Biologically plausible online PCA without recurrent neural dynamics

Victor Minden[1], Dmitri Chklovskii[1,2], Cengiz Pehlevan[1]

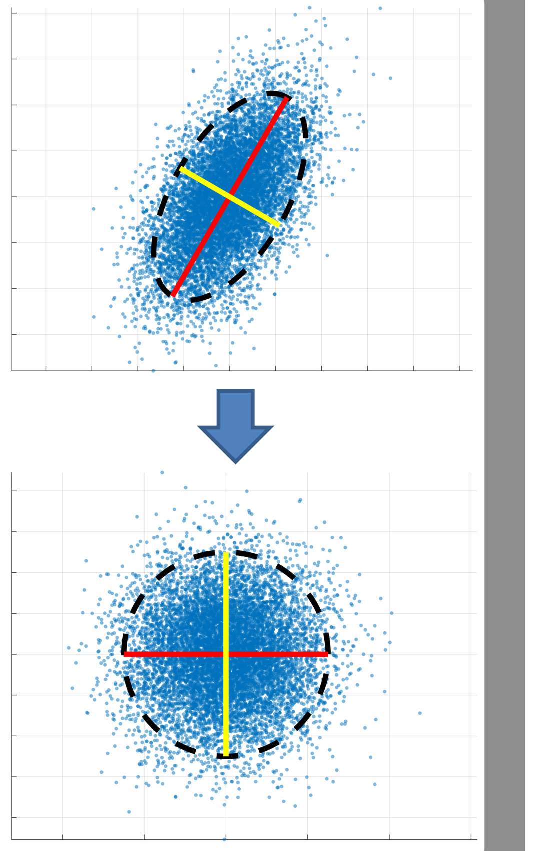
1. Center for Computational Biology, Flatiron Institute, Simons Foundation   2. NYU Medical Center

Summary

Starting from a similarity matching objective function, we derive an online algorithm for multi-component Principal Component Analysis (PCA) with local Hebbian learning rules in which synaptic plasticity and neural dynamics both occur on the same time scale.

Background

Early stages of information processing in the brain necessarily involve some form of dimensionality reduction. It has been hypothesized that this may be performed using PCA, which gives the optimal linear subspace for reducing dimensionality of input sensory data (and optionally whitening the result).



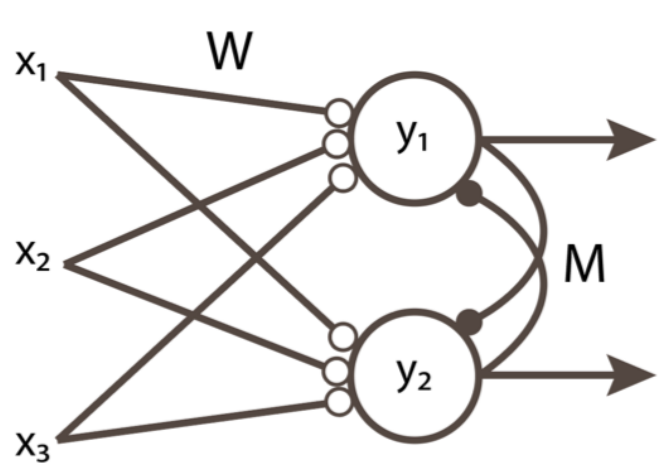
Method	Non-hierarchical	Local updates	Non-recurrent	Normative
Our approach	✓	✓	✓	✓
SNL [5]	✓		✓	✓
GHA [7]			✓	
SGA [6]			✓	✓
Földiák [3]	✓	✓		
APEX [4]		✓	✓	

Neural PCA from similarity matching [1,2]

- For $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_T]$ find smaller $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_T]$ with whitened components by solving constrained similarity matching.
$$\min_{\mathbf{Y}\in\mathbb{R}^{k\times T}} \|\mathbf{X}^\top\mathbf{X} - \mathbf{Y}^\top\mathbf{Y}\|_F^2 \quad \text{s.t.} \quad \mathbf{Y}\mathbf{Y}^T = \mathbf{I}$$
- Expand norm and introduce feed-forward synaptic weights $\mathbf{W}$.
$$\min_{\mathbf{W}\in\mathbb{R}^{k\times n}, \mathbf{Y}\in\mathbb{R}^{k\times T}} -4\,\mathrm{Tr}\left(\mathbf{X}^\top\mathbf{W}^\top\mathbf{Y}\right) + \mathrm{Tr}\left(\mathbf{W}^\top\mathbf{W}\right) \quad \text{s.t.} \quad \mathbf{Y}\mathbf{Y}^T = \mathbf{I}$$
- Obtain equivalent min-max problem by writing Lagrangian in terms of lateral synaptic weights $\mathbf{M}$.
$$\min_{\mathbf{W}\in\mathbb{R}^{k\times n}} \max_{\mathbf{M}\in\mathbb{R}^{k\times k}} \min_{\mathbf{Y}\in\mathbb{R}^{k\times T}} -4\,\mathrm{Tr}\left(\mathbf{X}^\top\mathbf{W}^\top\mathbf{Y}\right) + 2\,\mathrm{Tr}\left(\mathbf{W}^\top\mathbf{W}\right) + \mathrm{Tr}\left(\mathbf{M}[\mathbf{Y}\mathbf{Y}^T - \mathbf{I}]\right)$$

