NATURAL-GRADIENT LEARNING FOR SPIKING NEURONS

NICE 2021

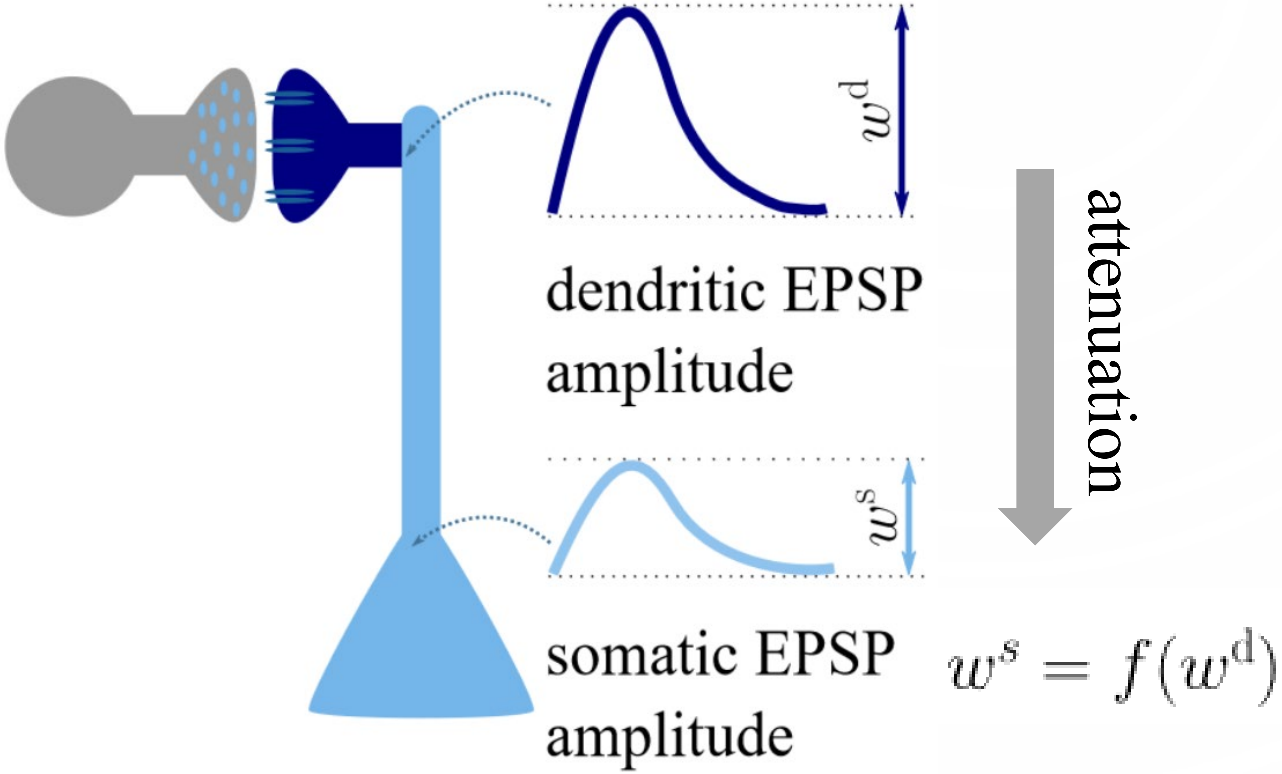
ELENA KREUTZER

COMPUTATIONAL NEUROSCIENCE GROUP

UNIVERSITY OF BERN

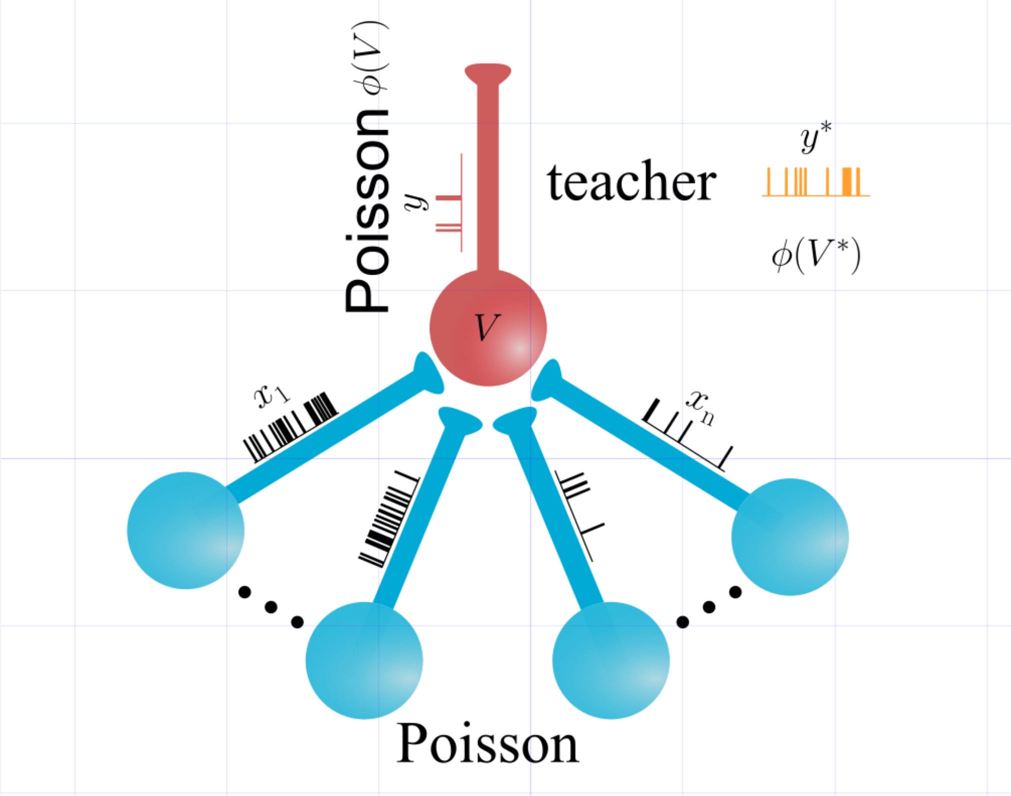
MARCH 18, 2021

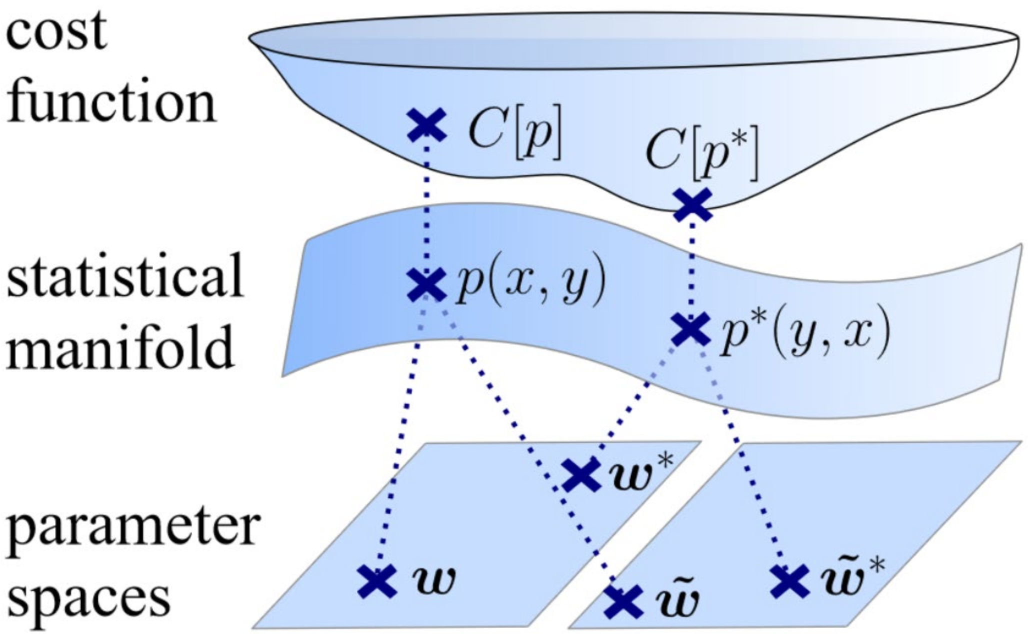
What is synaptic strength?



