Synergies Between Intrinsic and Synaptic Plasticity Mechanisms

Jochen Triesch

Frankfurt Institute for Advanced Studies
http://fias.uni-frankfurt.de

Telluride, 2007-07-11
• basic theoretical research in physics, chemistry, biology, neuroscience, and computer science
• Frankfurt International Graduate School for Science
• ~110 people and growing

http://fias.uni-frankfurt.de
Different “Perspectives” on the Brain

**Perspective A:** The brain is a computation device. It finds solutions to certain computational problems. Sometimes these solutions are only approximate. ("top-down, computational (functional) view")

**Perspective B:** The brain is a complex dynamical system with many non-linearly interacting parts. The behavior emerging from these interactions is often difficult to predict ("bottom-up, physical view")
Forms of Neuronal Plasticity

- long-term potentiation (LTP) and depression (LTD) of synapses; modeled as Hebbian learning rules
- spike-timing-dependent plasticity (STDP); relation to LTP/LTD
- homeostatic synaptic scaling
- short term facilitation and depression of synapses
- modulation of synaptic plasticity by neuromodulators; reinforcement learning, attention
- adult neurogenesis
- short-term neuronal adaptation
- intrinsic plasticity (IP)
  - ...
What is Intrinsic Plasticity (IP)?

- not only synapses are plastic, but also soma (and dendrites) change properties through modification of voltage-gated channels [review in Zhang & Linden, 2003]
- changes in neuron’s frequency-current (f-I) curve
- found across large number of organisms and brain areas; multiple time scales [e.g. van Welie et al., 2004]
- some evidence consistent with idea of homeostasis: may keep activity level in desired regime [Desai et al., 1999]
Intrinsic Plasticity Example

- after activity deprivation, cultured pyramidal cells are more excitable [Desai et al., 1999]
Adapted from Yang Dan:

"sparse" coding of natural stimuli
Sensory Coding

• **sparse codes** are efficient way of representing sensory information [Barlow 1989; Földiak, 1995; Field, 1994, Olshausen&Field, 1996; Bell&Sejnowski, 1997; Rao&Ballard, 1999…]

• [Baddeley et al., 1997]: neurons in visual cortex of cat and monkey have close to exponential activity distributions (lifetime sparseness); maximize entropy for fixed energy consumption: “energy efficient coding”

Could intrinsic plasticity contribute to efficient coding? [Stemmler&Koch, 1999]

How does it interact with synaptic plasticity?
The right level of abstraction?

“Essentially, all models are wrong, but some are useful.”
George E. P. Box

“Make everything as simple as possible, but not simpler.”
Albert Einstein

Simplest kind of model:

\[ y = S(w^T x) = S(h) \]
**A Simple Model of Intrinsic Plasticity**

**Idea:** use parametric sigmoid nonlinearity, two adjustable parameters \( a \in \mathbb{R}^{>0}, b \in \mathbb{R} \)

\[
y = S_{ab}(h) = \frac{1}{1 + \exp\left(- (ah + b)\right)}
\]

Effect of varying \( a \) and \( b \): change shape of sigmoid

- **b shifts left/right**
  - \( b = -1 \)
  - \( b = 0 \)
  - \( b = 1 \)

- **a changes slope**
  - \( a = 1/2 \)
  - \( a = 1 \)
  - \( a = 2 \)
Excursion: Kullback-Leibler Divergence (relative entropy)

- Concept from information theory measuring how different two probability distributions \( f_a \) and \( f_b \) are:

\[
d(f_a \| f_b) \equiv \int f_a(y) \log \left( \frac{f_a(y)}{f_b(y)} \right) dy
\]

- Properties:
  - \( d(f_a \| f_b) \geq 0 \) and \( d(f_a \| f_b)=0 \) if and only if \( f_a=f_b \), i.e., if the two distributions are the same, their KL-divergence is zero otherwise it’s bigger
  - \( d(f_a \| f_b) \) in general is not equal to \( d(f_b \| f_a) \) (i.e. \( d(.\|.) \) is not a metric)
Learning Rule for IP

**Gradient rule:** consider Kullback-Leibler divergence between activity distribution and desired exponential of mean $\mu$:

$$D = d\left(f_y\|f_{\exp}^{\mu}\right) = \int f_y(y) \log \left(\frac{f_y(y)}{\frac{1}{\mu} \exp\left(-\frac{y}{\mu}\right)}\right) dy$$

$$= -H(y) + \frac{1}{\mu} E(y) + \log \mu$$

$H(y)$ and $E(y)$ depend on sigmoid parameters $a, b$.

Stochastic gradient descent rule for $a, b$ to minimize $D$:

$$\Delta a = \eta \left(a^{-1} + h - \left(2 + \mu^{-1}\right)y + \mu^{-1}hy^2\right)$$

$$\Delta b = \eta \left(1 - \left(2 + \mu^{-1}\right)y + \mu^{-1}hy^2\right)$$

this rule is strictly local!
**Example 1:** Gaussian input $h$ to the transfer function

The image shows the distribution of $h$ and the output distribution. The activation function is labeled as learned and optimal. The graph on the right plots the frequency against the activity $y$.
Example 2: uniform input $h$ to the transfer function

**Distribution of $h$**

**Activation fct.: learned, optimal**

**Output distribution**

frequency

activity $y$
Example 3: exponential input $h$ to the transfer function

distribution of $h$

output distribution

activation fct.: learned, optimal
Example 4: Recovery from “sensory deprivation”
Combination with Hebbian Learning

**Question:** [Triesch, NIPS 2005; Neural Comp., 2007] What is the interaction of Hebbian synaptic plasticity and intrinsic plasticity?

