Deriving neural circuits from first principles

Dmitri "Mitya" Chklovskii
Simons Center for Data Analysis
Simons Foundation, New York City
&
NYU Medical School

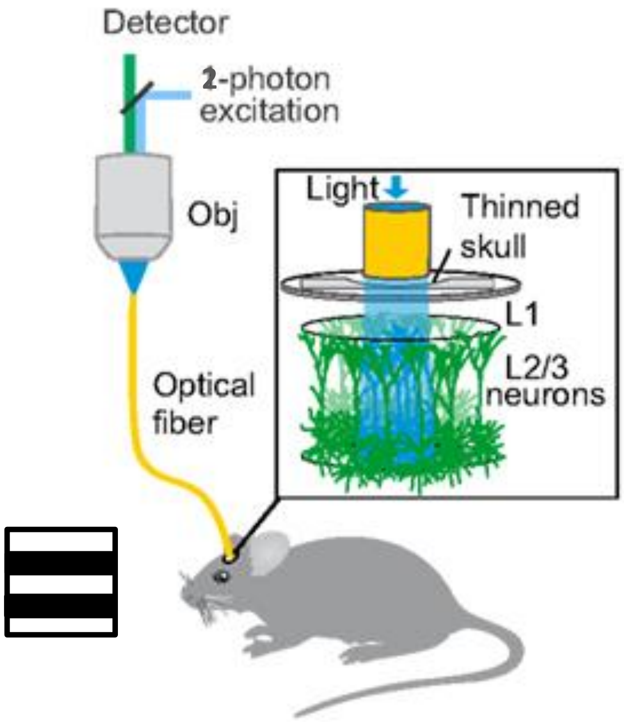
Richard Feynman
1918-1988

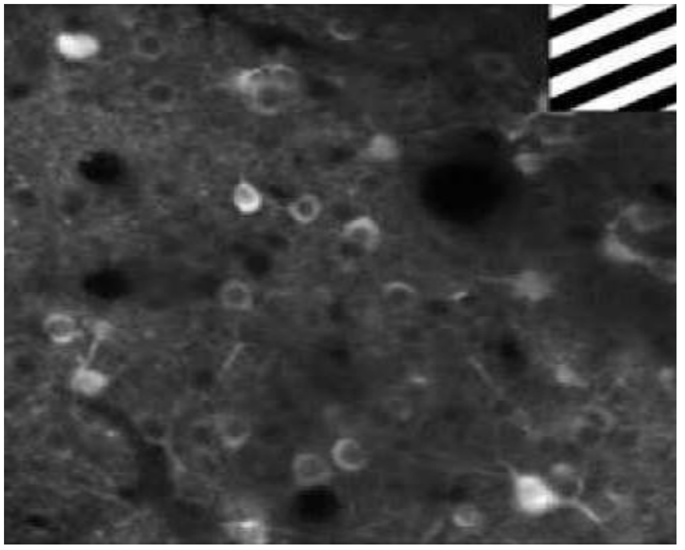
"In general we look for a new law by the following process. First we guess it. Then we compute the consequences of the guess to see what would be implied if this law that we guessed is right. Then we compare the result of the computation to nature, with experiment or experience, compare it directly with observation, to see if it works… if it disagrees with experiment it is wrong. That is all there is to it."

Deriving neural circuits from first principles

- Representation/decoding approach
  - Single-neuron PCA
  - Soft-thresholding neuron
  - Multiple neurons
- Similarity matching approach
  - Linear dimensionality reduction
  - Nonnegative output

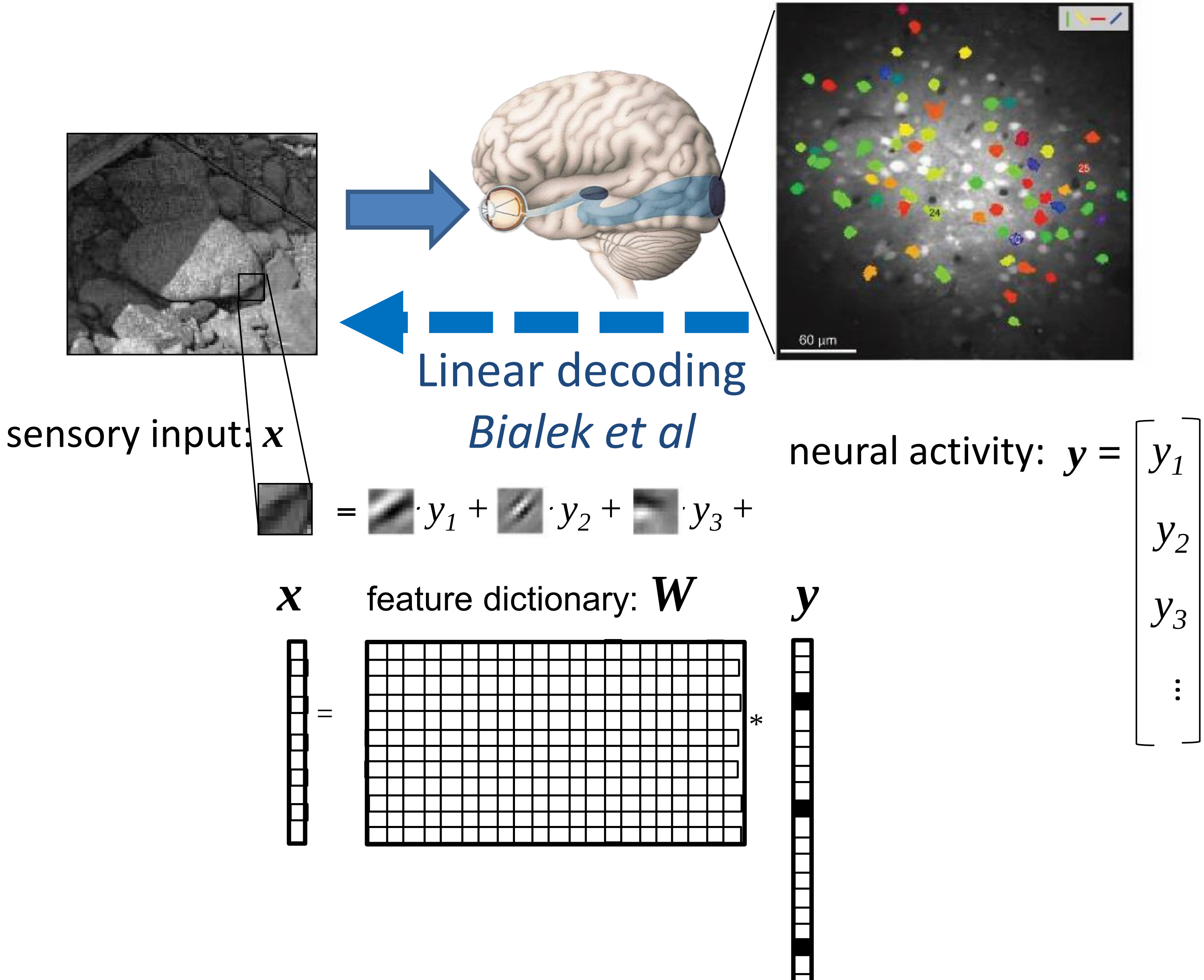
Imaging *in vivo* activity of neuronal populations



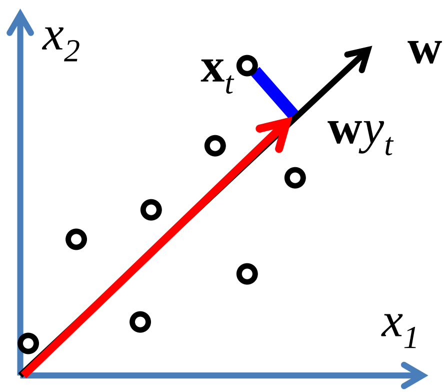


Representation/decoding approach

What does neural activity represent?

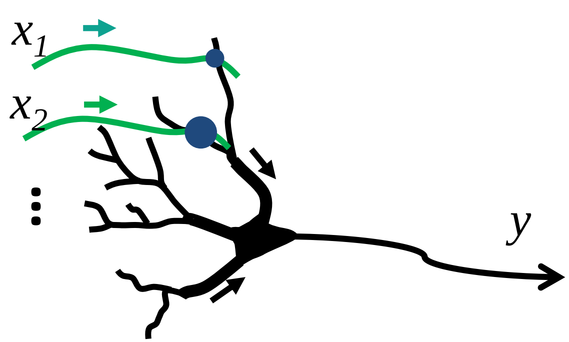


Principal component analysis (PCA)



$$\min_{\mathbf{w}, y_{1..T}} \sum_{t=1}^{T} \left\| \mathbf{x}_t - \mathbf{w} y_t \right\|_2^2$$

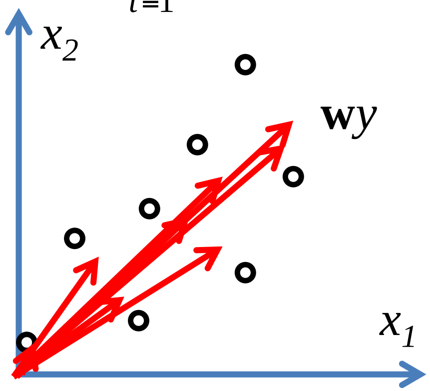
Linear neuron



Online PCA

Oja, 1982; Yang, 1995

$$\min_{\mathbf{w}, y_{1..T}} \sum_{t=1}^{T} \left\| \mathbf{x}_t - \mathbf{w} y_t \right\|_2^2$$