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

Supervised learning

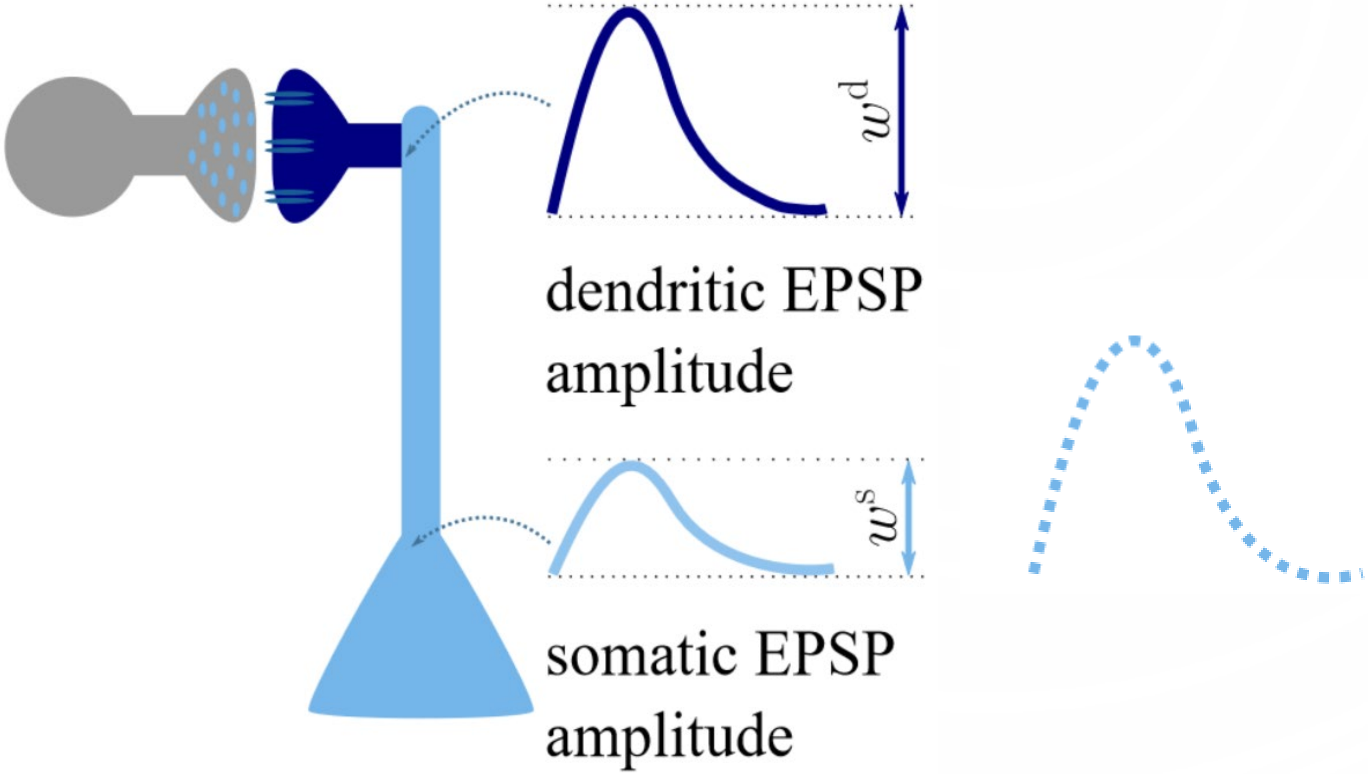




$V = \sum_i^n w_i^s x_i^\epsilon$ somatic membrane potential

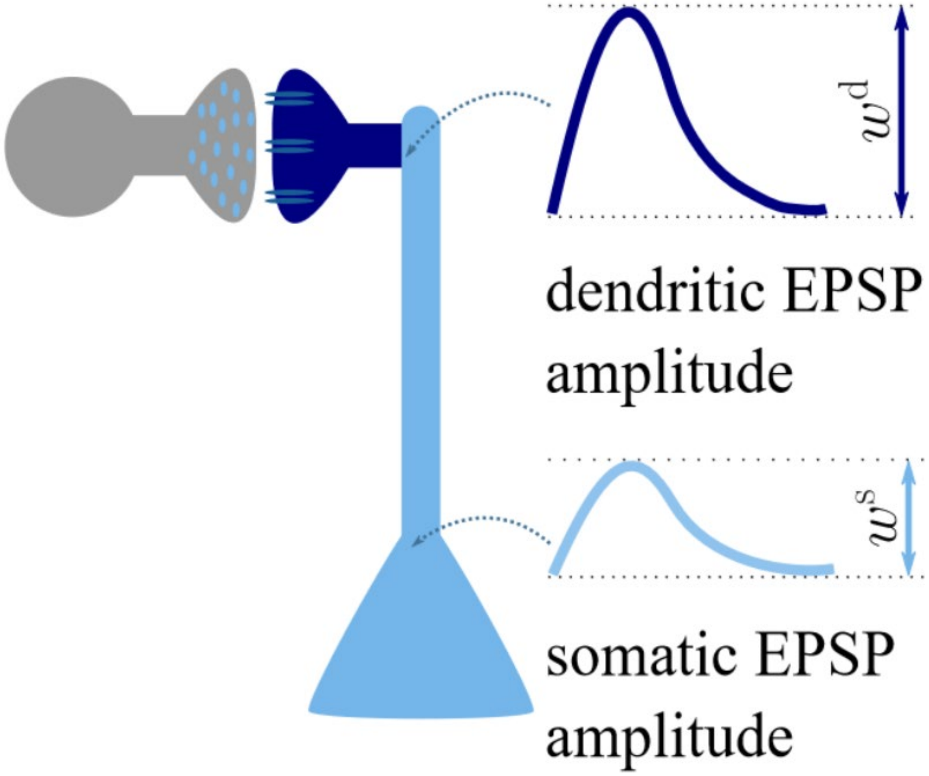
$x_i^\epsilon(t) = [x_i * \epsilon](t)$ low pass filtered input spike trains

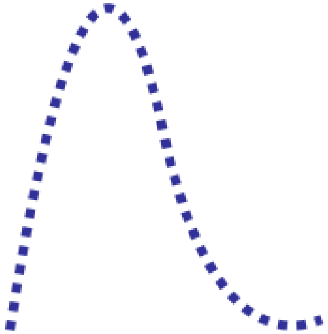
Euclidean-gradient-based-learning depends on parametrization



$$\Delta w^{\mathrm{s}} = -\frac{\partial C^{\mathrm{s}}}{\partial w^{\mathrm{s}}}$$

