

Estimation of two-layer statistical model of natural images using score matching leads to complex cell properties

Urs Köster¹, Aapo Hyvarinen¹, and Silke Dreissigacker²

¹University of Helsinki, ²Universität Göttingen

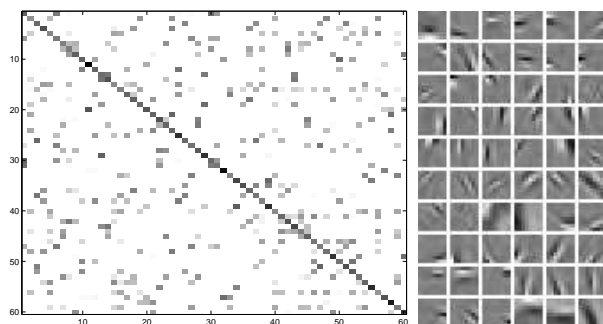
We present a two-layer model of natural images that produces outputs similar to Complex Cells. Unsupervised learning using the novel method *Score Matching (SM)*[1] is used to estimate the model. In contrast to our previous work, both layers are freely learned from the data. The second layer has a very sparse connectivity as predicted by the Hubel and Wiesel Model of primary visual cortex, and describes local dependencies between first layer features.

Score Matching is a locally consistent estimation method for non-normalized statistical models. By making use of the Score Function (the gradient of the log-probability), an estimation without knowledge of the partition function is possible. This is in contrast to Monte Carlo methods such as Contrastive Divergence, which rely on slow random sampling, or methods based on approximations. We apply the method to a model of the form

$$\log p(\mathbf{x}|\mathbf{W}, \mathbf{V}) = \sum_i \sqrt{\mathbf{v}_i (\mathbf{W}\mathbf{x})^2} - \log Z \quad (1)$$

where x is a data vector, \mathbf{V} and \mathbf{W} are weight matrices and Z is the intractable partition function. The nonlinearities are applied elementwise. We can compute an objective function in terms of derivatives of the score function, which is then optimized by stochastic gradient descent.

When applied to natural image patches, the first layer converges to Gabor-type features. The second layer, however exhibits a very sparse connectivity, where each row has significant activity only for a few similar first layer features. These features that are pooled by the second layer have similar size, position and orientation, but differ in spacial frequencies. This local pooling is in contrast to similar two-layer models, in which the second layer describes more global structure. Similar results are obtained for a range of overcomplete model specifications and data dimensionalities.



Left: The matrix \mathbf{V} with white representing zero and darker colors representing stronger activation. The sparse connectivity is evident. In the non-overcomplete case the number of units in the second layer is the same as in the first layer, so several output units share a similar receptive field.

Right: The features given by rows of \mathbf{W} . The features \mathbf{W} have been sorted according to subspaces as identified by \mathbf{V} .

References

- [1] A. Hyvärinen. Estimation of non-normalized statistical models using score matching. *Journal of Machine Learning Research*, 6:695709, 2005.