NATURAL-GRADIENT LEARNING FOR SPIKING NEURONS

NICE 2021
ELENA KREUTZER
COMPUTATIONAL NEUROSCIENCE GROUP
UNIVERSITY OF BERN
MARCH 18, 2021
What is synaptic strength?

Many equivalent ways to describe the strength of a synapse.
What really matters is the neuron's firing wrt. synaptic input.
Supervised learning

\[ V = \sum_{i}^{n} w^s_i x^\epsilon_i \]  
\[ x^\epsilon_i(t) = [x_i * \epsilon](t) \]

somatic membrane potential

low pass filtered input spike trains
Euclidean-gradient-based-learning depends on parametrization

\[ \Delta w^s = -\frac{\partial C^s}{\partial w^s} \]

dendritic EPSP amplitude

somatic EPSP amplitude
Euclidean-gradient-based-learning depends on parametrization

$$\Delta w^d = -f'(w^d) \frac{\partial C^s}{\partial w^s}$$

**dendritic EPSP amplitude**

**somatic EPSP amplitude**

$$C^d[w^d] = C^s[f(w^d)]$$
Euclidean-gradient-based-learning depends on parametrization

\[ \Delta w^d = -f'(w^d) \frac{\partial C^s}{\partial w^s} \]

\[ \tilde{\Delta} w^s = -f'(w^d)^2 \frac{\partial C^s}{\partial w^s} \]
Euclidean-gradient-based-learning depends on parametrization.

\[ \Delta w^d = -f'(w^d) \frac{\partial C^s}{\partial w^s} \]

\[ \tilde{\Delta} w^s = -f'(w^d)^2 \frac{\partial C^s}{\partial w^s} \neq \frac{\partial C^s}{\partial w^s} \]

Different parametrizations lead to different predictions → inconsistent and inefficient!
Suboptimal choice of parameterization leads to inefficient Euclidean-gradient-based-learning
Natural gradient descent follows steepest descent direction of cost function on manifold of output distributions.
Learning rule

Euclidean gradient descent

\[ \dot{w} = -\eta [Y^* - \phi(V)] \frac{\phi'(V)}{\phi(V)} x^c \]
Learning rule

Euclidean gradient descent
\[
\dot{\mathbf{w}} = -\eta [Y^* - \phi(V)] \frac{\phi'(V)}{\phi(V)} \mathbf{x}^e
\]

Natural gradient descent
\[
\dot{\mathbf{w}} = \eta \gamma_s [Y^* - \phi(V)] \frac{\phi'(V)}{\phi(V)} \frac{1}{f'(\mathbf{w})} \left[ c_e \frac{\mathbf{x}^e}{r} - \gamma_u + \gamma_w f(\mathbf{w}) \right]
\]
Learning rule

Euclidean gradient descent
\[ \dot{w} = -\eta \left[ Y^* - \phi(V) \right] \frac{\phi'(V)}{\phi(V)} x^e \]

Natural gradient descent
\[ \dot{w} = \eta \gamma_s \left[ Y^* - \phi(V) \right] \frac{\phi'(V)}{\phi(V)} \frac{1}{f'(w)} \left[ c_e \frac{\dot{x}^e}{r} - \gamma_u + \gamma_w f(w) \right] \]

- Keeps error term and homosynaptic term of EGD-learning.
- Introduces global and synapse-specific learning rate scaling and heterosynaptic plasticity.
- Global terms can in many cases be locally approximated.
Performance
Synaptic democracy
Interplay of homo- and heterosynaptic plasticity
Conclusion

- Natural gradient yields a parametrization-independent plasticity rule.

- Learning with the Natural gradient rule is faster than with the standard Euclidean gradient descent rule.

- The natural gradient learning rule predicts the existence of "synaptic democracy" and heterosynaptic plasticity.
Acknowledgements

Walter Senn
Mihai Petrovici

Computational Neuroscience and
Neuroscience TMA Groups
University of Bern
Parametrizations

output $p(x, y)$

synaptic weight $w$, $\tilde{w}$
What is synaptic strength?

dendritic EPSP amplitude

somatic EPSP amplitude
What is synaptic strength?

\[ w^s = f(w^d) \]
Approximation

\[ \dot{w}_a = \eta \gamma_s \left[ Y^* - \phi(V) \right] \frac{\phi'(V)}{\phi(V)} f'(w)^{-1} \left[ \frac{c_c x^c}{r} - c_c c_u + c_w V f(w) \right] \]
Backup
Different types of plasticity

\[ \Delta w = \text{error} \cdot (\Delta w_{\text{hom}} + \Delta w_{\text{het}u} + \Delta w_{\text{het}w}) \]

A. Homosynaptic plasticity
   \[ \Delta w_{\text{hom}} \propto \frac{x^c}{\sigma_{\text{CSP}}} \]

B. Uniform heterosynaptic plasticity
   \[ \Delta w_{\text{het}u} \propto -\gamma_w \]

C. Weight proportional heterosynaptic plasticity
   \[ \Delta w_{\text{het}w} \propto \gamma_w f(w) \propto V f(w) \]