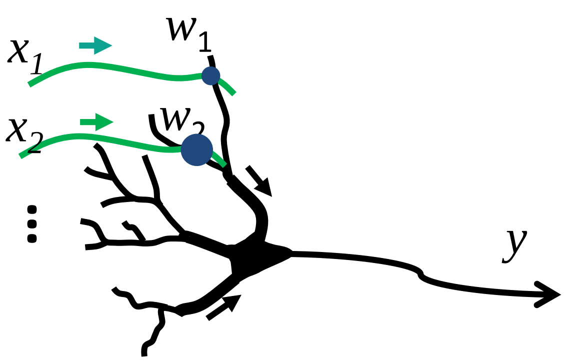
$$y_T = \arg\min_y \left\| \mathbf{x}_T - \mathbf{w}_{T-1} y \right\|_2^2$$

$$= \arg\min_y \left[-2\mathbf{w}_{T-1}^{\mathsf{T}} \mathbf{x}_T y + \left\| \mathbf{w}_{T-1} \right\|_2^2 y^2 \right]$$

$$= \arg\min_y \left[\frac{\mathbf{w}_{T-1}^{\mathsf{T}} \mathbf{x}_T}{\left\| \mathbf{w}_{T-1} \right\|_2^2} - y \right]^2 \left\| \mathbf{w}_{T-1} \right\|_2^2$$

$$y_T = \mathbf{w}_{T-1}^{\mathsf{T}} \mathbf{x}_T \Big/ \left\| \mathbf{w}_{T-1} \right\|_2^2$$

Output = weighted input



Online PCA learning rule

Oja, 1982; Yang, 1995

$$\min_{\mathbf{w}, y_{1..T}} \sum_{t=1}^{T} \left\| \mathbf{x}_t - \mathbf{w} y_t \right\|_2^2$$

$$\mathbf{w}_T = \arg\min_{\mathbf{w}} \sum_{t=1}^{T} \left\| \mathbf{x}_t - \mathbf{w} y_t \right\|_2^2$$

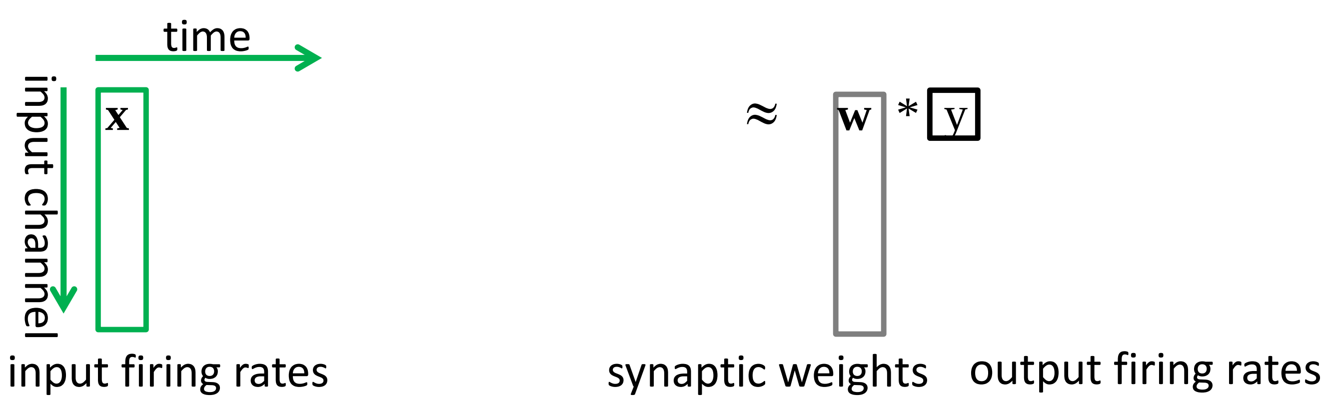
$$= \arg\min_{\mathbf{w}} \sum_{t=1}^{T} \left[-\mathbf{w}^{\mathrm{T}} \mathbf{x}_t y_t + \left\| \mathbf{w} \right\|_2^2 y_t^2 \right]$$

$$= \arg\min_{\mathbf{w}} \left\| \frac{\sum_{t=1}^{T} \mathbf{x}_t y_t}{\sum_{t=1}^{T} y_t^2} - \mathbf{w} \right\|_2^2 = \frac{\sum_{t=1}^{T} \mathbf{x}_t y_t}{\sum_{t=1}^{T} y_t^2}$$

$$\mathbf{w}_T = \mathbf{w}_{T-1} + y_T \left(\mathbf{x}_T - \mathbf{w}_{T-1} y_T \right) \Big/ \sum_{t=1}^{T} y_t^2$$

Hebbian synaptic learning rule

Online matrix factorization



Online PCA learning rule

Oja, 1982; Yang, 1995

$$\min_{\mathbf{w}, y_{1..T}} \sum_{t=1}^{T} \left\| \mathbf{x}_t - \mathbf{w} y_t \right\|_2^2$$

$$\mathbf{w}_T = \arg\min_{\mathbf{w}} \sum_{t=1}^{T} \left\| \mathbf{x}_t - \mathbf{w} y_t \right\|_2^2$$

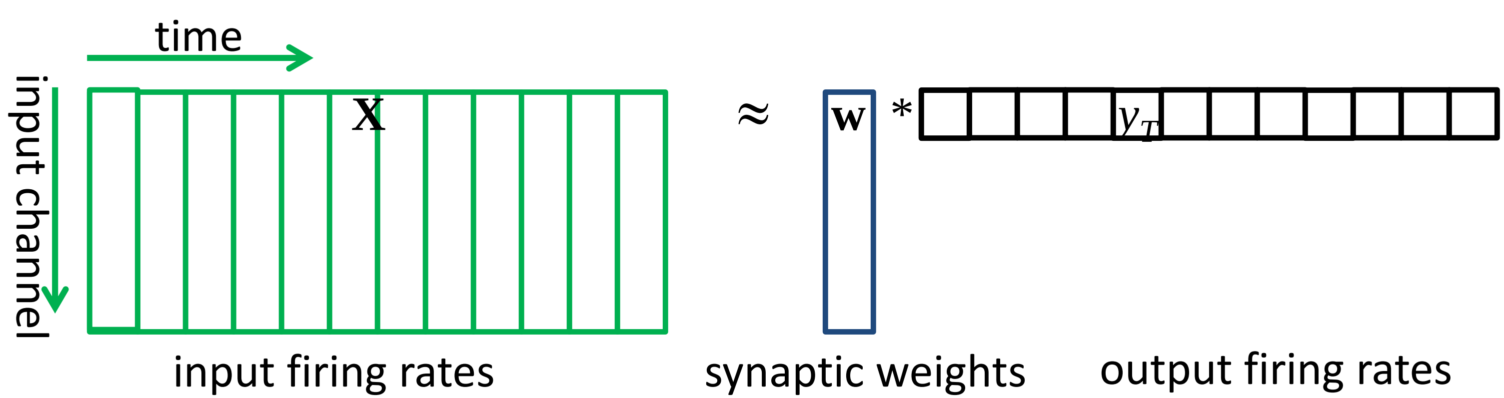
$$= \arg\min_{\mathbf{w}} \sum_{t=1}^{T} \left[-\mathbf{w}^{\mathrm{T}} \mathbf{x}_t y_t + \left\| \mathbf{w} \right\|_2^2 y_t^2 \right]$$

$$= \arg\min_{\mathbf{w}} \left\| \frac{\sum_{t=1}^{T} \mathbf{x}_t y_t}{\sum_{t=1}^{T} y_t^2} - \mathbf{w} \right\|_2^2 = \frac{\sum_{t=1}^{T} \mathbf{x}_t y_t}{\sum_{t=1}^{T} y_t^2}$$

$$\mathbf{w}_T = \mathbf{w}_{T-1} + y_T \left(\mathbf{x}_T - \mathbf{w}_{T-1} y_T \right) \Big/ \sum_{t=1}^{T} y_t^2$$