Euclidean-gradient-based-learning depends on parametrization

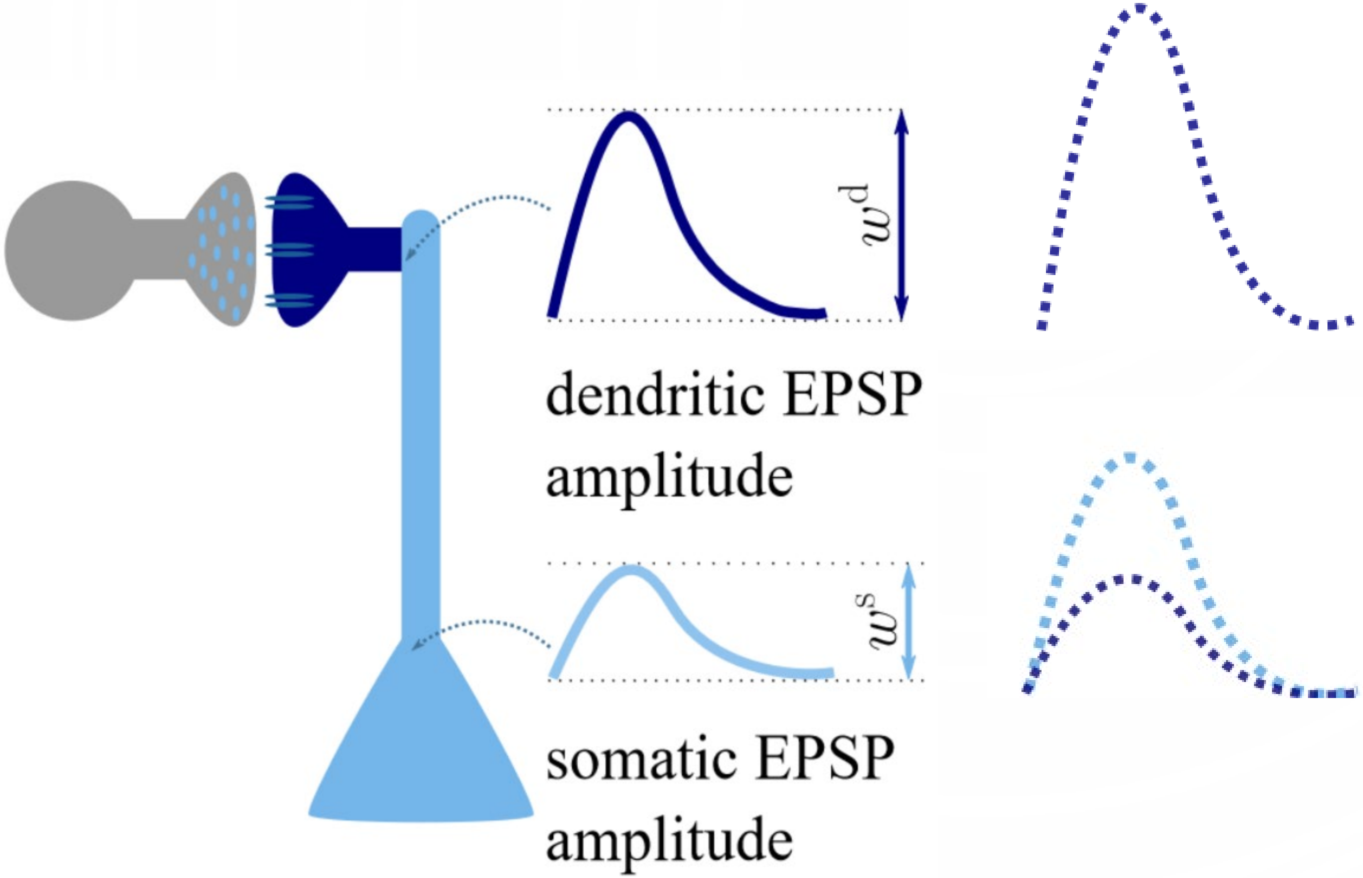




$$\Delta w^{\mathrm{d}} = -f'(w^{\mathrm{d}}) \frac{\partial C^{\mathrm{s}}}{\partial w^{\mathrm{s}}}$$

$$C^{\mathrm{d}}\left[w^{\mathrm{d}}\right] = C^{\mathrm{s}}\left[f\left(w^{\mathrm{d}}\right)\right]$$

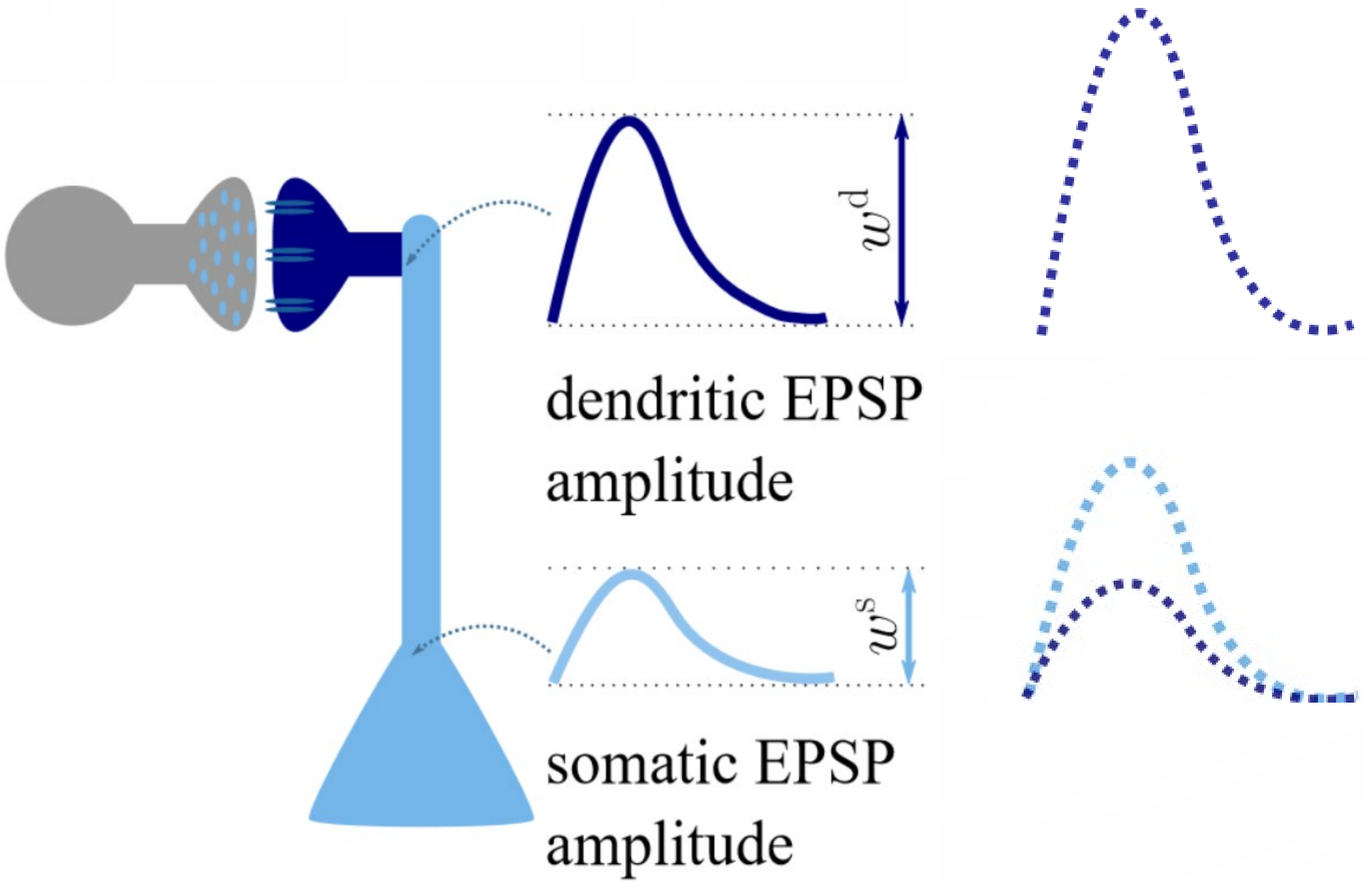
Euclidean-gradient-based-learning depends on parametrization



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

$$\tilde{\Delta} w^{\mathrm{s}} = -f'(w^{\mathrm{d}})^2 \frac{\partial C^{\mathrm{s}}}{\partial w^{\mathrm{s}}}$$

Euclidean-gradient-based-learning depends on parametrization



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

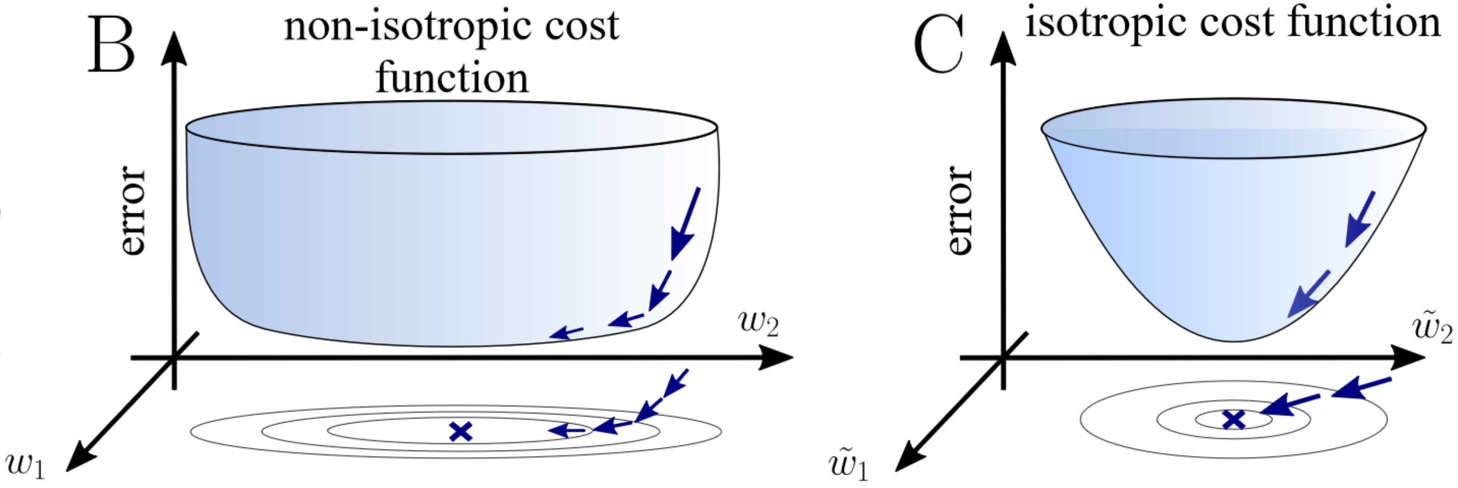
$$\tilde{\Delta} w^{\mathrm{s}} = -f'(w^{\mathrm{d}})^2 \frac{\partial C^{\mathrm{s}}}{\partial w^{\mathrm{s}}}$$

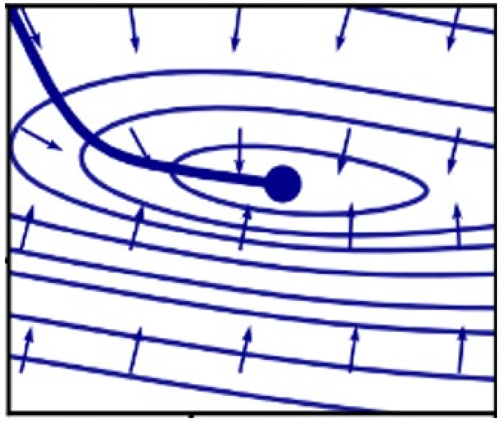
$$\neq$$

$$\Delta w^{\mathrm{s}} = -\frac{\partial C^{\mathrm{s}}}{\partial w^{\mathrm{s}}}$$

Different parametrizations lead to different predictions→ inconsistent and inefficient!

