fig. 6 with Fig. 2).

3.4 Objective Function of BCM Learning in the Compartment Model

To form a better understanding of why the BCM-type learning rule in combination with the implemented compartment model, we can formalize the learning rule for the proximal weights in terms of an objective function. For this purpose, we further simplify (1) by replacing the sigmoid functions $\sigma(x)$ by a simple step function $\Theta(x)$. This does not change the overall shape or topology of the activation in the (I_p, I_d) space but merely makes the smooth transitions sharp and instantaneous. Using $\Delta w_i \propto y\,(y - \theta_M)\,x_i$, we find in this case

$$\begin{aligned}\Delta w_i \propto\ &[(1-\alpha)\Theta(I_d - \theta_d)\Theta(p - \theta_{p1}) \\ &+\alpha(\alpha - 1)\Theta(\theta_d - I_d)\Theta(p - \theta_{p0})]x_i\ .\end{aligned} \tag{20}$$

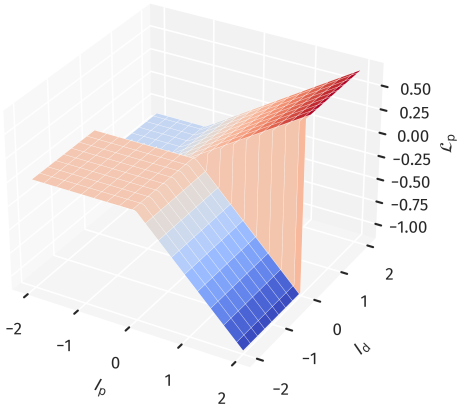

Figure 7: **Objective Function for the Proximal Weight Update.** The approximate objective function for the proximal weights as given in (22). This corresponds to a combination of using (1) together with (19). Note the ridge-like structure along the I_p-I_d diagonal, which supports the alignment between proximal and distal input.

Noting that $\Theta(x)$ is the first derivative of the ReLu function $[x]^+ \equiv \max(0, x)$, we find that this update rule can be written as

$$\Delta w_i \propto \frac{\partial \mathcal{L}_p}{\partial w_i} \tag{21}$$

$$\begin{aligned}\mathcal{L}_p \equiv\ & (1-\alpha)\Theta(I_d - \theta_d)[p - \theta_{p1}]^+ \\ & + \alpha(\alpha - 1)\Theta(\theta_d - I_d)[p - \theta_{p0}]^+ \ .\end{aligned} \tag{22}$$

The objective function is shown in Fig. 7. One can observe that states closer to the I_p-I_d diagonal are preferred, while the opposite is the case for off-diagonal states. This provides an explanation why the BCM-rule can induce an alignment between proximal and distal inputs when used in combination with the nonlinear compartment model.

It should be noted, though, that to a certain extent, the objective function is not scale-invariant (as would be e.g. if the squared error was used) in the sense that the prior distributions of both proximal and distal inputs need a certain mean and variance to cover a region of input states for which the described effects can take place. As a counterexample, one could imagine that the input samples only covered a flat of $\mathcal{L}_p$, as for example in Fig. 7 on the left, leading to a zero average gradient. This is prevented, however, by the homeostatic processes acting simultaneously on the gains and biases.

4 Discussion

We have demonstrated that in a simple supervised learning scheme, the proposed two-compartment transfer function significantly increases the robustness of the learning process against distracting components in the proximal input space. This was most prominently found when combined with a BCM-like learning rule.

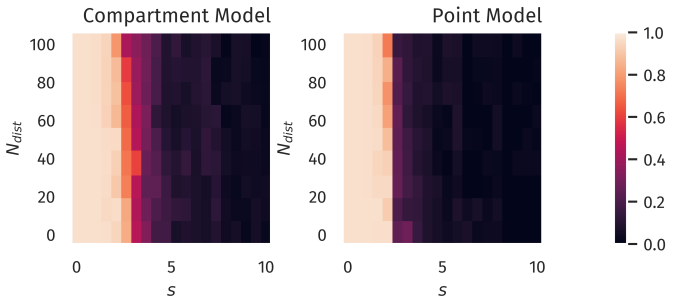

Figure 8: **Alignment between Basal and Apical Input after Binary Classification Learning.** Correlation between proximal and distal Input after training as described in Sect. 3.2.

The idea of target-backpropagation in multi-layered networks has already been proposed in different variants [1, 17, 7]. Yet, all of these approaches assume a learning rule being dependent on an explicit error term between top-down and bottom up signals guiding plasticity in some form or another. In this work, we considered an alternative approach, that is, the correlation between proximal and distal input as a maximization objective, in combination with homeostatic adaptation rules that regulate proximal and distal inputs into a desired "working regime", see Sect. 3.4. Since I_p is a linear projection of the proximal input space, maximizing the correlation between I_p and I_d can be regarded as a form of canonical correlation analysis (CCA) [10]. The idea of investigating CCA as a possible mode of synaptic learning has previously been investigated by Haga and Fukai [9]. Interestingly, according to the authors, a BCM-learning term in the plasticity dynamics accounts for principal component analysis in the input space, while CCA requires an additional multiplicative term between local basal and apical activity. In contrast, our results indicate that such a multiplicative term is not required to drive basal synaptic plasticity towards a maximal alignment between basal and apical input, even in the presence of distracting principal components.