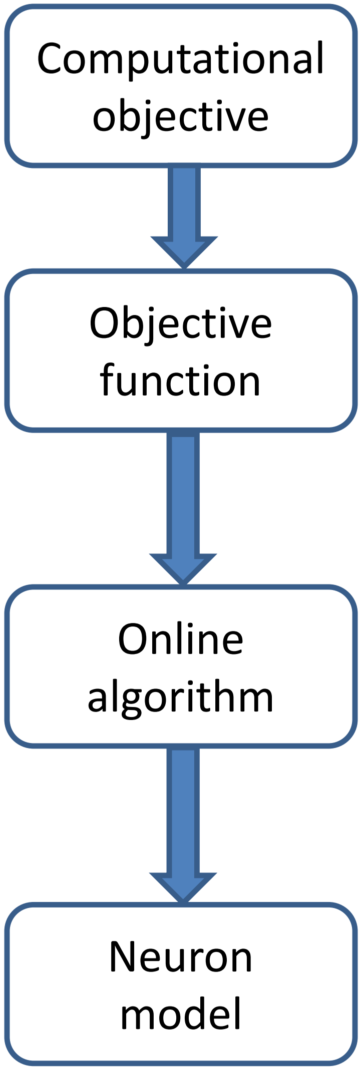
Hebbian synaptic learning rule

Online matrix factorization



Online PCA

cf. Levels of analysis *(Marr & Poggio, 1976)*

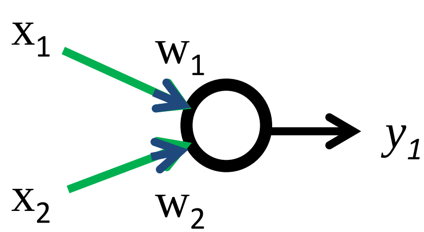


Principal component analysis (PCA)

$$\min_{\mathbf{w}, y_{1..T}} \sum_{t=1}^{T} \left\| \mathbf{x}_t - \mathbf{w} y_t \right\|_2^2$$

$$y_T = \mathbf{w}_{T-1}^{\mathsf{T}} \mathbf{x}_T \Big/ \left\| \mathbf{w}_{T-1} \right\|_2^2$$

$$\mathbf{w}_T = \mathbf{w}_{T-1} + y_T \left(\mathbf{x}_T - \mathbf{w}_{T-1} y_T \right) \Big/ \sum_{t=1}^{T} y_t^2$$



Dependence of synaptic plasticity on age and activity

Theory

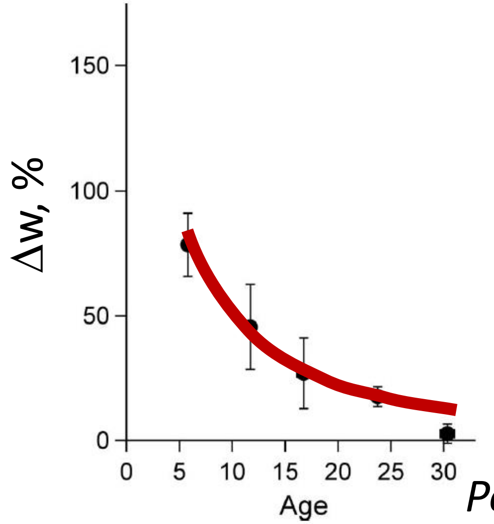
$$\mathbf{w}_T = \mathbf{w}_{T-1} + y_T \left(\mathbf{x}_T - \mathbf{w}_{T-1} y_T \right) \Big/ \sum_{t=1}^{T} y_t^2$$

Nonstationary input statistics -> discounting of past errors or forgetting:

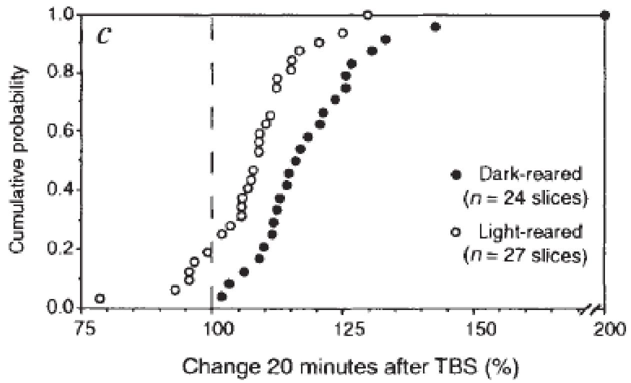
$$\mathbf{w}_T = \arg\min_{\mathbf{w}} (1-\alpha) \sum_{t=1}^{T} \alpha^{T-t} \left\| \mathbf{x}_t - \mathbf{w} y_t \right\|_2^2, \qquad \mathbf{w}_T = \alpha \mathbf{w}_{T-1} + y_t \left(\mathbf{x}_t - \mathbf{w} y_t \right) \Big/ \sum_{t=1}^{T} \alpha^{T-t} y_t^2$$

$\alpha = e^{-1/\tau_w} < 1, \quad \tau_w$ - autocorrelation time of $\mathbf{w}$

Experiment



Poo & Isaacson 2007

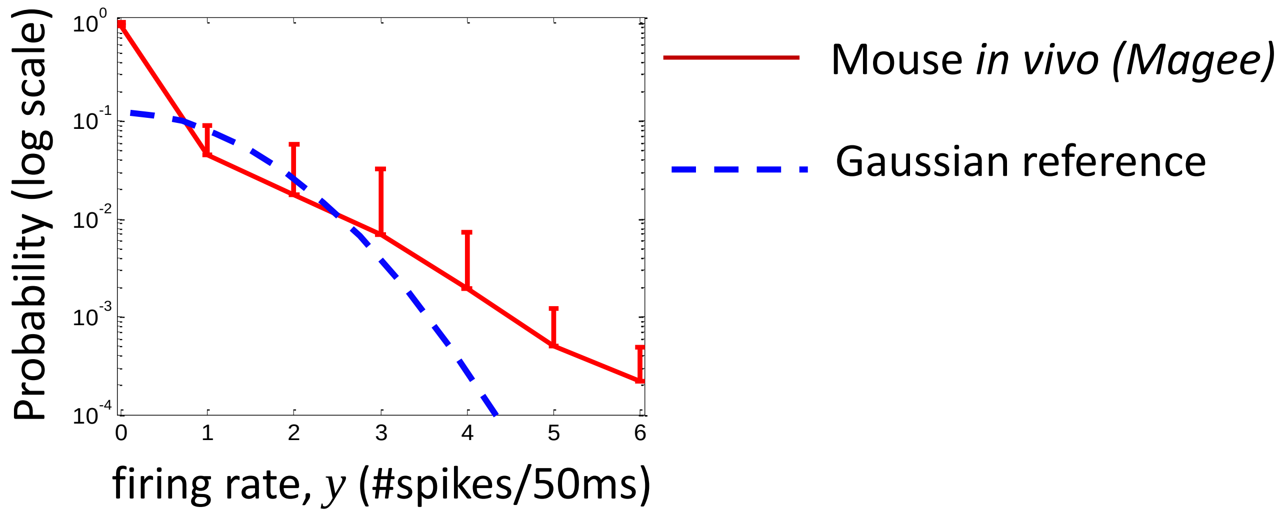


Kirkwood, Lee, Bear 1995

Deriving neural circuits from first principles

- Representation/decoding approach
  - Single-neuron PCA
  - Soft-thresholding neuron
  - Multiple neurons
- Similarity matching approach
  - Linear dimensionality reduction
  - Nonnegative output

Statistics of firing rates



Firing rate distribution is nonGaussian: sparse and heavy-tailed → Add a sparsity-inducing regularizer to the cost function

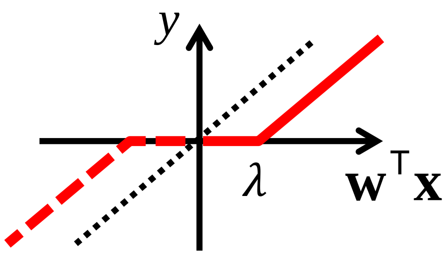
$$\min_{\mathbf{w}, y_{1..T}} \sum_{t=1}^{T} \Big[\underbrace{\frac{1}{2} \left\| \mathbf{x}_t - \mathbf{w} y_t \right\|_2^2}_{\text{log-likelihood}} + \lambda \underbrace{|y_t|}_{\text{log-prior}} \Big]$$

Nonlinear output function

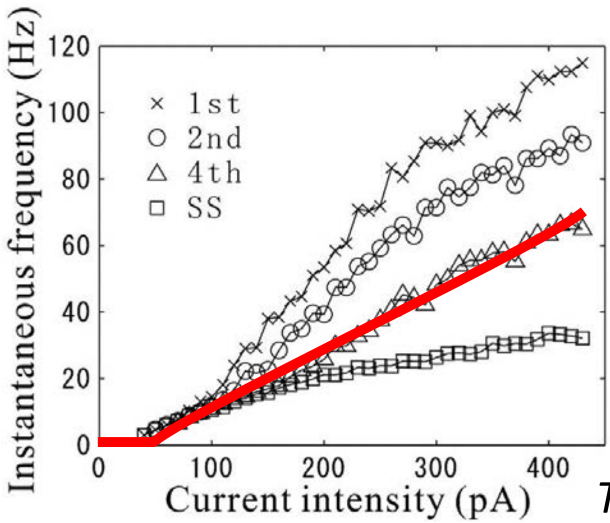
Theory

$$y_T = \arg\min_y \left[\frac{1}{2} \left\| \mathbf{x}_T - \mathbf{w}_{T-1} y \right\|_2^2 + \lambda |y| \right]$$

$$= \frac{1}{\left\| \mathbf{w}_{T-1} \right\|_2^2} \mathrm{SoftThresh}\left(\mathbf{w}_{T-1}^{\mathsf{T}} \mathbf{x}_T, \lambda \right)$$



Experiment



Type I:

