Beyond ICA

Chris Williams

School of Informatics, University of Edinburgh

January 15, 2018

Outline

- Independent Components aren't Independent!
- Gaussian Scale Mixtures
- Hierarchical Models for Capturing Dependencies among Sparse Components
- Independent Subspaces and Complex Cells
- Topographic ICA
- Further work

Sparse distribution of coefficients

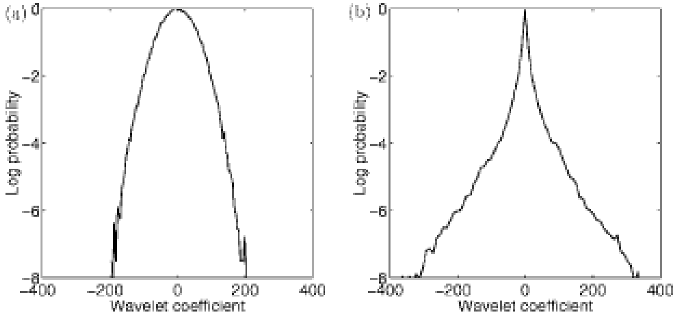
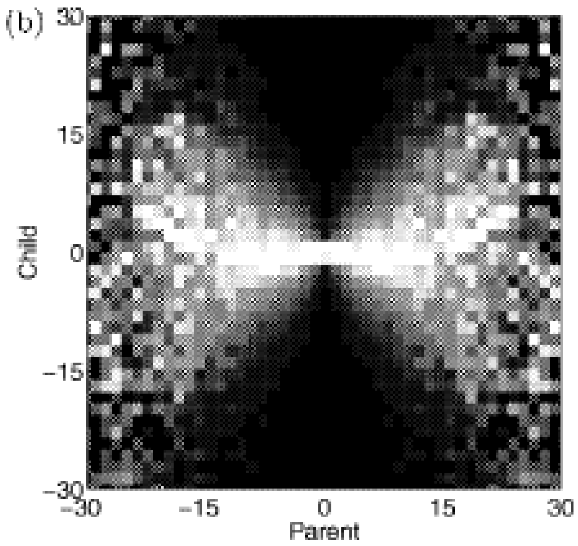


FIG. 1. Histograms of wavelet marginal distributions for (a) Gaussian noise; and (b) a typical natural image. Vertical axis gives log probability (rescaled).

[Wainwright, Simoncelli, Willsky 2001]

Independent Components aren't Independent!

- Consider two Gabor functions at the same spatial position and orientation, but with different scales
- Coarser scale is denoted "parent", finer scale is denoted child
- Plot the conditional histogram $p(\text{child}|\text{parent})$.
- "bowtie" structure for a natural image (Buccigrossi and Simoncelli, 1999)



[Wainwright, Simoncelli, Willsky 2001]

Gaussian Scale Mixtures

- A non-Gaussian distribution can be produced as a scale mixture of Gaussians

$$p(v) = \int \mathcal{N}(v; 0, \sigma^2) p(\sigma^2) d\sigma^2$$

- Example: Student-*t* distribution

$$p(v) = c(1 + v^2/\nu)^{-(\nu+1)/2}$$

  is obtained as a GSM from an inverse-χ^2 distribution on the variance

- GSM construction is not limited to univariate Gaussians. Construction

$$v_i = \sigma z_i$$

  with $z_i \sim \mathcal{N}(0, 1)$ gives

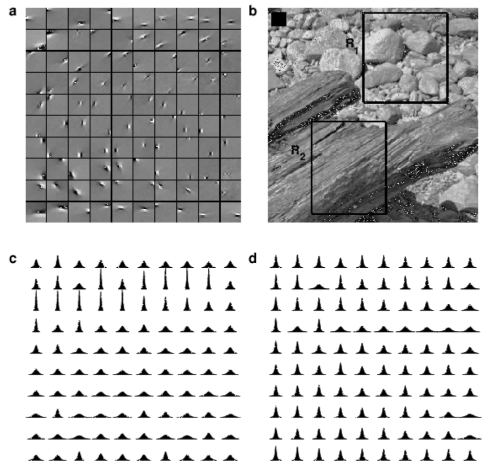
$$p(\mathbf{v}) = \int \mathcal{N}(\mathbf{v}; \mathbf{0}, \sigma^2 I) p(\sigma^2) d\sigma^2$$