As we did not include higher-dimensional distal input patterns, it remains an open question how target signals would be formed in multi-layered network structures. However, as previous works have indicated, random top-down weights can be sufficient for successful credit assignment and learning [19, 7]. Therefore, our results could potentially also be transferred to deeper network structures, where plasticity is guided by local errors between top-down and bottom-up signals.

5 Supplementary Material

References

[1] Yoshua Bengio. How Auto-Encoders Could Provide Credit Assignment in Deep Networks via Target Propagation. jul 2014.

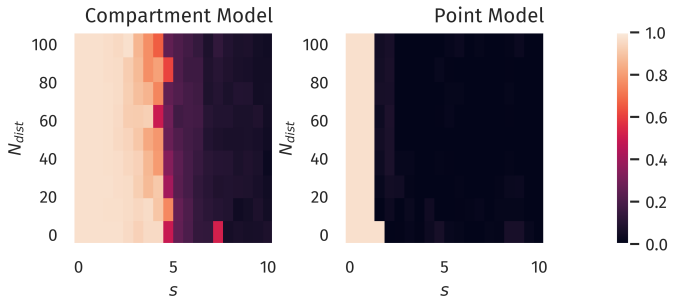

Figure 9: **Alignment between Basal and Apical Input after Binary Classification Learning using a BCM-like Rule.** Correlation between proximal and distal Input after linear classification training as described in Sect. 3.2, but using the plasticity rules given in Sect. 3.3.

[2] Guo Qiang Bi and Mu Ming Poo. Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of Neuroscience*, 18 (24):10464–10472, dec 1998. ISSN 02706474. doi: 10.1523/jneurosci.18-24-10464.1998.

[3] E. L. Bienenstock, L. N. Cooper, and P. W. Munro. Theory for the development of neuron selectivity: Orientation specificity and binocular interaction in visual cortex. *Journal of Neuroscience*, 2(1):32–48, jan 1982. ISSN 02706474. doi: 10.1523/jneurosci.02-01-00032.1982.

[4] Tiago Branco and Michael Häusser. Synaptic Integration Gradients in Single Cortical Pyramidal Cell Dendrites. *Neuron*, 69(5):885–892, mar 2011. ISSN 08966273. doi: 10.1016/j.neuron.2011.02.006.

[5] Dominique Debanne, Beat H. Gähwiler, and Scott M. Thompson. Asynchronous pre- and postsynaptic activity induces associative long-term depression in area CA1 of the rat hippocampus in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, 91(3):1148–1152, feb 1994. ISSN 00278424. doi: 10.1073/pnas.91.3.1148.

[6] Christian Ebner, Claudia Clopath, Peter Jedlicka, and Hermann Cuntz. Unifying Long-Term Plasticity Rules for Excitatory Synapses by Modeling Dendrites of Cortical Pyramidal Neurons. *Cell Reports*, 29(13):4295–4307.e6, dec 2019. ISSN 22111247. doi: 10.1016/j.celrep.2019.11.068.

[7] Jordan Guerguiev, Timothy P. Lillicrap, and Blake A. Richards. Towards deep learning with segregated dendrites. *eLife*, 6, dec 2017. ISSN 2050084X. doi: 10.7554/eLife.22901.

[8] B. Gustafsson, H. Wigstrom, W. C. Abraham, and Y. Y. Huang. Long-term potentiation in the hippocampus using depolarizing current pulses as the conditioning stimulus to single volley synaptic potentials. *Journal of Neuroscience*, 7(3):774–780, 1987. ISSN 02706474. doi: 10.1523/jneurosci.07-03-00774.1987.

[9] Tatsuya Haga and Tomoki Fukai. Dendritic processing of spontaneous neuronal sequences for single-trial learning. *Scientific Reports*, 8(1):15166, dec 2018. ISSN 20452322. doi: 10.1038/s41598-018-33513-9.

[10] Wolfgang Härdle and Leopold Simar. Canonical Correlation Analysis. In *Applied Multivariate Statistical Analysis*, pages 321–330. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007. doi: 10.1007/978-3-540-72244-1_14.

[11] M. Häusser, N. Spruston, and G. J. Stuart. Diversity and dynamics of dendritic signaling, oct 2000. ISSN 00368075.

[12] Etay Hay, Sean Hill, Felix Schürmann, Henry Markram, and Idan Segev. Models of Neocortical Layer 5b Pyramidal Cells Capturing a Wide Range of Dendritic and Perisomatic Active Properties. *PLoS Computational Biology*, 7(7):e1002107, jul 2011. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1002107.