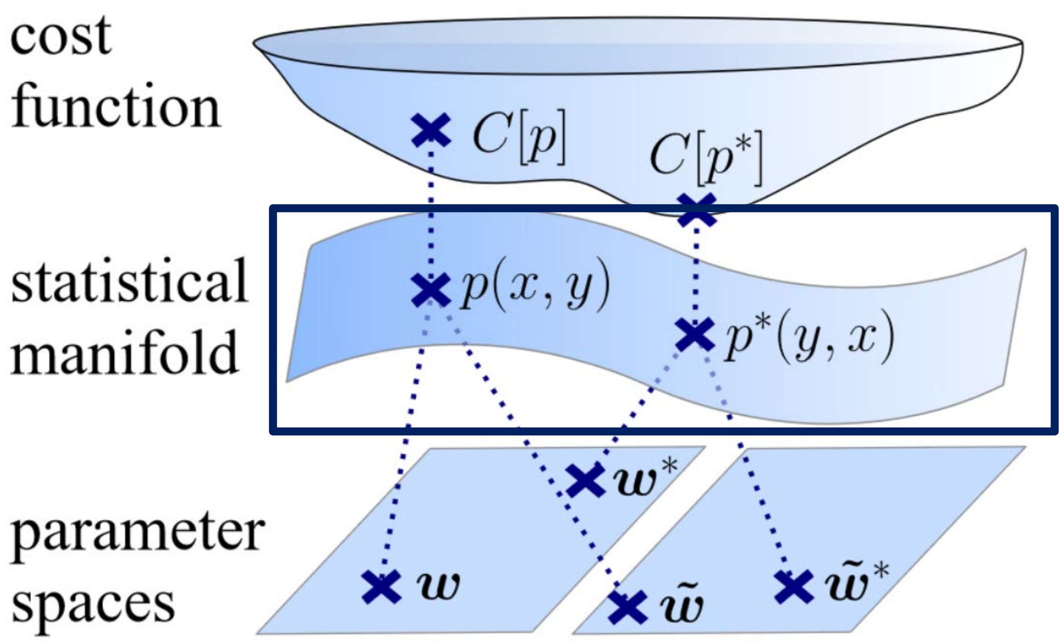
Suboptimal choice of paramterization leads to inefficient Euclidean-gradient-based-learning





Natural-gradient learning

Natural gradient descent follows steepest descent direction of cost function on manifold of output distributions.



Learning rule

Euclidean gradient descent

$$\dot{\boldsymbol{w}} = -\eta \left[Y^* - \phi(V) \right] \frac{\phi'(V)}{\phi(V)} \boldsymbol{x}^{\epsilon}$$

Learning rule

Euclidean gradient descent

$$\dot{\boldsymbol{w}} = -\eta \left[Y^* - \phi(V) \right] \frac{\phi'(V)}{\phi(V)} \boldsymbol{x}^\epsilon$$

Natural gradient descent

$$\dot{\boldsymbol{w}} = \eta \, \gamma_{\mathrm{s}} \left[Y^* - \phi(V) \right] \frac{\phi'(V)}{\phi(V)} \frac{1}{f'(\boldsymbol{w})} \left[c_\epsilon \frac{\boldsymbol{x}^\epsilon}{\boldsymbol{r}} - \gamma_{\mathrm{u}} + \gamma_{\mathrm{w}} f(\boldsymbol{w}) \right]$$

Learning rule

Euclidean gradient descent

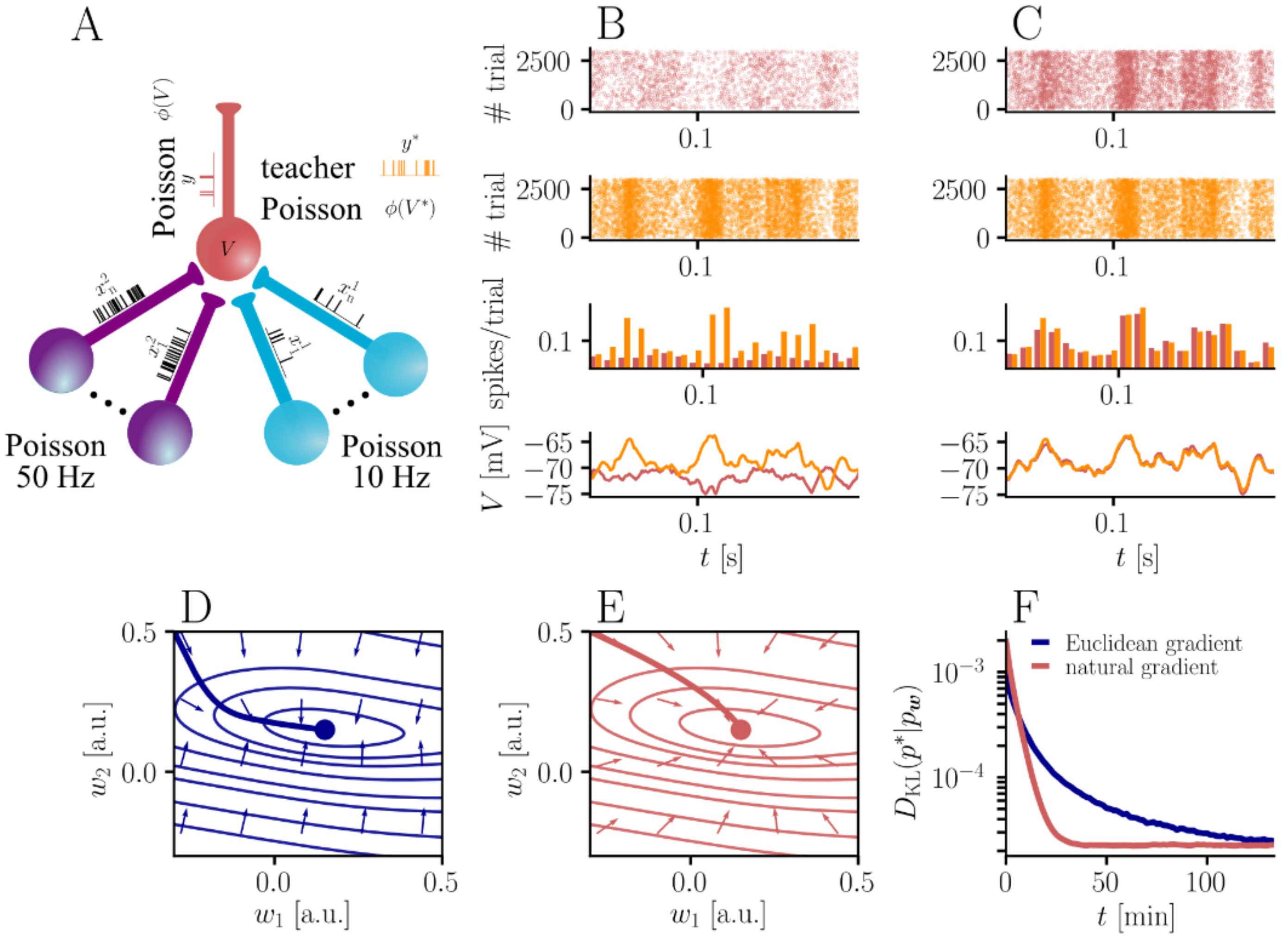
$$\dot{\boldsymbol{w}} = -\eta \left[Y^* - \phi(V) \right] \frac{\phi'(V)}{\phi(V)} \boldsymbol{x}^\epsilon$$