Neural dynamics
$$\mathbf{y}_t := \mathbf{M}_t^{-1}\mathbf{W}_t\mathbf{x}_t$$

Synaptic plasticity
$$\mathbf{W}_{t+1} := \mathbf{W}_t + \eta_{\mathbf{W},t}\left(\mathbf{y}_t\mathbf{x}_t^\top - \mathbf{W}_t\right)$$
$$\mathbf{M}_{t+1} := \mathbf{M}_t + \eta_{\mathbf{M},t}\left(\mathbf{y}_t\mathbf{y}_t^\top - \mathbf{I}_t\right)$$



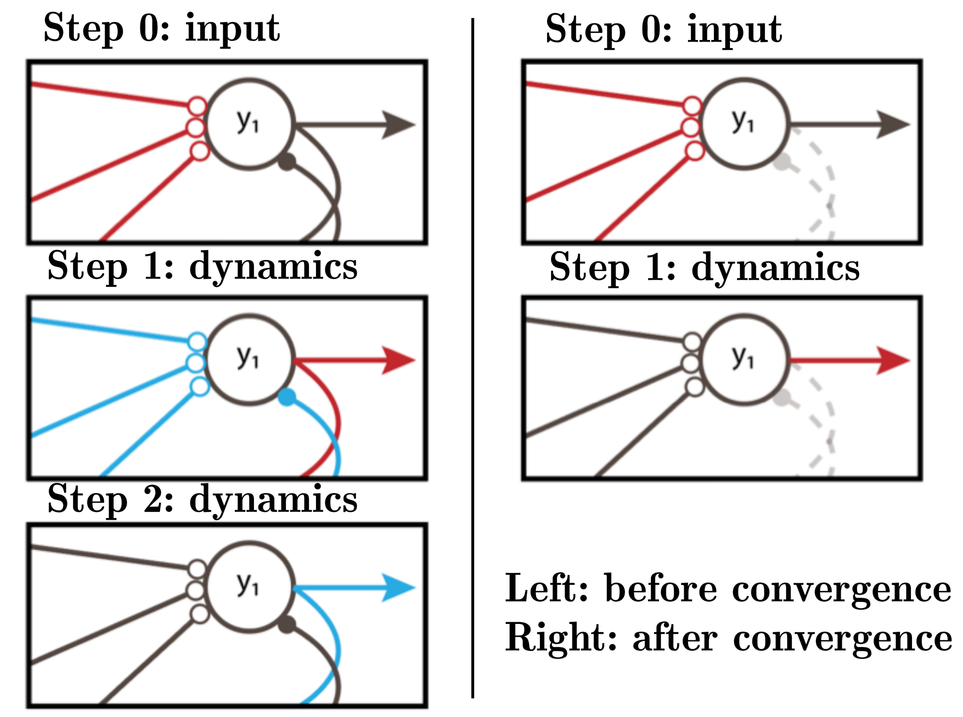
- Computing the neural dynamics requires matrix inversion, which is not local. To implement biologically, **need to apply matrix inverse using recurrent neural dynamics** on a substantially faster time scale compared to time scale of input.
- Synaptic plasticity only occurs after recurrent neural dynamics are resolved, so **plasticity is not occurring during most of the computation.**
- **Components of y are not projections onto individual principal vectors** (due to rotational degeneracy of solution). This makes decoding difficult downstream.