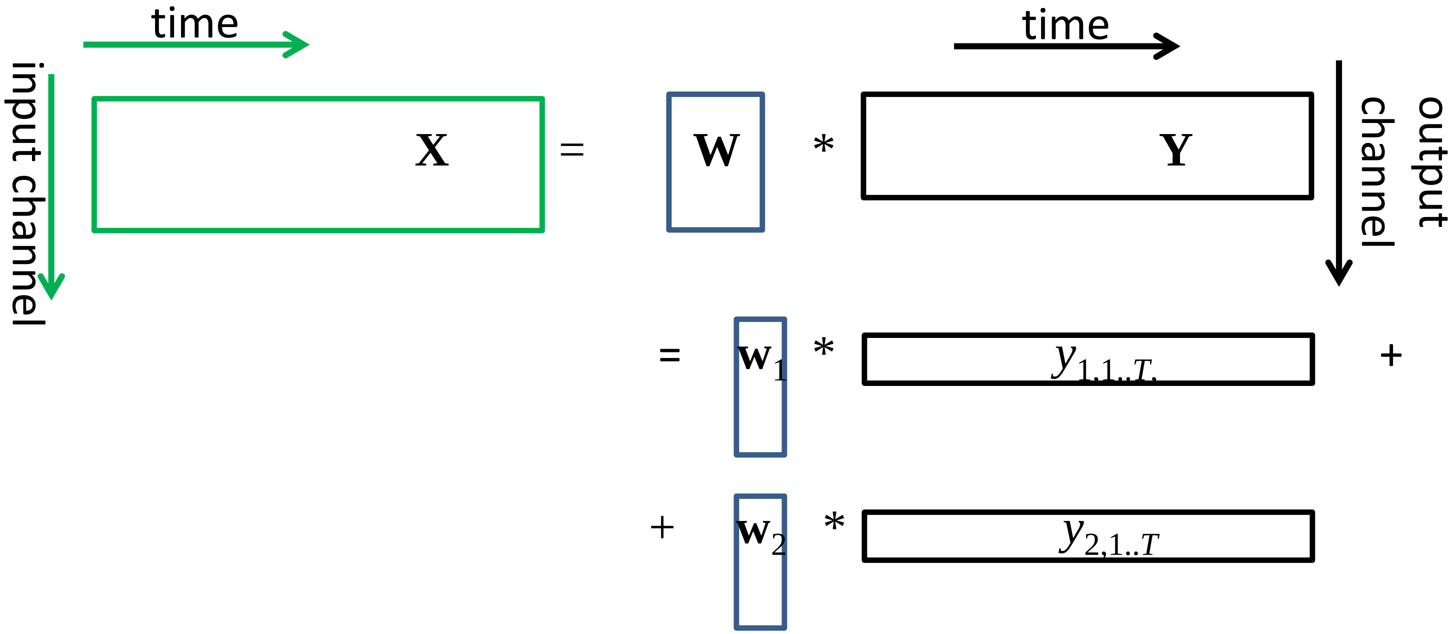
Pyramidal cell

Tateno et al. (2004)

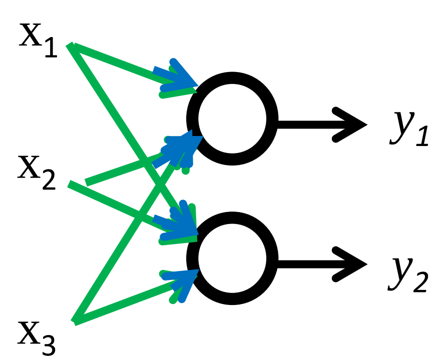
Deriving neural circuits from first principles

- Representation/decoding approach
  - Single-neuron PCA
  - Soft-thresholding neuron
  - Multiple neurons
- Similarity matching approach
  - Linear dimensionality reduction
  - Nonnegative output

Multiple output channels



$$\min_{\mathbf{W}} \sum_t \min_{y_t} \left\| \mathbf{x}_t - \mathbf{W} \boldsymbol{y}_t \right\|_2^2$$



multiple neurons

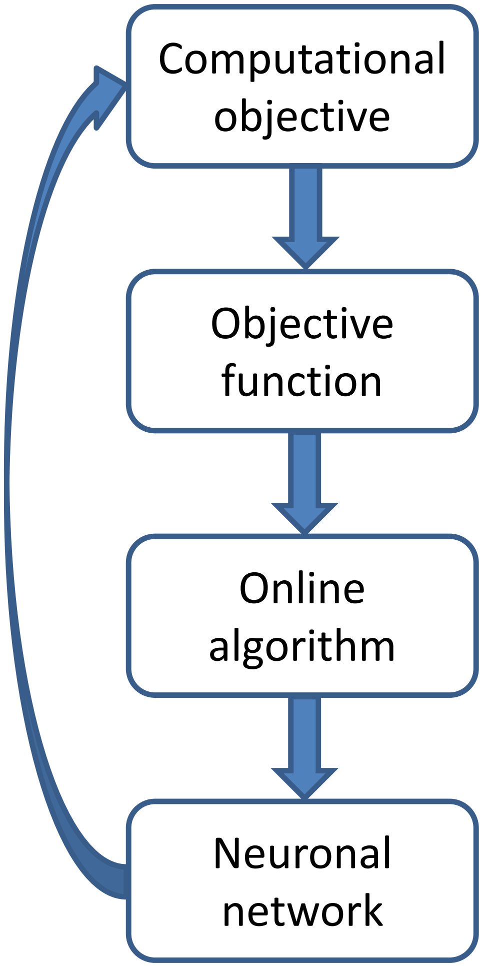
Sparse dictionary learning

(Olshausen & F

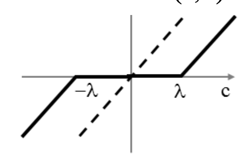
No objective function — *Foldiak (1990), Oja (1992), Zylberberg et al (2011)*

or

Biologically implausible nonlocal learning rules — *(Bell & Sejnowski(1997), Gerhard et al(2009), Falconbridge et al(2005)*

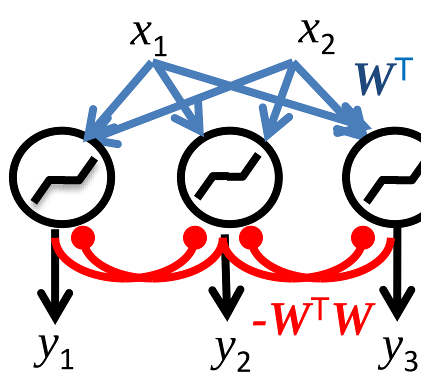


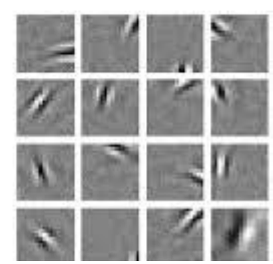
$$\min_{\boldsymbol{W}} \sum_t \min_{y_t} \left(\overbrace{\frac{1}{2}\left\|\boldsymbol{x}_t - \boldsymbol{W}\boldsymbol{y}_t\right\|_2^2}^{\text{representation error}} + \lambda \overbrace{\left\|\boldsymbol{y}_t\right\|_1}^{\text{sparsity}} \right)$$



neural activity: $\boldsymbol{y}_t \leftarrow \mathrm{SoftThresh}\left(\boldsymbol{W}^{\top}\boldsymbol{x}_t - \boldsymbol{W}^{\top}\boldsymbol{W}\boldsymbol{y}_t, \lambda\right)$

synaptic weight: $W_{j,i,t+1} \leftarrow W_{j,i,t} + y_{i,t}\left[x_{j,t} - \sum_k W_{j,k,t} y_{k,t} \right]\eta$





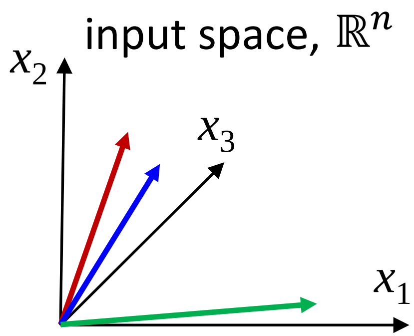
learned dictionary, $\boldsymbol{W}$

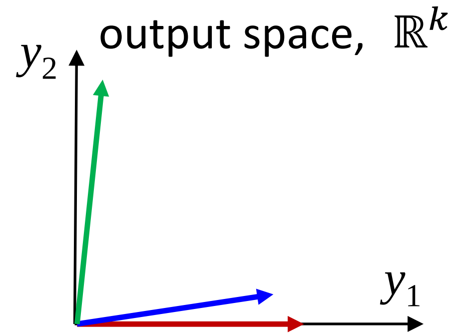
Alternative two-layer reciprocal circuit: Olshausen & Field, 1997; Koulakov & Rinberg, 2011; Druckmann, Hu, Chklovskii, 2012

Deriving neural circuits from first principles

- Representation/decoding approach
  - Single-neuron PCA
  - Soft-thresholding neuron
  - Multiple neurons
- Similarity matching approach
  - Linear dimensionality reduction
  - Nonnegative output

Similar inputs produce similar outputs





Quantify similarities by scalar products

$$\mathbf{x}_1^\top \mathbf{x}_2 = \sum_{i=1}^{n} x_{i,1} x_{i,2} \qquad \mathbf{y}_1^\top \mathbf{y}_2 = \sum_{i=1}^{k} y_{i,1} y_{i,2}$$

Matrix notation

$$\mathbf{X} = [\mathbf{x}_1 \ \cdots \ \mathbf{x}_t] \qquad \mathbf{Y} = [\mathbf{y}_1 \ \cdots \ \mathbf{y}_t]$$