Natural gradient descent

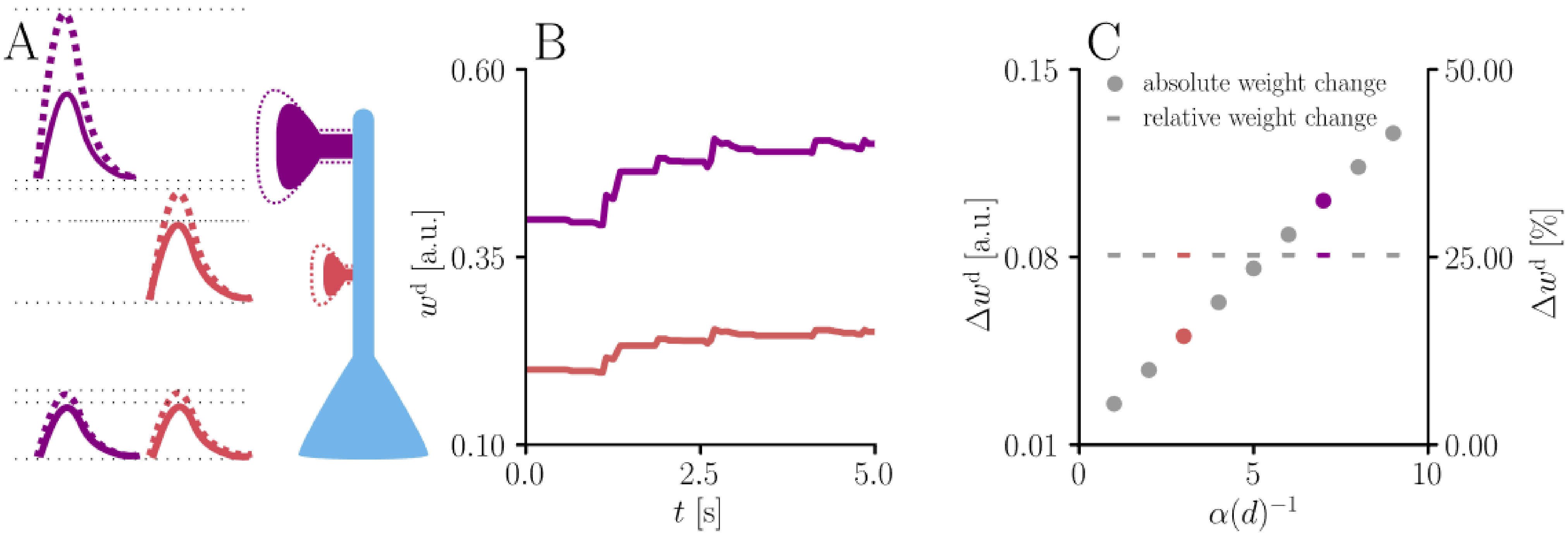
$$\dot{\boldsymbol{w}} = \eta \, \gamma_{\mathrm{s}} \left[Y^* - \phi(V) \right] \frac{\phi'(V)}{\phi(V)} \frac{1}{f'(\boldsymbol{w})} \left[c_\epsilon \frac{\boldsymbol{x}^\epsilon}{\boldsymbol{r}} - \gamma_{\mathrm{u}} + \gamma_{\mathrm{w}} f(\boldsymbol{w}) \right]$$

- 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.

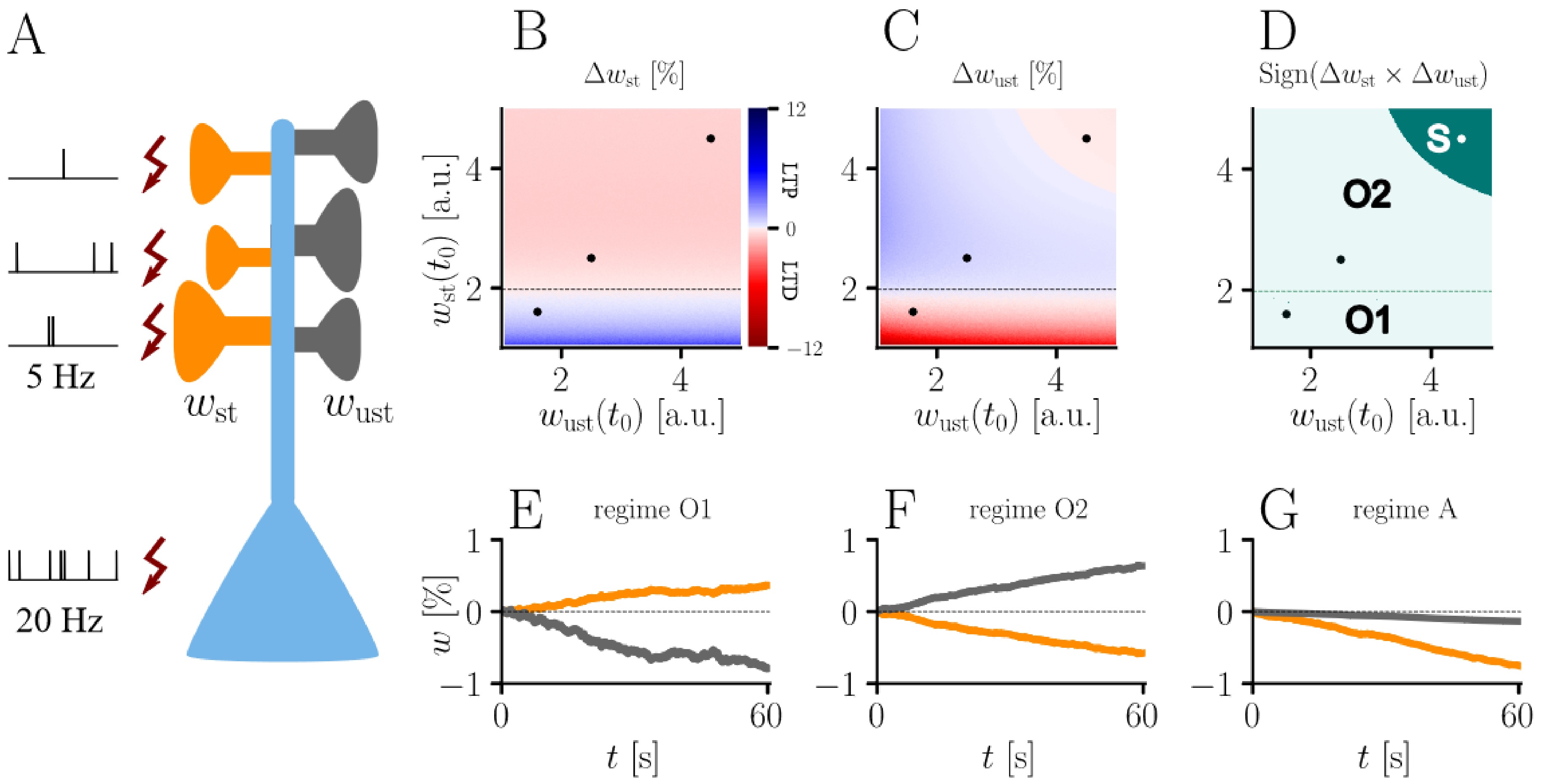
Performance



Synaptic democracy



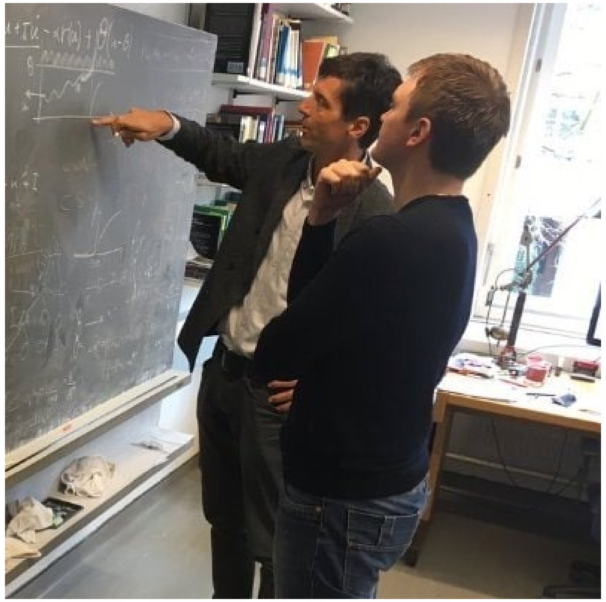
Interplay of homo- and heterosynaptic plasticity