Neural PCA based on similarity matching without recurrent neural dynamics

- To remove recurrent dynamics and degeneracy of solution, start from a modified objective. For diagonal $\mathbf{\Lambda} \in \mathbb{R}^{k\times k}$ with positive decreasing diagonal entries start with
$$\min_{\mathbf{Y}\in\mathbb{R}^{k\times T}} \|\mathbf{X}^\top\mathbf{X} - \mathbf{Y}^\top\mathbf{Y}\|_F^2 \quad \text{s.t.} \quad \mathbf{Y}\mathbf{Y}^T = \mathbf{\Lambda}$$
- Follow same steps as before to get min-max problem
$$\min_{\mathbf{W}\in\mathbb{R}^{k\times n}} \max_{\mathbf{M}\in\mathbb{R}^{k\times k}} \min_{\mathbf{Y}\in\mathbb{R}^{k\times T}} -4\,\mathrm{Tr}\left(\mathbf{X}^\top\mathbf{W}^\top\mathbf{Y}\right) + 2\,\mathrm{Tr}\left(\mathbf{W}^\top\mathbf{W}\right) + \mathrm{Tr}\left(\mathbf{M}[\mathbf{Y}\mathbf{Y}^T - \mathbf{\Lambda}]\right)$$
- Optimal $\mathbf{M}$ is diagonal, so we split into diagonal-plus-off-diagonal parts and for $\mathbf{M}=\mathbf{M}_d+\mathbf{M}_o$ with $\mathbf{M}$ *close to diagonal* we have from Taylor series that
$$\mathbf{M}^{-1}\mathbf{W}\mathbf{x} \approx \mathbf{M}_d^{-1}\mathbf{W}\mathbf{x} - \mathbf{M}_d^{-1}\mathbf{M}_o\mathbf{M}_d^{-1}\mathbf{W}\mathbf{x}$$

Neural dynamics (two-step)
$$\tilde{\mathbf{y}}_t := \mathbf{M}_{d,t}^{-1}\mathbf{W}_t\mathbf{x}_t$$
$$\mathbf{y}_t := \mathbf{M}_{d,t}^{-1}\mathbf{W}_t\mathbf{x}_t - \mathbf{M}_{d,t}^{-1}\mathbf{M}_{o,t}\tilde{\mathbf{y}}_t$$

Synaptic plasticity
$$\mathbf{W}_{t+1} := \mathbf{W}_t + \eta_{\mathbf{W},t}\left(\mathbf{y}_t\mathbf{x}_t^\top - \mathbf{W}_t\right)$$
$$\mathbf{M}_{t+1} := \mathbf{M}_t + \eta_{\mathbf{M},t}\left(\mathbf{y}_t\mathbf{y}_t^\top - \mathbf{\Lambda}_t\right)$$



Left: before convergence
Right: after convergence

- Note that in component form we have, e.g., $[\tilde{\mathbf{y}}_t]_i = \frac{1}{[\mathbf{M}_t]_{ii}}\sum_j [\mathbf{W}_t]_{ij}[\mathbf{x}_t]_j$, requiring only division by local variable $[\mathbf{M}_t]_{ii}$ (**no real matrix inversion**).
- Instead of recurrent dynamics, during training we have **two-step feed-forward dynamics** with one step of within-layer communication.
- **Synaptic plasticity, neural dynamics, and input variation all occur on the same time scale**
- **Components of y are projections onto individual principal vectors** (fixed rotational degeneracy)

Discussion and extensions

- By starting with a different objective function, we can obtain a similar algorithm for the case of unconstrained similarity matching [1], i.e., projection onto the principal subspace with no whitening.
$$\min_{\mathbf{Y}\in\mathbb{R}^{k\times T}} \|\mathbf{X}^\top\mathbf{X} - \mathbf{Y}^\top\mathbf{Y}\|_F^2 \quad \Rightarrow \quad \min_{\mathbf{Y}\in\mathbb{R}^{k\times T}} -2\,\mathrm{Tr}\left(\mathbf{X}^\top\mathbf{X}\mathbf{Y}^\top\mathbf{Y}\right) + \mathrm{Tr}\left(\mathbf{Y}^\top\mathbf{\Lambda}^{-1}\mathbf{Y}\mathbf{Y}^\top\mathbf{\Lambda}^{-1}\mathbf{Y}\right)$$
- Biologically, the model predicts that lateral synaptic weights could be weak compared to feed-forward weights, as long as the network is operating near equilibrium.
- To prove convergence of stochastic gradient dynamics in online algorithm, need more than offline theorem – possible that recent results on Generative Adversarial Networks (GANs) might apply.

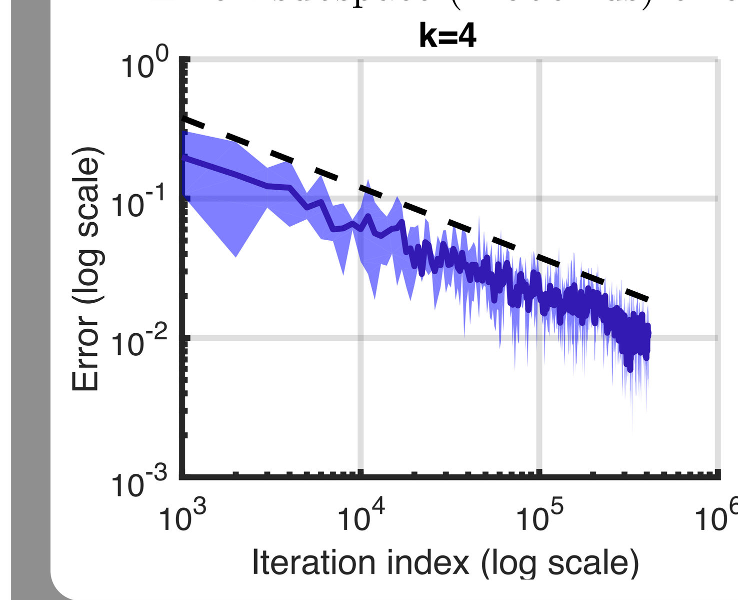
Theoretical results (offline)