- Exercise: show that $\mathrm{cov}(v_i^2, v_j^2) \geq 0$ for GSM
- Exercise: show that $\mathrm{kurt}(v_i) \geq 0$ for GSM
- The idea is that the local variance σ^2 can be different in different parts of the image. Flat surfaces have no variation (“blue-sky effect”)
- Using a $\mathrm{Gamma}(\beta/2, \beta/2)$ prior on $\tau = \sigma^{-2}$ it can be shown that

$$\mathrm{var}(v_2|v_1) = \frac{\beta + v_1^2}{\beta - 1}$$

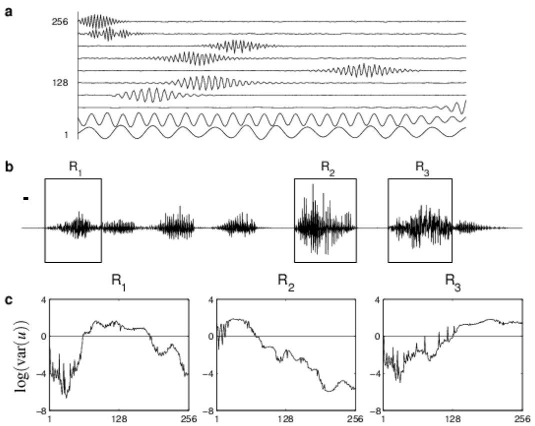
  Note that $\mathrm{var}(v_2|v_1)$ increases as a function of $|v_1|$

Non-stationary statistics of ICA coefficients 1



[Fig 1, Karklin and Lewicki (2005)]

Non-stationary statistics of ICA coefficients 2



[Fig 2, Karklin and Lewicki (2005)]

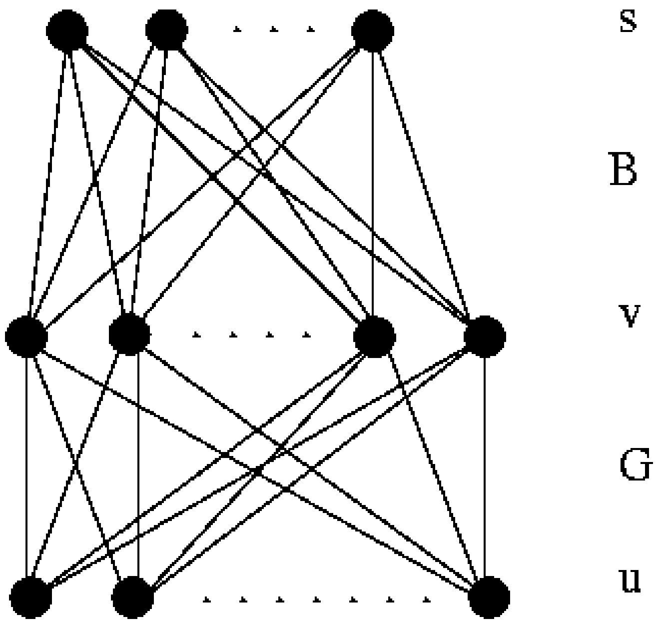
Hierarchical models for capturing dependencies among sparse components

$$v_i = \sigma_i z_i$$

$$\sigma_i = f(\sum_j b_{ij} s_j)$$

e.g.

- Wainwright, Simoncelli, Willsky (2001)
- Hyvärinen, Hoyer, Inki (2001)
- Karklin and Lewicki (2003)



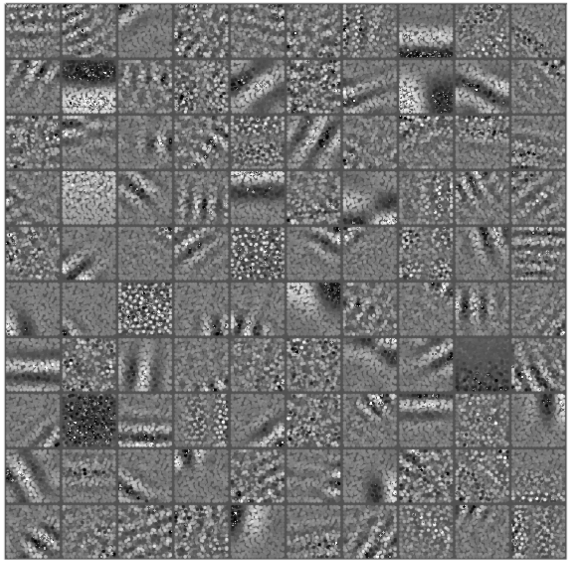
[after Lewicki, 2004]

Karklin and Lewicki (2003)