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



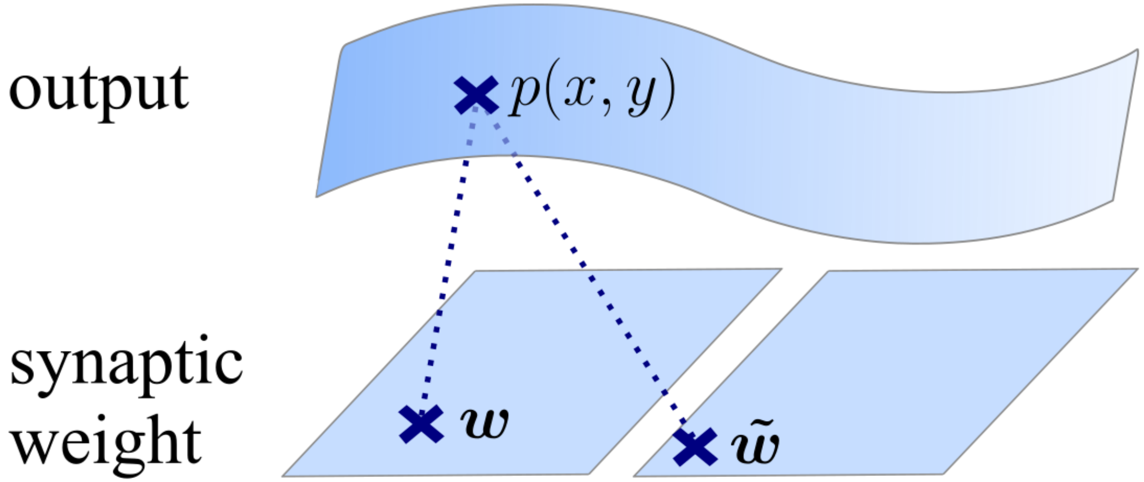
Walter Senn
Mihai Petrovici

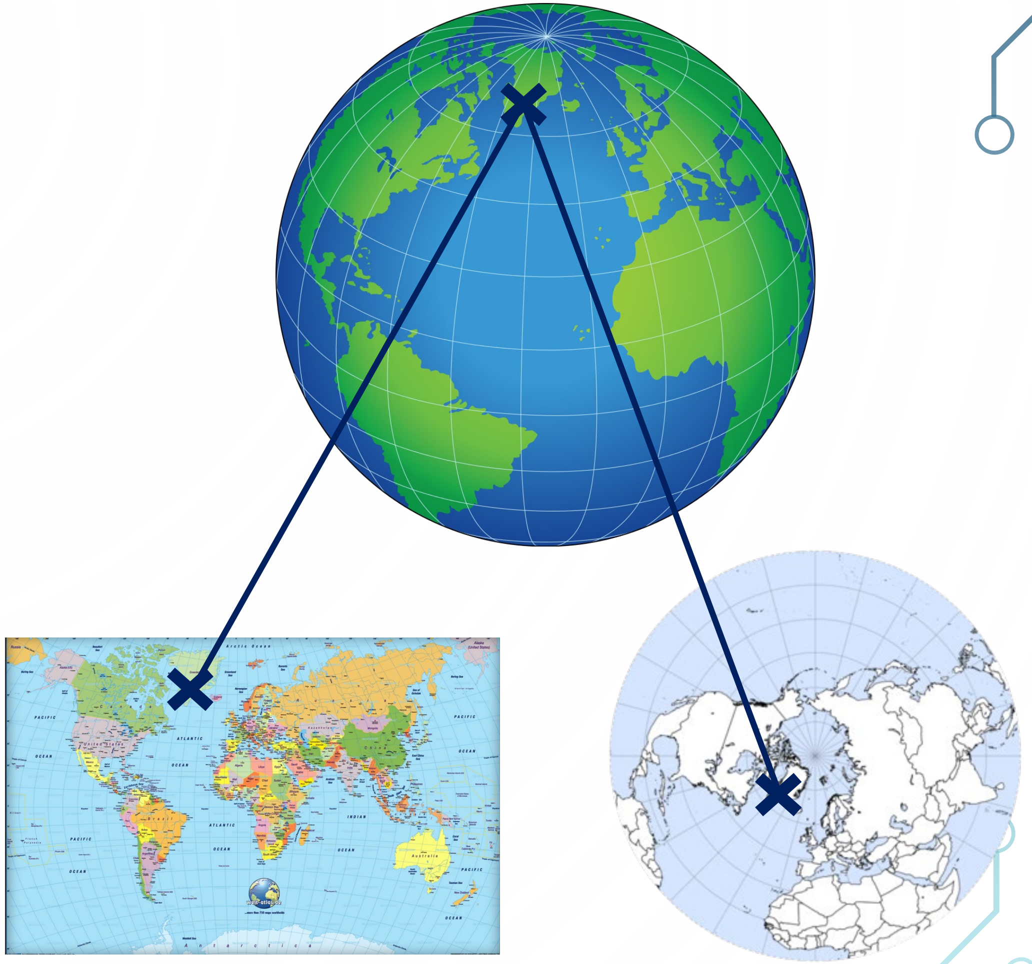
Computational Neuroscience and
Neuroscience TMA Groups
University of Bern

Human Brain Project

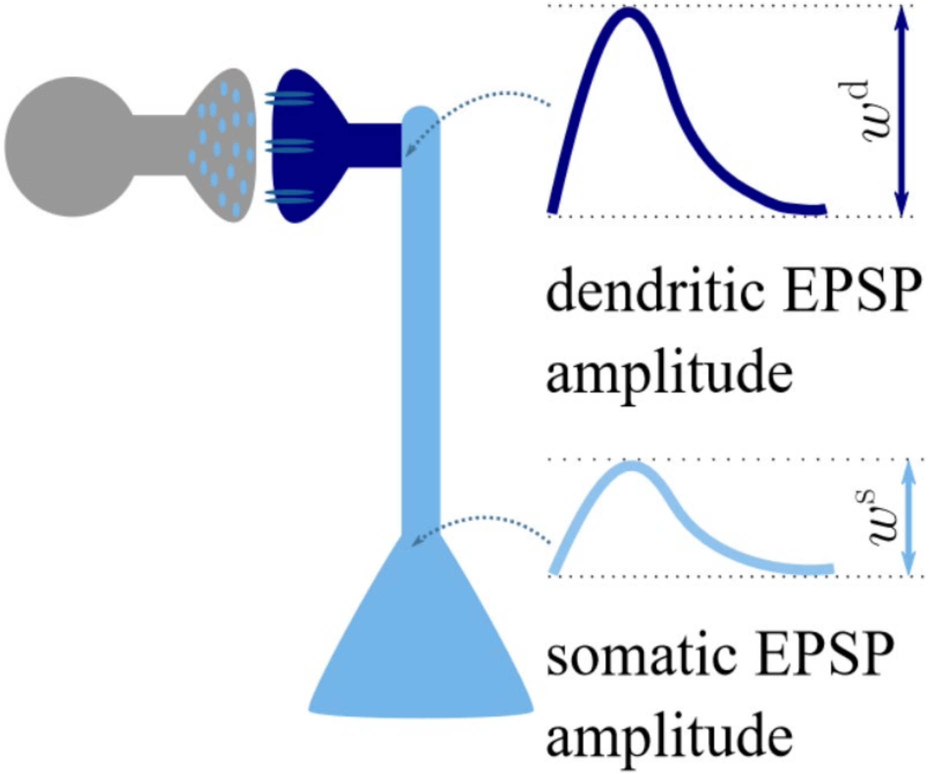
FNSNF
FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION