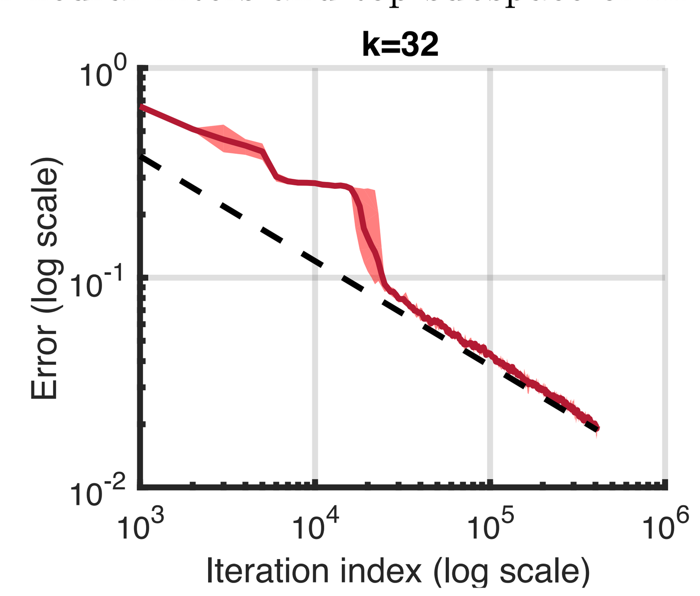
Theorem (informal). Consider the continuous time, offline dynamical system
$$\mathbf{Y}(t) = \left[\mathbf{I} - \mathbf{M}_d^{-1}(t)\mathbf{M}_o(t)\right]\mathbf{M}_d^{-1}(t)\mathbf{W}(t)\mathbf{X}$$
$$\frac{d\mathbf{W}(t)}{dt} = \mathbf{Y}(t)\mathbf{X}^\top - \mathbf{W}(t)$$
$$\tau\frac{d\mathbf{M}(t)}{dt} = \mathbf{Y}(t)\mathbf{Y}(t)^\top - \mathbf{\Lambda}$$
and assume the input data $\mathbf{X}$ has unique singular values. Then, if τ is small enough we have:

- At any fixed point $\mathbf{M}$ is diagonal and each row of $\mathbf{Y}$ is given by projecting $\mathbf{X}$ onto a singular vector of $\mathbf{X}$ and then normalizing.
- At any stable fixed point, row i of $\mathbf{Y}$ corresponds to scaled projection onto the i-th principal component of $\mathbf{X}$.

Numerical results (online)

- Data parameters: n = 160, T = 409600, k nontrivial principal values linearly spaced between 1 and 0.05, remaining singular values are all 0.0001
- Error: subspace (Frobenius) error between neural filters and top subspace of $\mathbf{X}$.





References

1. C. Pehlevan, T. Hu, and D. B. Chklovskii. "A Hebbian/Anti-Hebbian Neural Network for Linear Subspace Learning: A Derivation from Multidimensional Scaling of Streaming Data". In: Neural Computation 27.7 (2015), pp. 1461– 1495.
2. C. Pehlevan, A. M. Sengupta, and D. B. Chklovskii. "Why Do Similarity Matching Objectives Lead to Hebbian/Anti-Hebbian Networks?". In: Neural Computation 30.1 (2018), pp. 84– 124.
3. P. Földiák. "Adaptive network for optimal linear feature extraction". In: International 1989 Joint Conference on Neural Networks. 1989, 401–405 vol.1.
4. S. Y. Kung and K. I. Diamantaras. "A neural network learning algorithm for adaptive principal component extraction (APEX)". In: International Conference on Acoustics, Speech, and Signal Processing. 1990, 861–864 vol.2.
5. E. Oja. "Neural networks, principal components, and subspaces". In: International Journal of Neural Systems 01.01 (1989), pp. 61–68.
6. E. Oja. "Principal components, minor components, and linear neural networks". In: Neural Networks 5.6 (1992), pp. 927 –935.
7. T. D. Sanger. "Optimal unsupervised learning in a single-layer linear feedforward neural network". In: Neural Networks 2.6 (1989), pp. 459–473.
8. R. Linsker. "Improved local learning rule for information maximization and related applications". In: Neural Networks 18.3 (2005), pp. 261–265