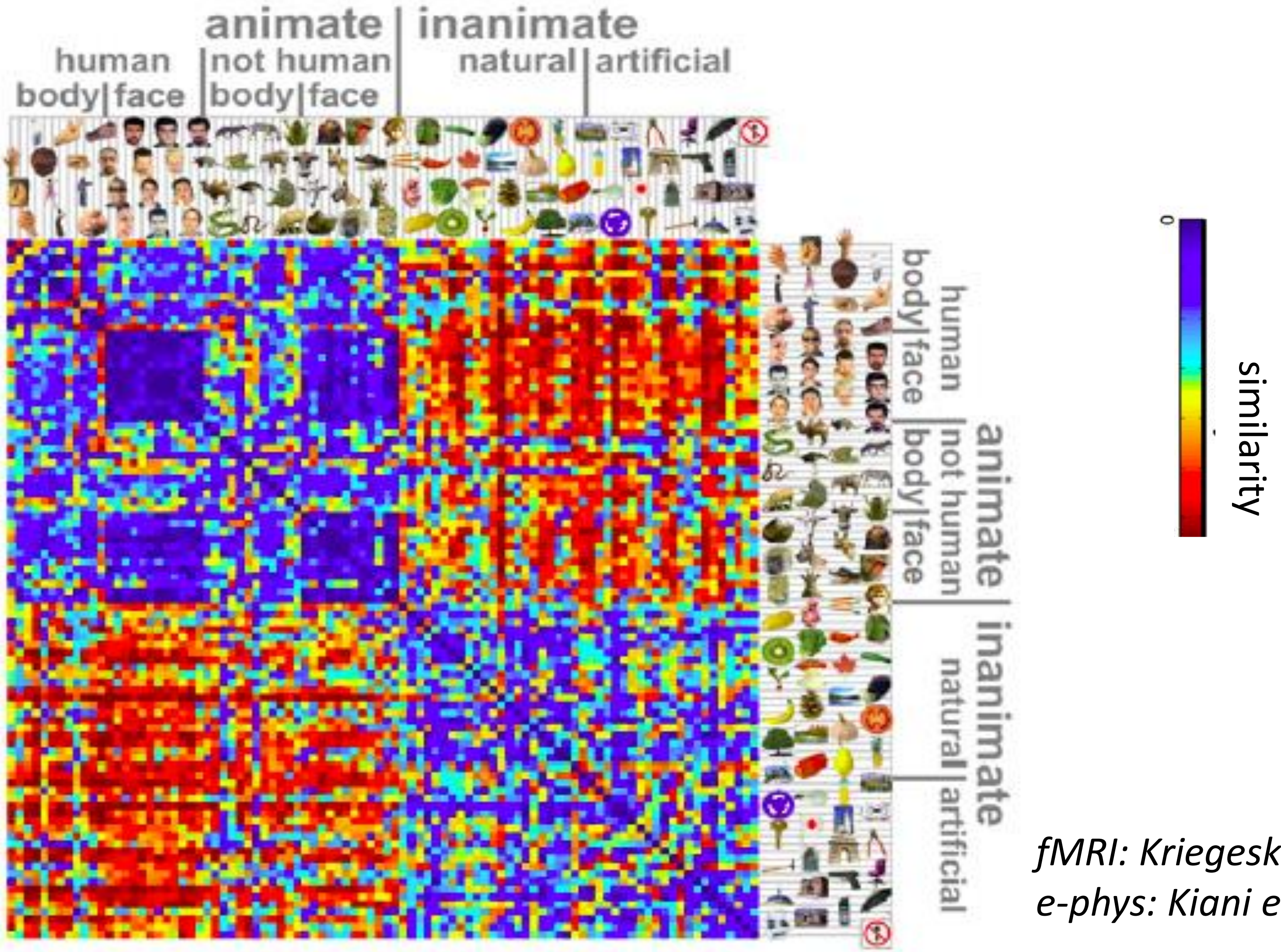
$$\mathbf{X}^\top \mathbf{X} = \begin{bmatrix} \mathbf{x}_1^\top \mathbf{x}_1 & \mathbf{x}_1^\top \mathbf{x}_2 & \cdots & \mathbf{x}_1^\top \mathbf{x}_t \\ \mathbf{x}_2^\top \mathbf{x}_1 & \mathbf{x}_2^\top \mathbf{x}_2 & \cdots & \mathbf{x}_2^\top \mathbf{x}_t \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_t^\top \mathbf{x}_1 & \mathbf{x}_t^\top \mathbf{x}_2 & \cdots & \mathbf{x}_t^\top \mathbf{x}_t \end{bmatrix} \qquad \mathbf{Y}^\top \mathbf{Y} = \begin{bmatrix} \mathbf{y}_1^\top \mathbf{y}_1 & \mathbf{y}_1^\top \mathbf{y}_2 & \cdots & \mathbf{y}_1^\top \mathbf{y}_t \\ \mathbf{y}_2^\top \mathbf{y}_1 & \mathbf{y}_2^\top \mathbf{y}_2 & \cdots & \mathbf{y}_2^\top \mathbf{y}_t \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{y}_t^\top \mathbf{y}_1 & \mathbf{y}_t^\top \mathbf{y}_2 & \cdots & \mathbf{y}_t^\top \mathbf{y}_t \end{bmatrix}$$

$$\boxed{\mathbf{X}^\top \mathbf{X} \approx \mathbf{Y}^\top \mathbf{Y}}$$

cf. Multidimensional scaling (MDS)

Similarity matching approach

Experiment: cognitively similar sensory stimuli evoke similar neural activity patterns in IT



fMRI: Kriegeskorte et al
e-phys: Kiani et al

Similarity matrix of neural activity patterns in human inferior temporal (IT) cortex

Objective functions for similarity matching

$$\min_{\boldsymbol{Y}} \left\| \boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{Y}^\top \boldsymbol{Y} \right\|_F^2 \qquad \boldsymbol{X} \in \mathbb{R}^{n \times T}$$

$$\min_{\substack{\boldsymbol{Y} \in \mathbb{R}^{k \times T} \\ k<n}} \left\| \boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{Y}^\top \boldsymbol{Y} \right\|_F^2$$

Linear dimensionality reduction

$$\min_{\boldsymbol{Y} \in \mathbb{R}_{\geq 0}^{k \times T}} \left\| \boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{Y}^\top \boldsymbol{Y} \right\|_F^2$$

Soft clustering and feature extraction

Offline similarity matching

Problem: $\min_{\boldsymbol{Y}\in\mathbb{R}^{k\times T}} \left\| \boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{Y}^\top \boldsymbol{Y} \right\|_F^2$

Solution: $\boldsymbol{Y}$ is a projection of $\boldsymbol{X}$ onto its k-dimensional principal subspace

Mardia '80

Proof sketch:

$$\boldsymbol{X} = \boldsymbol{U}_X \boldsymbol{S}_X \boldsymbol{V}_X^\top, \qquad \boldsymbol{Y} = \boldsymbol{U}_Y \boldsymbol{S}_Y \boldsymbol{V}_Y^\top$$

$$\left\| \boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{Y}^\top \boldsymbol{Y} \right\|_F^2 = \left\| \boldsymbol{V}_X \boldsymbol{S}_X^2 \boldsymbol{V}_X^\top - \boldsymbol{V}_Y \boldsymbol{S}_Y^2 \boldsymbol{V}_Y^\top \right\|_F^2$$

$$\boldsymbol{V}_Y = \boldsymbol{V}_{X(K)} \Rightarrow \left\| \boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{Y}^\top \boldsymbol{Y} \right\|_F^2 = \left\| \boldsymbol{S}_X^2 - \boldsymbol{S}_Y^2 \right\|_F^2$$

$$\boldsymbol{S}_Y = \boldsymbol{S}_{X(K)}$$

$$\boldsymbol{Y} = \boldsymbol{U}_{Y(K)} \boldsymbol{S}_{X(K)} \boldsymbol{V}_{X(K)}^\top$$

Online dimensionality reduction

$$\min_{\boldsymbol{Y}\in\mathbb{R}^{k\times T}} \left\| \boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{Y}^\top \boldsymbol{Y} \right\|_\mathrm{F}^2$$

$$0 = \frac{\partial}{\partial \boldsymbol{Y}} \left\| \boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{X}^\top \boldsymbol{X} \right\|_\mathrm{F}^2$$

$$= \frac{\partial}{\partial \boldsymbol{Y}} \mathrm{Tr}\left(\boldsymbol{Y}^\top \boldsymbol{Y} \boldsymbol{Y}^\top \boldsymbol{Y} - 2\boldsymbol{X}^\top \boldsymbol{X} \boldsymbol{Y}^\top \boldsymbol{Y} \right)$$

$$= 4\left(\boldsymbol{Y}\boldsymbol{Y}^\top \boldsymbol{Y} - \boldsymbol{Y}\boldsymbol{X}^\top \boldsymbol{X} \right)$$

Group:

$$\left[\boldsymbol{Y}\boldsymbol{Y}^\top\right]\boldsymbol{Y} = \left[\boldsymbol{Y}\boldsymbol{X}^\top\right]\boldsymbol{X}$$