Simple Hebb rule with weight normalization:

\[
\Delta w = \eta y x = \eta S_{ab} (w^T x) x
\]

**Result:** The combination of intrinsic with Hebbian plasticity can result in the neuron discovering heavy-tailed (interesting) directions in the input.

IP+Hebb can support Independent Component Analysis (ICA)
Intuition

Limiting case:

- intrinsic plasticity much faster than Hebbian plasticity
- assume intrinsic plasticity achieves approximately exponential firing rate distribution before weight changes much

Consequence: \[ \langle \Delta w \rangle = \eta \langle yx \rangle \]

Weight update is exponentially weighted sum of all the inputs.

\[ p(y) \approx \frac{1}{\mu} \exp\left(-\frac{y}{\mu}\right) \]
**Illustration:**
- consider neuron with two inputs
- Hebbian learning in linear unit wanders aimlessly
- Hebbian learning + IP discovers heavy-tailed direction

“Laplace band” : \[ \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \]
The “Bars Problem” (non-linear ICA)

[Földiak, 1990]:
• bars on $R \times R$ retina (horiz. & vert.)
• each shows up independently with certain probability; $\sim 2^{2R}$ patterns
• problem is to find the individual bars
Is the time scale of IP important?

Answer: No, as long as it is present at all, it can be much faster or much slower than synaptic plasticity.
Full Bars Problem

- de-correlation mechanism [Butko&Triesch, Neurocomp. 2007]:
  - "winning" unit learns in Hebbian fashion
  - other units in anti-Hebbian fashion

\[
\Delta w_i^{(a)} = x y N(y_i; y)
\]

\[
N_{bars}(y_i; y) = \begin{cases} 
1 : y_i = \max(y) \\
-\beta : \text{else}
\end{cases}
\]
learning works reliably over a range of $\beta$-values
Learning Orientation Maps

- Hebb rule with additional center-surround gating around the most activated unit [Butko&Triesch, Neurocomp. 2007]:

\[
\Delta w_i^{(a)} = x y N(y_i; y) \\
N_{\text{map}}(y_i; y) = \frac{1}{2\pi\sigma_c^2} \exp\left(\frac{-d_i^2}{2\sigma_c^2}\right) - \frac{1}{2\pi\sigma_s^2} \exp\left(\frac{-d_i^2}{2\sigma_s^2}\right)
\]

- \(N_{\text{map}}\) is a Mexican hat centered at most activated unit
- natural images (10x10 patches) of van Hateren database
- preprocessing by whitening
15x15 units, 2.25 times overcomplete representation:

- network learns maps of oriented filters
- low average dependence between units (normalized mutual information measure, 3% of max.)
Receptive fields, non-linearity, and activity histograms
Generative Model of Natural Scenes

- joint work with Cornelius Weber [in revision]
- based on “Helmholtz machine” [Dayan et al., 1995; Hinton et al., 1995]
- exploits multiple time scales of plasticity
- model develops simple cell-like receptive fields
- offers explanation of tilt after-effect (TAE) as adaptive response to maintain efficient coding [Wainwright, 1999]
"wake" phase:  
\[ \hat{x} = \tilde{x} - W^{td} \tilde{z} \]
\[ \Delta w_{ji}^{td} = \eta_w \hat{x}_j z_i \]

"sleep" phase:  
\[ \hat{z} = \tilde{z} - g(W^{bu} \tilde{x}) \]
\[ \Delta w_{ij}^{bu} = \eta_w \hat{z}_i \tilde{x}_j \]
Learned Receptive Fields
Tilt after-effect (TAE)

- after adapting to a grating for several seconds or minutes, nearby orientations are perceived as tilted away from the adapting orientation

[He & McLeod, 2001]

Diagram: Adapt (5 s) → Test (250 ms) → Perception (subjective horizontal)

250 ms
Tilt After-Effect in the Model

**Adaptation phase:**
- present gratings of same orientation but different phase to the network
- synaptic weights stay fixed, but a and/or b parameters are allowed to adapt according to IP learning rule

**Test phase:**
- present test gratings of varying orientations
- use population vector decoding to estimate “perceived orientation” of test gratings; plot change of orientation as a function of orientation difference between adaptation orientation and test orientation
Tilt After-Effect Results

- gain changes of the model (parameter a) produce good fit to TAE data from human subjects
- roughly logarithmic scaling with duration (as in humans)
Conclusions

• Different plasticity mechanisms may interact synergistically in the brain; emergence of interesting learning properties
• Combination of IP with Hebbian rules allows learning of efficient sensory codes (close ties to sparse coding, ICA)
• In order to understand cortical plasticity we have to study the interaction of different plasticity mechanisms

Current/Future work:
• learning hierarchical representations; inference
• optimal predictability, slowness, ...
• IP in recurrent networks, “liquid computing” etc.
• IP for spiking model neurons, interaction with STDP
• relation to criticality, avalanches in recurrent networks?
Thank you!
modeling cognitive development in infants

models of plasticity and learning in cortical networks

cognitive robotics

computer vision

visual psychophysics

Open positions: 1 post-doc, 1 PhD student