Parametrizations

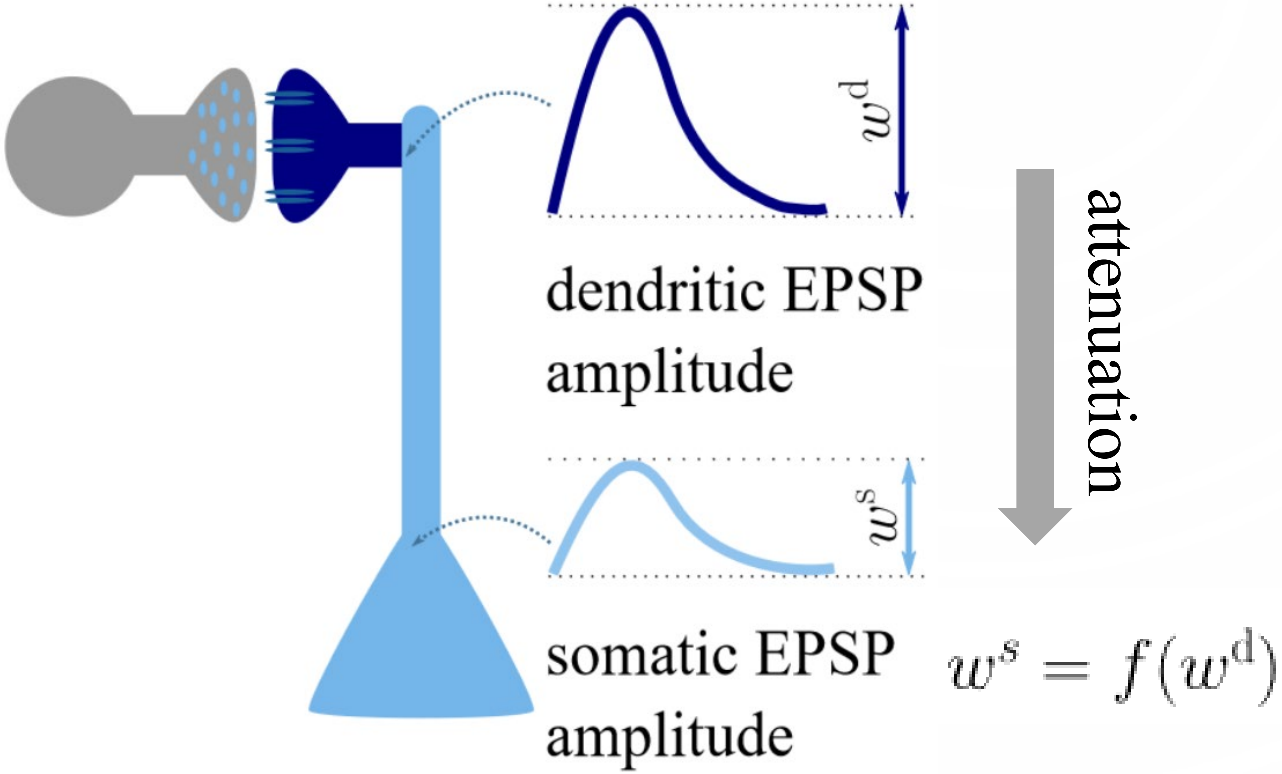




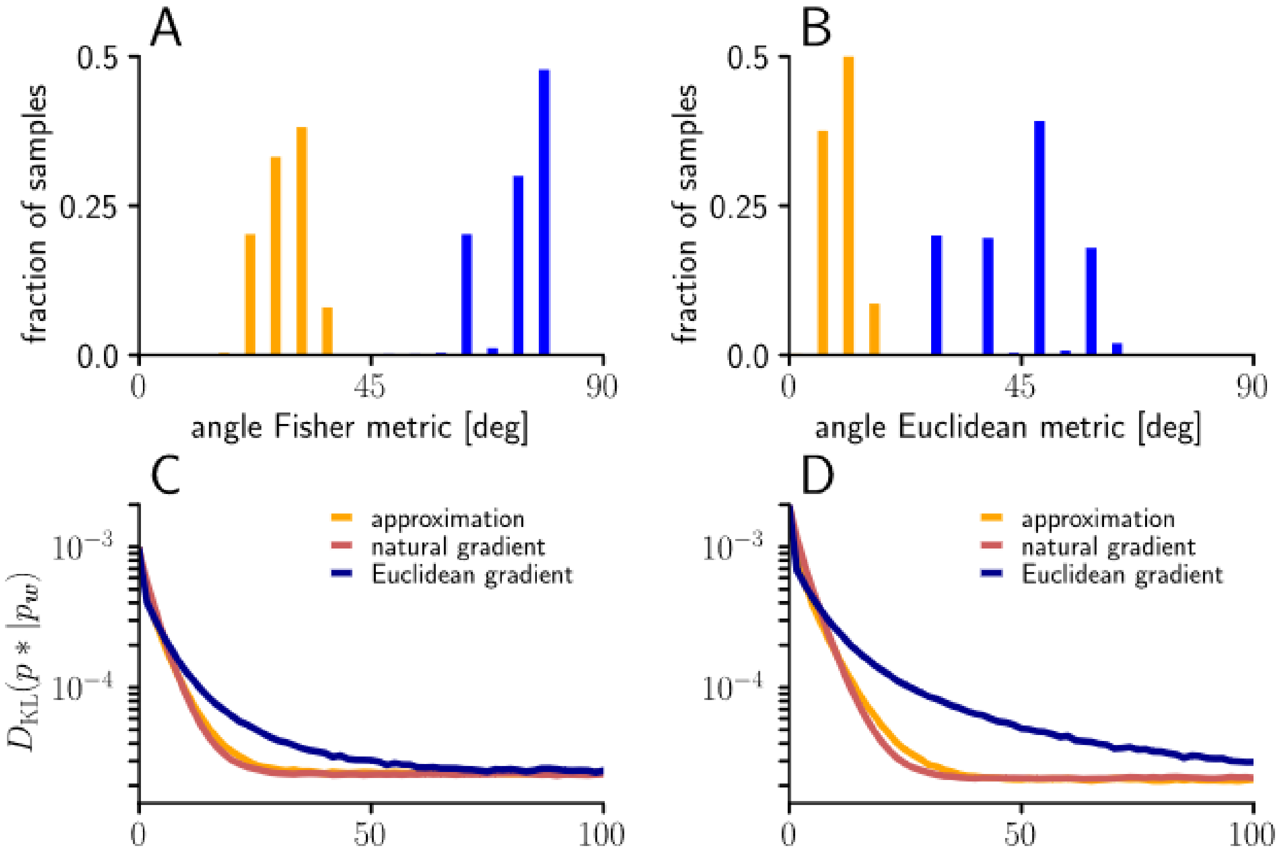
What is synaptic strength?



What is synaptic strength?