Online,

$$\boldsymbol{y}_t \leftarrow \arg\min_{\boldsymbol{y}_t\in\mathbb{R}^k} \left\| \boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{Y}^\top \boldsymbol{Y} \right\|_\mathrm{F}^2 : \quad \left[\boldsymbol{Y}\boldsymbol{Y}^\top\right]\boldsymbol{y}_t = \left[\boldsymbol{Y}\boldsymbol{X}^\top\right]\boldsymbol{x}_t$$

Large-*t* limit:

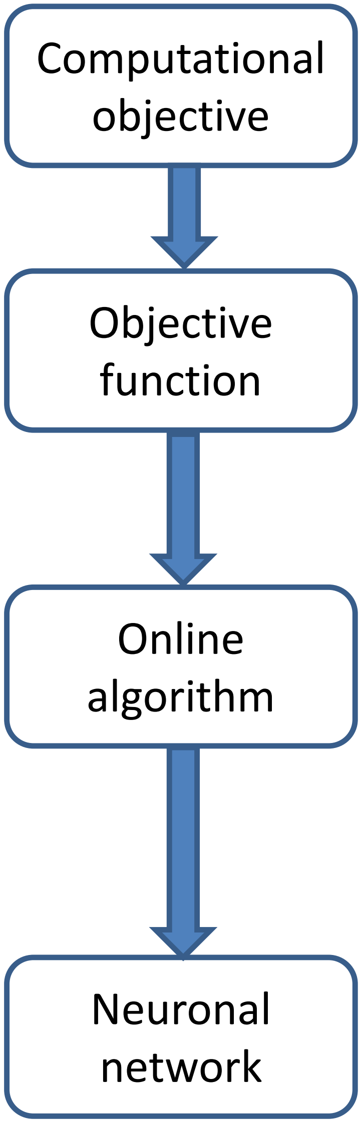
$$\left[\boldsymbol{Y}\boldsymbol{Y}^\top\right]_{t-1}\boldsymbol{y}_t = \left[\boldsymbol{Y}\boldsymbol{X}^\top\right]_{t-1}\boldsymbol{x}_t$$

$$y_{i,t} \leftarrow \boldsymbol{W}_{i,t}^{\boldsymbol{YX}}\boldsymbol{x}_t - \boldsymbol{W}_{i,t}^{\boldsymbol{YY}}\boldsymbol{y}_t \qquad W_{i,j,t}^{YX} = \frac{\sum_{\tau=1}^{t-1} y_{i,\tau}x_{j,\tau}}{\sum_{\tau=1}^{t-1} y_{i,\tau}^2};$$

$$W_{i,j\neq i,t}^{YY} = \frac{\sum_{\tau=1}^{t-1} y_{i,\tau}y_{j,\tau}}{\sum_{\tau=1}^{t-1} y_{i,\tau}^2}; \quad W_{i,i,t}^{YY} = 0$$

Pehlevan, Hu, & Chklovskii (2015)

Deriving a neural circuit



Linear dimensionality reduction

$$\min_{\boldsymbol{Y}\in\mathbb{R}^{k\times T}} \left\|\boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{Y}^\top \boldsymbol{Y}\right\|_F^2$$

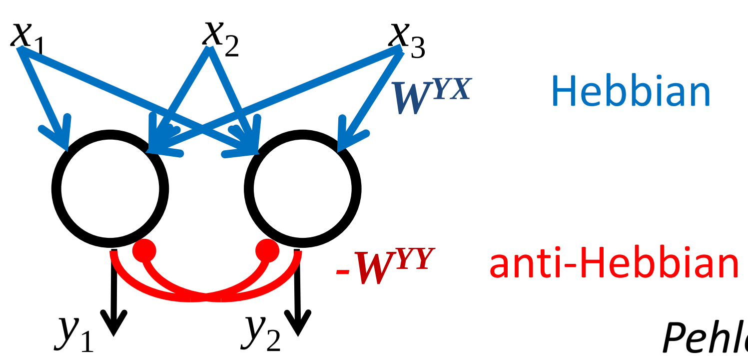
$$\boldsymbol{y}_t \leftarrow \arg\min_{\boldsymbol{y}_t\in\mathbb{R}^k} \left\|\boldsymbol{X}_t^\top \boldsymbol{X}_t - \boldsymbol{Y}_t^\top \boldsymbol{Y}_t\right\|_F^2$$

$$\boldsymbol{y}_t \leftarrow \boldsymbol{W}_t^{\boldsymbol{YX}} \boldsymbol{x}_t - \boldsymbol{W}_t^{\boldsymbol{YY}} \boldsymbol{y}_t$$

$$W^{YX}_{i,j,t} = \frac{\sum_{\tau=1}^{t-1} y_{i,\tau} x_{j,\tau}}{\sum_{\tau=1}^{t-1} y_{i,\tau}^2}$$

$$W^{YX}_{i,j,t+1} \leftarrow W^{YX}_{i,j,t} + y_{i,t}\left(x_{j,t} - W^{YX}_{i,j,t} y_{i,t}\right) \Big/ \sum_\tau y_{i,\tau}^2$$

$$W^{YY}_{i,j\neq i,t+1} \leftarrow W^{YY}_{i,j,t} + y_{i,t}\left(y_{j,t} - W^{YY}_{i,j,t} y_{i,t}\right) \Big/ \sum_\tau y_{i,\tau}^2$$



Pehlevan, Hu, & Chklovskii (2015)

Online similarity matching learns principal subspace on synthetic data

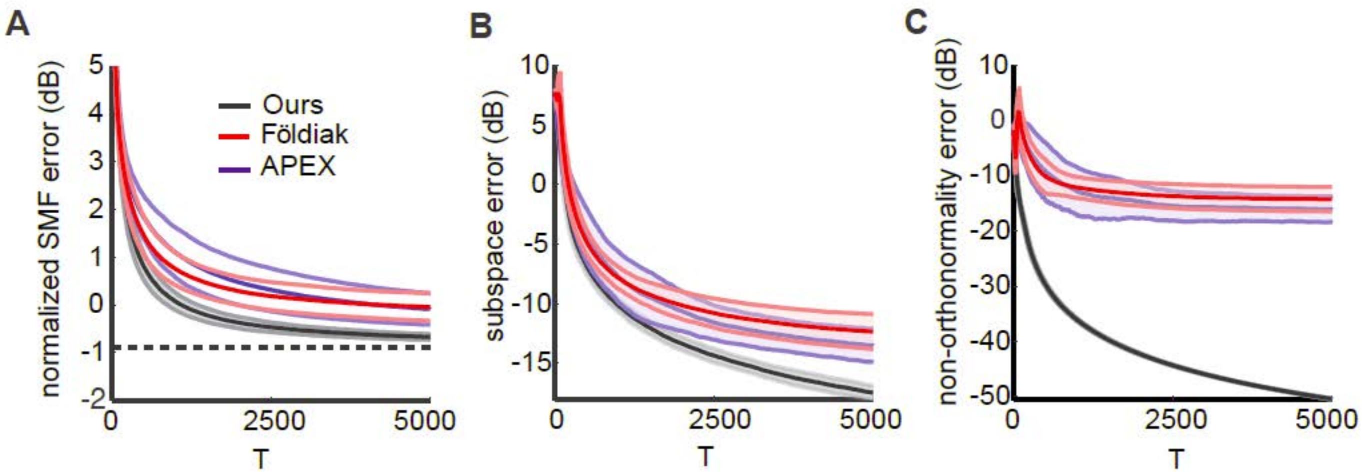


Fig. 2: Performance of the neural network compared with existing algorithms.

A. Normalized SMF error:

$$\frac{1}{T^2}\left\|\mathbf{X}^\top\mathbf{X} - \mathbf{Y}^\top\mathbf{Y}\right\|_F^2$$

B. Subspace error:

$$\mathbf{F}_T = (I_k + W_T^{YY})^{-1}\, W_T^{YX}$$

$$\left\|\mathbf{F}_T^\top\mathbf{F}_T - \mathbf{V}^\top\mathbf{V}\right\|_F^2$$

C. Non-orthonormality error:

$$\left\|\mathbf{F}_T\mathbf{F}_T^\top - \mathbf{I}_m\right\|_F^2$$

Pehlevan, Hu, & Chklovskii (2015)

Deriving neural circuits from first principles

- Representation/decoding approach
  - Single-neuron PCA
  - Soft-thresholding neuron
  - Multiple neurons
- Similarity matching approach
  - Linear dimensionality reduction
  - Nonnegative output

Objective functions for similarity matching

$$\min_{\boldsymbol{Y}} \left\| \boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{Y}^\top \boldsymbol{Y} \right\|_F^2 \qquad \boldsymbol{X} \in \mathbb{R}^{n \times T}$$

$$\min_{\substack{\boldsymbol{Y} \in \mathbb{R}^{k \times T} \\ k<n}} \left\| \boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{Y}^\top \boldsymbol{Y} \right\|_F^2$$