- $\sigma_i = \exp(\sum_j b_{ij} s_j)$, with a factorized Laplacian prior on **s**
- They learn G weights, then learn B weights
- Again use MAP inference

$$\hat{\mathbf{s}} = \mathrm{argmax}_{\mathbf{s}}\ p(\hat{\mathbf{v}}|B, \mathbf{s})p(\mathbf{s})$$

- B is learned by maximum likelihood
- Vision expts: 100 higher level (**s**) units, 20×20 input patches

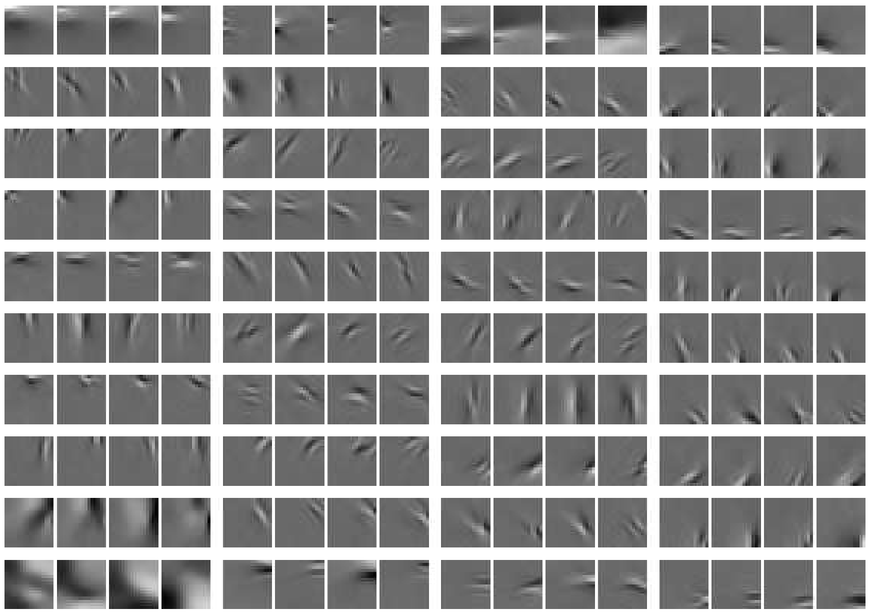


[Fig 6, Karklin and Lewicki, 2005]

Independent Subspaces

Hyvärinen and Hoyer (2000)

- The v_is are not assumed to be independent. Divide the **v** variables into groups, and allow dependencies within groups, but not between groups
- Note that v_i and v_j in a group are uncorrelated, but $\mathrm{cov}(v_i^2, v_j^2) \neq 0$
- (As viewed using K & L model) Each group of v variables has a single s parent, and all B weights from the s-parent to the group are 1. Optimize G under this model.
- ISA is described in detail in HHH chap 10



[Fig 2, Hyvärinen and Hoyer, 2000]

Models of Complex Cells

- Energy model for complex cells (Adelson and Berger, 1985). Let v_1 and v_2 be responses of even (cosine) and odd (sine) Gabor functions at same location, orientation and scale. Then

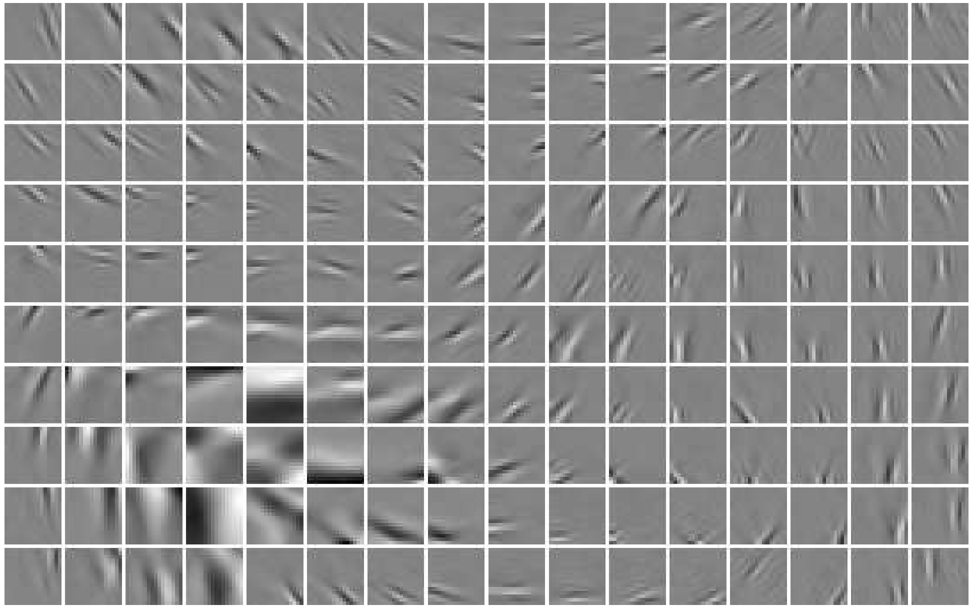
$$r = r_0 + c(v_1^2 + v_2^2)$$

- Close correspondence between the variance of the group in the generative model and r