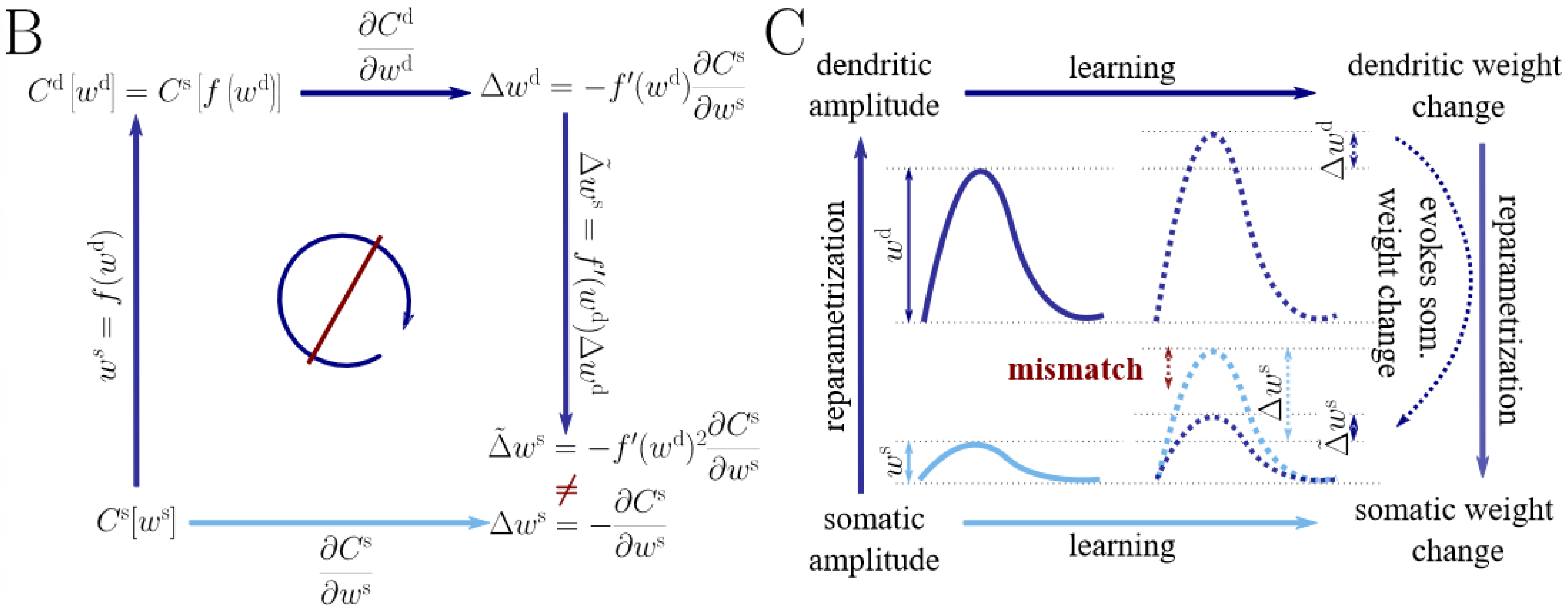


Approximation



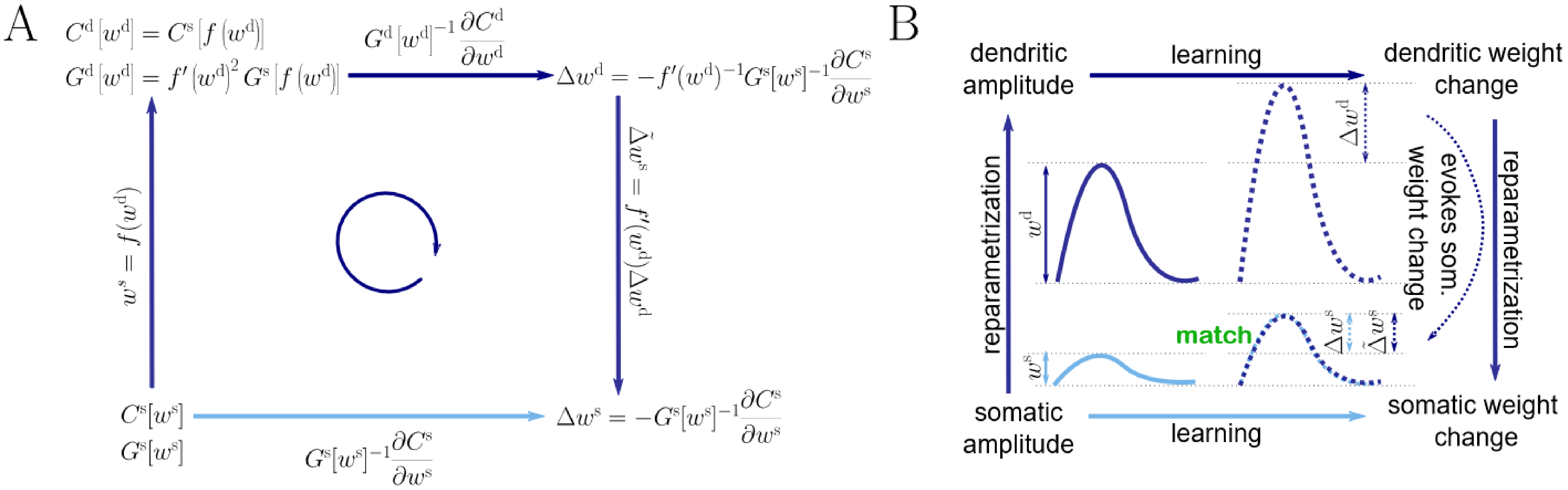
$$\dot{\boldsymbol{w}}_{\mathrm{a}} = \eta\,\gamma_{\mathrm{s}}\left[Y^{*} - \phi(V)\right]\frac{\phi'(V)}{\phi(V)}\,f'\left(\boldsymbol{w}\right)^{-1}\left[\frac{c_{\epsilon}\boldsymbol{x}^{\epsilon}}{\boldsymbol{r}} - c_{\epsilon}c_{\mathrm{u}} + c_{\mathrm{w}}V f\left(\boldsymbol{w}\right)\right]$$

Backup



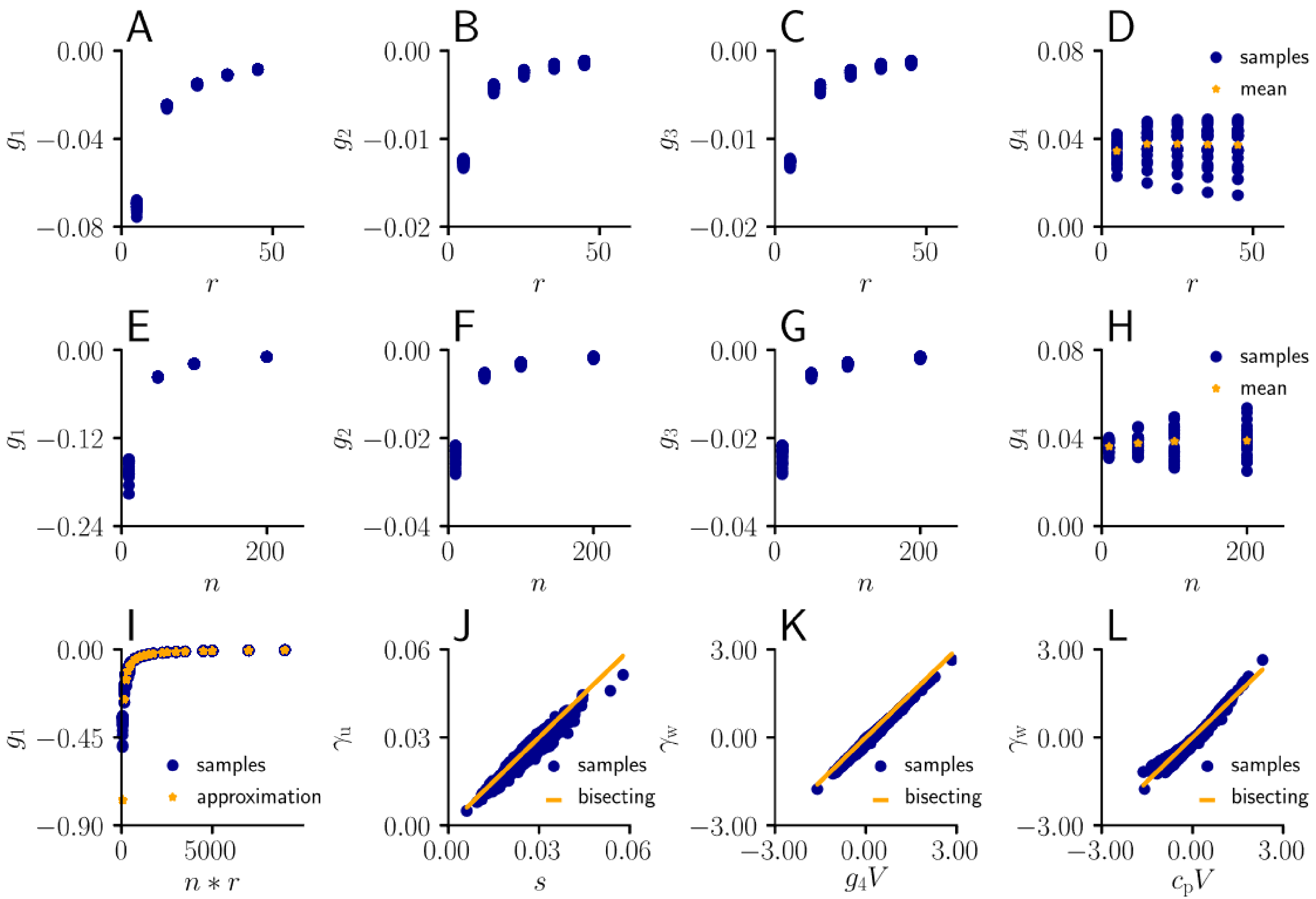
B
$C^d[w^d] = C^s[f(w^d)]$
$\frac{\partial C^d}{\partial w^d}$
$\Delta w^d = -f'(w^d)\frac{\partial C^s}{\partial w^s}$
$w^s = f(w^d)$
$\tilde{\Delta} w^s = f'(w^d)\Delta w^d$
$\tilde{\Delta} w^s = -f'(w^d)^2\frac{\partial C^s}{\partial w^s}$
$\neq$
$C^s[w^s]$
$\frac{\partial C^s}{\partial w^s}$
$\Delta w^s = -\frac{\partial C^s}{\partial w^s}$
C
dendritic amplitude
learning
dendritic weight change
reparametrization
w^d
Δw^d
weight change evokes som.
reparametrization
mismatch
Δw^s
$\tilde{\Delta} w^s$
w^s
somatic amplitude
learning
somatic weight change

Backup



A
$C^{d}[w^{d}] = C^{s}[f(w^{d})]$
$G^{d}[w^{d}] = f'(w^{d})^2 G^{s}[f(w^{d})]$
$G^{d}[w^{d}]^{-1}\frac{\partial C^{d}}{\partial w^{d}}$
$\Delta w^{d} = -f'(w^{d})^{-1}G^{s}[w^{s}]^{-1}\frac{\partial C^{s}}{\partial w^{s}}$
$w^{s} = f(w^{d})$
$\tilde{\Delta w}^{s} = f'(w^{d})\Delta w^{d}$
$C^{s}[w^{s}]$
$G^{s}[w^{s}]$
$G^{s}[w^{s}]^{-1}\frac{\partial C^{s}}{\partial w^{s}}$
$\Delta w^{s} = -G^{s}[w^{s}]^{-1}\frac{\partial C^{s}}{\partial w^{s}}$
B
dendritic amplitude
learning
dendritic weight change
reparametrization
w^{d}
Δw^{d}
weight change
evokes som.
reparametrization
match
w^{s}
Δw^{s}
$\tilde{\Delta w}^{s}$
somatic amplitude
learning
somatic weight change

Backup



Different types of plasticity

$$\Delta \boldsymbol{w} = \mathrm{error} \cdot \left(\Delta \boldsymbol{w}^{\mathrm{hom}} + \Delta \boldsymbol{w}^{\mathrm{het_u}} + \Delta \boldsymbol{w}^{\mathrm{het_w}}\right)$$