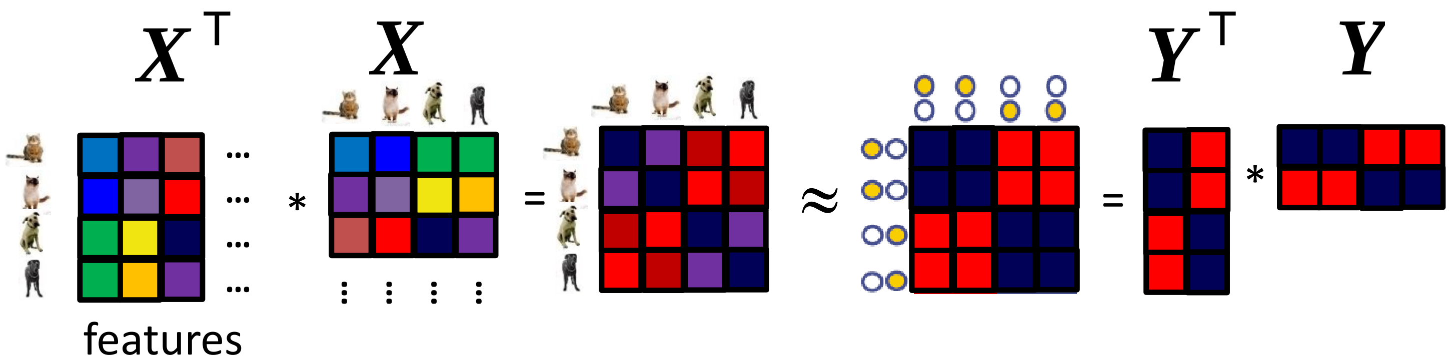
Linear dimensionality reduction

$$\min_{\boldsymbol{Y} \in \mathbb{R}^{k \times T}_{\geq 0}} \left\| \boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{Y}^\top \boldsymbol{Y} \right\|_F^2$$

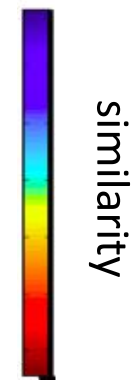
Soft clustering and feature extraction

Clustering with nonnegative similarity matching

$$\boldsymbol{X}^{\top}\boldsymbol{X} \approx \boldsymbol{Y}^{\top}\boldsymbol{Y}$$



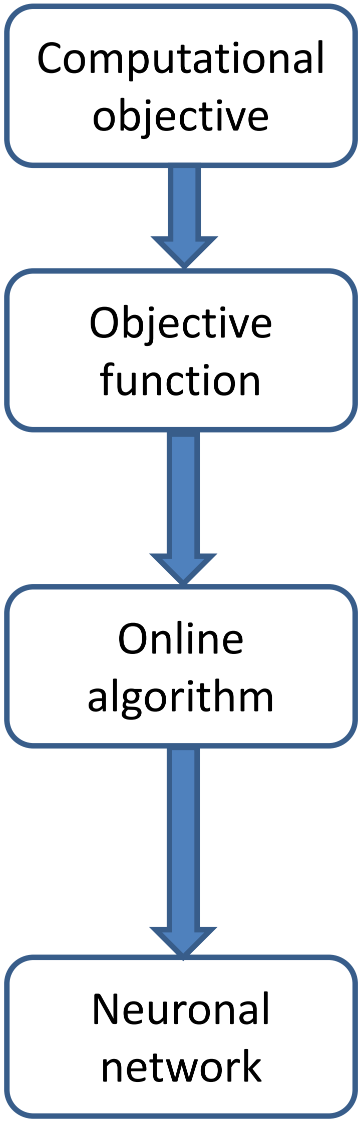
$$\min_{\boldsymbol{Y} \in \mathbb{R}_{\geq 0}^{k \times T}} \left\| \boldsymbol{X}^{\top}\boldsymbol{X} - \boldsymbol{Y}^{\top}\boldsymbol{Y} \right\|_{\mathrm{F}}^{2}$$



Nonnegative similarity matching objective function is a relaxation of *k*-means clustering

SNMF *(Ding et al, 2005)*

Deriving a neural circuit



Soft clustering and feature extraction

$$\min_{\boldsymbol{Y} \in \mathbb{R}^{k\times T}_{\geq 0}} \left\| \boldsymbol{X}^{\top}\boldsymbol{X} - \boldsymbol{Y}^{\top}\boldsymbol{Y} \right\|_F^2$$

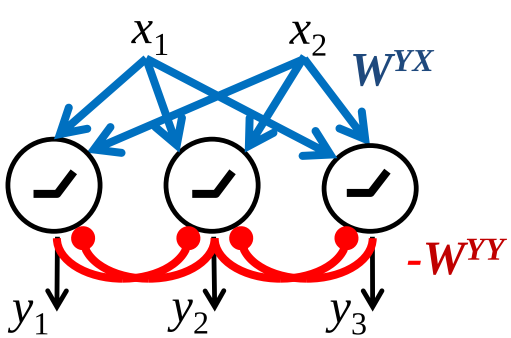
$$\boldsymbol{y}_t \leftarrow \arg\min_{\boldsymbol{y}_t \in \mathbb{R}^{k}_{\geq 0}} \left\| \boldsymbol{X}_t^{\top}\boldsymbol{X}_t - \boldsymbol{Y}_t^{\top}\boldsymbol{Y}_t \right\|_{\mathrm{F}}^2$$

$$\boldsymbol{y}_t \leftarrow \max\left(\boldsymbol{W}_t^{YX}\boldsymbol{x}_t - \boldsymbol{W}_t^{YY}\boldsymbol{y}_t, 0 \right)$$

$$W^{YX}_{i,j,t} = \frac{\sum_{\tau=1}^{t-1} y_{i,\tau} x_{j,\tau}}{\sum_{\tau=1}^{t-1} y_{i,\tau}^2}$$

$$W^{YX}_{i,j,t+1} \leftarrow W^{YX}_{i,j,t} + y_{i,t}\left(x_{j,t} - W^{YX}_{i,j,t} y_{i,t} \right) \Big/ \sum_{\tau} y_{i,\tau}^2$$

$$W^{YY}_{i,j\neq i,t+1} \leftarrow W^{YY}_{i,j,t} + y_{i,t}\left(y_{j,t} - W^{YY}_{i,j,t} y_{i,t} \right) \Big/ \sum_{\tau} y_{i,\tau}^2$$

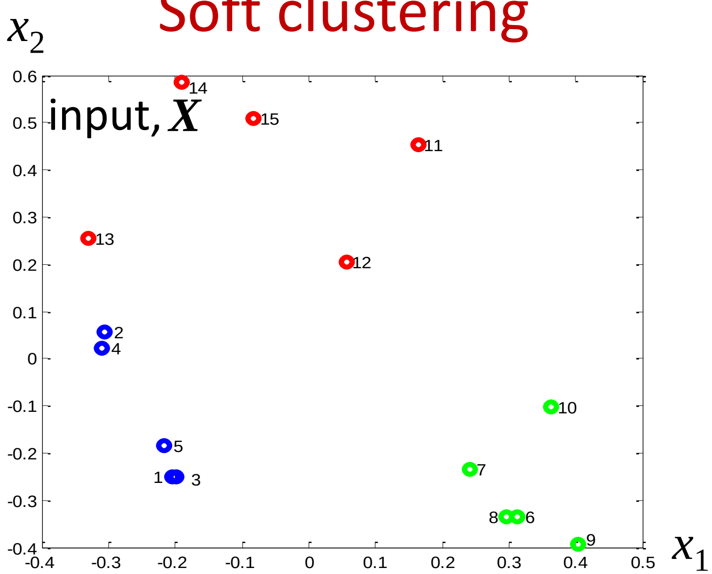


Pehlevan & Chklovskii (2014)

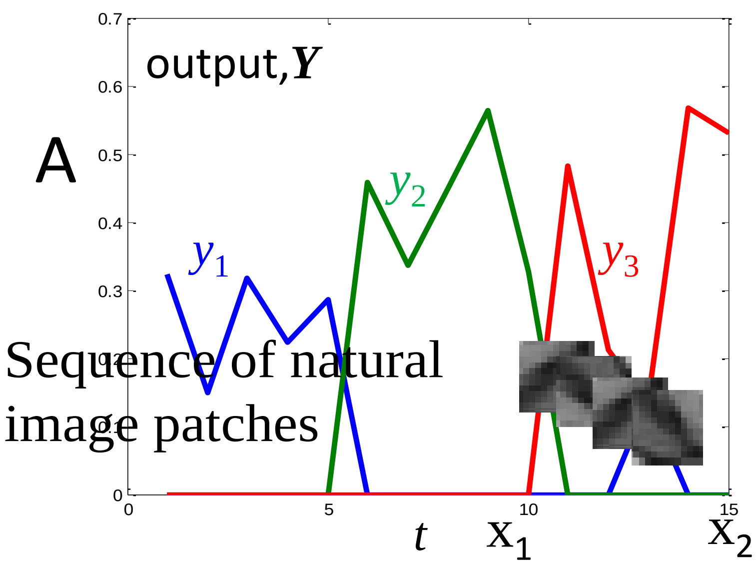
Online similarity matching solves computational objectives

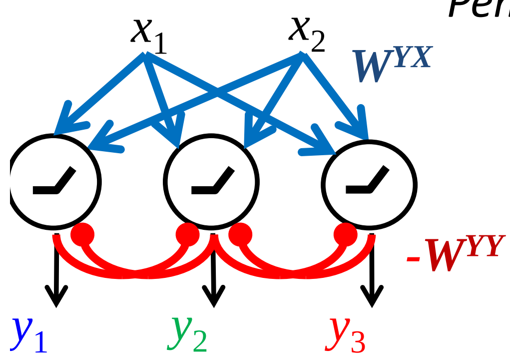
Pehlevan & Chklovskii (2014)