Topographic ICA

Hyvärinen, Hoyer and Inki (2001)

- More general than independent subspaces.
- (As viewed using K & L model) *B* matrix is not learned, but is specified to have a neighbourhood structure.
- Yields a topographic arrangement of basis functions
- Argument that topographic organization would minimize wiring length in the brain (HHH §11.5)
- TICA is described in detail in HHH chap 11

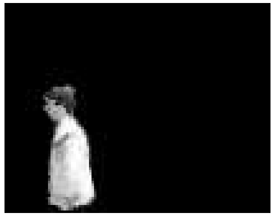


[Fig 11, Hyvärinen et al., 2005]

Further work

- Other routes to ICA: Slow feature analysis (Wiskott and Sejnowski, 2002; discussed in HHH §16.8)
- Spatio-temporal bubbles (Hyvärinen et al, 2003)
- Rao and Ballard (1999) consider correlations between filter outputs observed by higher level units that can "see" multiple patches. They observed end-stopping like effects due to feedback connections (see later lecture)

- Hyvärinen, Gutmann and Hoyer (2005) consider ICA analysis of the outputs of complex cells, and find edge-like pooling of spatial frequency channels. They predict that in V2 (or related area) cells will have optimal stimulus closer to a step edge (cf band-pass edges optimal for V1 simple and complex cells), and optimal stimulus will be more elongated. See HHH ch 12
- Linear superposition is not sufficient: occlusion



foreground (original)

mask (original)

foreground (transformed)

mask (transformed)

[Williams and Titsias, 2004]

References

- R. W. Buccigrossi and E. P. Simoncelli. Image Compression via Joint Statistical Characterization in the Wavelt Domain, *IEEE Trans. Signal Proc.*, 8(12) 1688-1701 (1999)
- A. Hyvärinen, M. Gutmann, P. O. Hoyer. Statistical model of natural stimuli predicts edge-like pooling of spatial frequency channels in V2, *BMC Neuroscience* 6(12) (2005)
- A. Hyvärinen and P. O. Hoyer. Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces, *N. Comp.* 12(7) 1705-1720 (2000)
- A. Hyvärinen, P. O. Hoyer and Inki, M. Topogrpahic Independent Components Analysis. *N. Comp.* 13(7) 1527-1558 (2001)

- A. Hyvärinen, P. O. Hoyer, J. Hurri, M. Gutmann. Statistical Models of Images and Early Vision, *Proceedings of the Int. Symposium on Adaptive Knowledge Representation and Reasoning (AKRR2005)* (2005)
- A. Hyvärinen, J. Hurri, J. Väyrynen. Bubbles: a unifying framework for low-level statistical properties of natural image sequences, *J. Optical Soc. America A*, 20(7) 1237-1252 (2003)
- A. Hyvärinen, P. O. Hoyer, J. Hurri, M. Gutmann. Statistical Models of Images and Early Vision, *Proceedings of the Int. Symposium on Adaptive Knowledge Representation and Reasoning (AKRR2005)* (2005)
- Y. Karlin and M. S. Lewicki. Learning higher-order structures in natural images. *Network: Computation in Neural Systems* 14 483-499 (2003)

- Y. Karlin and M. S. Lewicki. A Hierarchical Bayesian Model for Learning Nonlinear Statistical Regularities in Nonstationary Natural Signals *N. Comp.* 17(2) 397-424 (2005)
- R. P. N. Rao, D. H. Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects, *Nature Neuroscience* 2(1) 79-87 (1999)
- M. J. Wainwright, E. P. Simoncelli, A. S. Willsky. Random Cascades on Wavelet Trees and Their Use in Analyzing and Modeling Natural Images, *Applied and Computational Harmonic Analysis* 11, 89-123 (2001)
- L. Wiskott and T. J. Sejnowski. Slow feature analysis: unsupervised learning of invariances, *N. Comp.* 14(4), 715-770 (2002)