[13] Nathan Intrator and Leon N. Cooper. Objective function formulation of the BCM theory of visual cortical plasticity: Statistical connections, stability conditions. *Neural Networks*, 5(1):3–17, jan 1992. ISSN 08936080. doi: 10.1016/S0893-6080(05)80003-6.

[14] Matthew Larkum. A cellular mechanism for cortical associations: An organizing principle for the cerebral cortex, mar 2013. ISSN 01662236.

[15] Matthew E. Larkum, Thomas Nevian, Maya Sandier, Alon Polsky, and Jackie Schiller. Synaptic integration in tuft dendrites of layer 5 pyramidal neurons: A new unifying principle. *Science*, 325(5941):756–760, aug 2009. ISSN 00368075. doi: 10.1126/science.1171958.

[16] C. Charles Law and Leon N. Cooper. Formation of receptive fields in realistic visual environments according to the Bienenstock, Cooper, and Munro (BCM) theory. *Proceedings of the National Academy of Sciences of the United States of America*, 91 (16):7797–7801, aug 1994. ISSN 00278424. doi: 10.1073/pnas.91.16.7797.

[17] Dong Hyun Lee, Saizheng Zhang, Asja Fischer, and Yoshua Bengio. Difference target propagation. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 9284, pages 498–515. Springer Verlag, 2015. ISBN 9783319235271. doi: 10.1007/978-3-319-23528-8_31.

[18] Johannes J Letzkus, Björn M Kampa, and Greg J Stuart. Learning Rules for Spike Timing-Dependent Plasticity Depend on Dendritic Synapse Location. *Journal of Neuroscience*, 26(41):10420–10429, 2006. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.2650-06.2006.

[19] Timothy P. Lillicrap, Daniel Cownden, Douglas B. Tweed, and Colin J. Akerman. Random synaptic feedback weights support error backpropagation for deep learning. *Nature Communications*, 7(1):1–10, nov 2016. ISSN 20411723. doi: 10.1038/ncomms13276.

[20] Henry Markram, Joachim Lübke, Michael Frotscher, and Bert Sakmann. Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science*, 275(5297):213–215, jan 1997. ISSN 00368075. doi: 10.1126/science.275.5297.213.

[21] Panayiota Poirazi. Information processing in single cells and small networks: Insights from compartmental models. In *AIP Conference Proceedings*, volume 1108, pages 158–167. American Institute of Physics, 2009. doi: 10.1063/1.3117124.

[22] Srikanth Ramaswamy and Henry Markram. Anatomy and physiology of the thick-tufted layer 5 pyramidal neuron, jun 2015. ISSN 16625102.

[23] Mathieu Schiess, Robert Urbanczik, and Walter Senn. Somato-dendritic Synaptic Plasticity and Error-backpropagation in Active Dendrites. *PLoS Computational Biology*, 12(2):1004638, feb 2016. ISSN 15537358. doi: 10.1371/journal.pcbi.1004638.

[24] A S Shai, C A Anastassiou, M E Larkum, and C Koch. Physiology of Layer 5 Pyramidal Neurons in Mouse Primary Visual Cortex: Coincidence Detection through Bursting. *PLOS Computational Biology*, 11(3), 2015.

[25] A S Shai, C A Anastassiou, M E Larkum, and C Koch. Physiology of Layer 5 Pyramidal Neurons in Mouse Primary Visual Cortex: Coincidence Detection through Bursting. *PLOS Computational Biology*, 11(3), 2015.

[26] P J Sjöström and M Häusser. A Cooperative Switch Determines the Sign of Synaptic Plasticity in Distal Dendrites of Neocortical Pyramidal Neurons. *Neuron*, 51(2):227–238, jul 2006. doi: 10.1016/j.neuron.2006.06.017.

[27] N Spruston. Pyramidal neurons: dendritic structure and synaptic integration. *Nature Reviews Neuroscience*, 9(3):206–221, mar 2008. doi: 10.1038/nrn2286.

[28] Nelson Spruston, Yitzhak Schiller, Greg Stuart, and Bert Sakmann. Activity-dependent action potential invasion and calcium influx into hippocampal CA1 dendrites. *Science*, 268(5208):297–300, apr 1995. ISSN 00368075. doi: 10.1126/science.7716524.

[29] Greg J. Stuart and Michael Häusser. Dendritic coincidence detection of EPSPs and action potentials. *Nature Neuroscience*, 4(1):63–71, 2001. ISSN 10976256. doi: 10.1038/82910.

[30] Robert Urbanczik and Walter Senn. Learning by the Dendritic Prediction of Somatic Spiking. *Neuron*, 81(3):521–528, feb 2014. ISSN 08966273. doi: 10.1016/j.neuron.2013.11.030.