Soft clustering



A





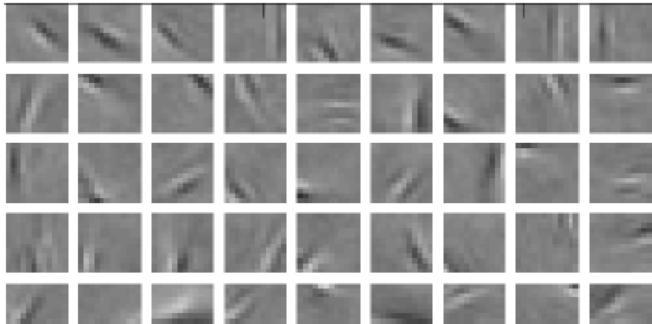
Sparse feature extraction

natural images, X

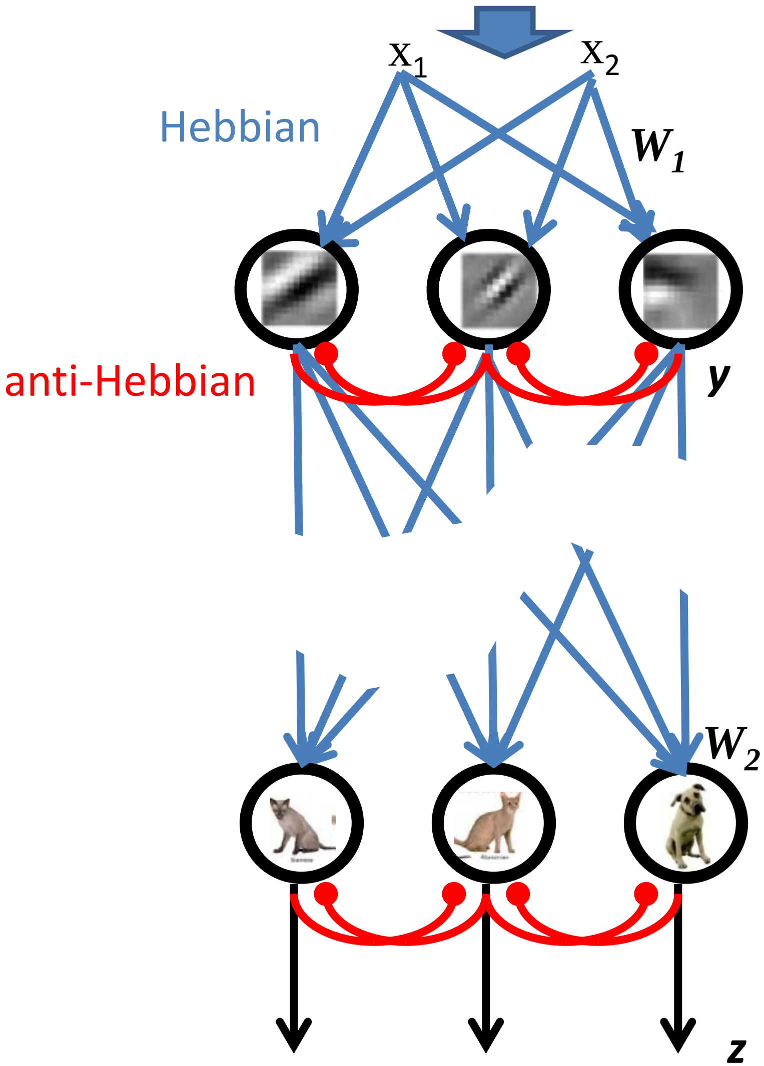
B

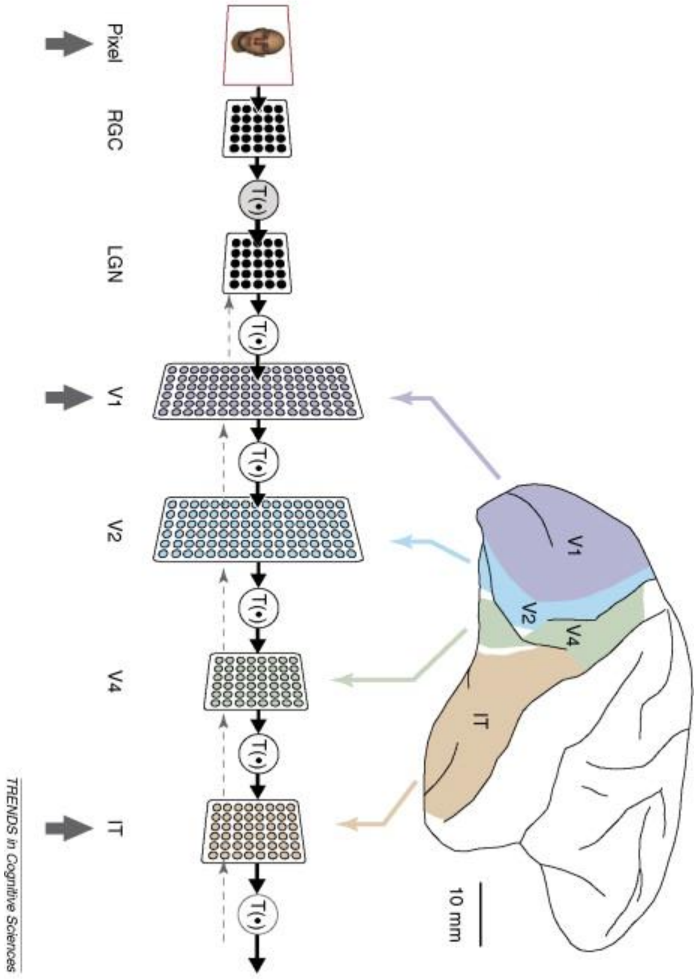
Learned feed-forward

dictionary, W^{YX}



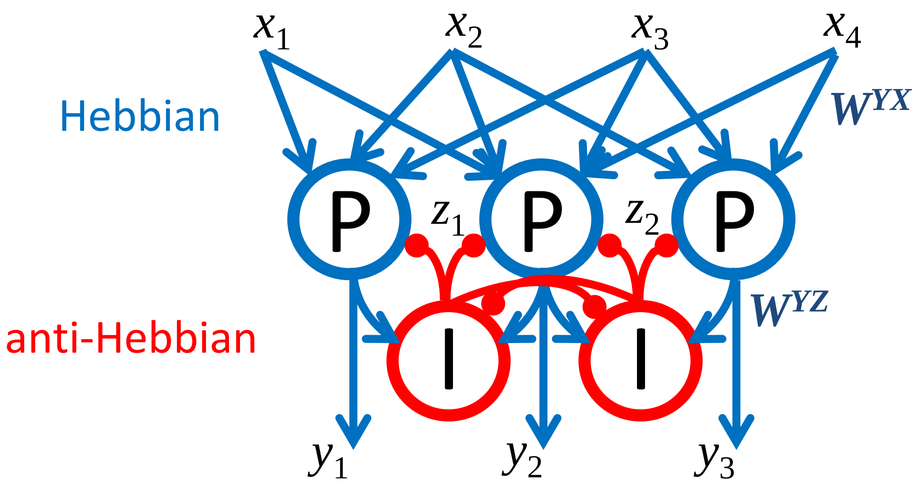
Deep network

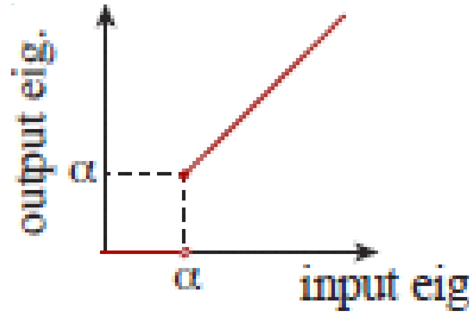




Inhibitory interneurons hard-threshold covariance eigenvalues

$$\min_{\boldsymbol{Y}\in\mathbb{R}^{k\times T}} \max_{\boldsymbol{Z}\in\mathbb{R}^{l\times T}} \left[\left\| \boldsymbol{X}^\top \boldsymbol{X} - \boldsymbol{Y}^\top \boldsymbol{Y} \right\|_F^2 - \left\| \boldsymbol{Y}^\top \boldsymbol{Y} - \boldsymbol{Z}^\top \boldsymbol{Z} - \alpha t \boldsymbol{I}_t \right\|_F^2 \right]$$





Pehlevan & Chklovskii, 2015

Deriving neural circuits from first principles

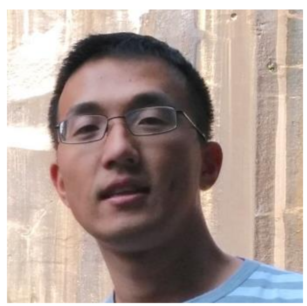
- Representation/decoding approach
  - Single-neuron PCA
  - Soft-thresholding neuron
  - Multiple neurons
- Similarity matching approach
  - Linear dimensionality reduction
  - Nonnegative output

Acknowledgements

Cengiz Pehlevan

Tao Hu -> Texas A&M

Summer students



Yuansi Chen

Reza Abbasi-Asl

Positions available in the Neuroscience Group
Simons Center for Data Analysis, New York City