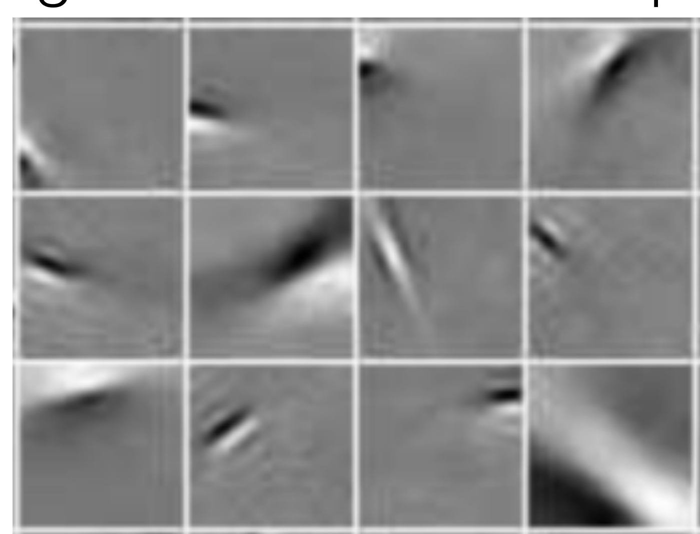
Independent Component Analysis

AI Helsinki Image Statistics Workshop — 31.10.2016

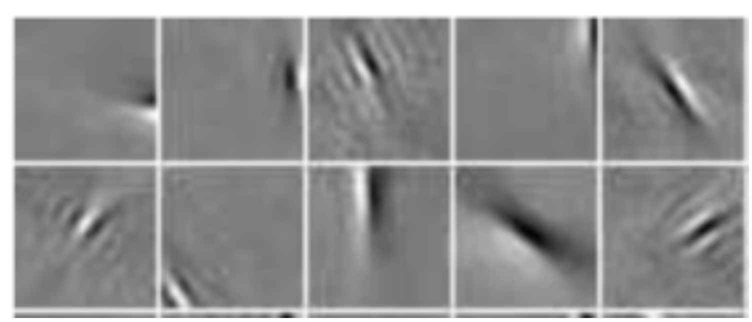


Isabeau Prémont-Schwarz

The Problem: Two Scenarios

- Sound Recording: Identify the different independent sources (different people speaking at a cocktail party)

- Natural Images: Find the most efficient sparse encoding.



Goals: Why ICA?

- Want Generative Model to be able to do Bayesian Inference
- Want to separate independent signals
- Want to do sparse coding of signal

The Statical Model

- Independent sources s_i with sparse distributions: p_s
- Image (patch): $I(x,y) = \sum_i A_i(x,y) \cdot s_i$
- Feature detectors/filters: $\sum_{x'} \sum_{y'} W_i(x', y') A_j(x', y') = \delta_{ij}$
- Each A_i should be seen as a vector with index $\alpha = (x, y)$ so that $A_i(x, y) = A_{i\alpha}$ is a (square) matrix and $W_i(x, y) = {A^{-1}}_{\alpha i}$

- In practice the signals (images/sound) are preprocessed with PCA keeping only the top n components which are then whitened.
- Calling these components z_i we have $z_j = \sum_i b_{ji} s_i$
- Defining $v = b^{-1}$ we have $s_i = \sum_j v_{ij} z_j$
- The goal is then to find v and p_s.

The Optimization Function

$$p(x)dx = p(y)dy \Rightarrow p(x) = p(y)\frac{dy}{dx}$$

- Maximizing probabilistic independence perspective:
  - $y_i = v_i \cdot z,\ r_i = P_s(y_i)$ $\quad$ $P_s(s) = \int_{-\infty}^{s} p_s(t)dt$
  - $H(r_i, r_j) = H(r_i) + H(r_j) - I(r_i, r_j)$
- Optimization: $H(r) = -\frac{1}{T}\sum_{t \in dataset} \log(p_r(r_t))$
  - $p_r(r) = \frac{p_y(y)}{|dr/dy|} = \frac{p_y(y)}{p_s(y)}$
  - $p_y(y) = \frac{p_z(z)}{|dy/dz|} = \frac{p_z(z)}{|\det(v)|}$ $\quad$ $p_r(r) = \frac{p_z(z)}{|\det(v)| \prod_i p_s(y_i)}$
  - $H(r) = |\det(v)| + \frac{1}{T}\sum_i \log(p_s(v_i \cdot z)) - H(z)$

- Maximizing sparcity perspective:
  - Max Likelihood: $L(v, p_s) = \prod_{t \in dataset} p(z_t) = \prod_{t \in dataset} |\det(v)| \prod_i p_s(v_i \cdot z_t)$

$$\log(L) = T \log(|\det(v)|) + \sum_i \sum_{t \in dataset} \log(p_s(v_i \cdot z_t))$$

$$\log(p_s(s)) = h_i(s^2)$$

  - If for example we start with

$$p_s(s) = \frac{1}{\sqrt{2}} \exp(-\sqrt{2}|s|) \implies h(u) = -\sqrt{u}$$

The Algorithm

- Objective function: $f(p_s, v) = \frac{1}{T}\sum_i \log(p_s(v_i \cdot z))$
- Alternate between maximizing with respect to v and p_s

The Result

- p_s is close to $p_s(s) = \frac{1}{\sqrt{2}} \exp(-\sqrt{2}|s|) \implies h(u) = -\sqrt{u}$
- $A_i(x, y)$ are similar to Gabor functions

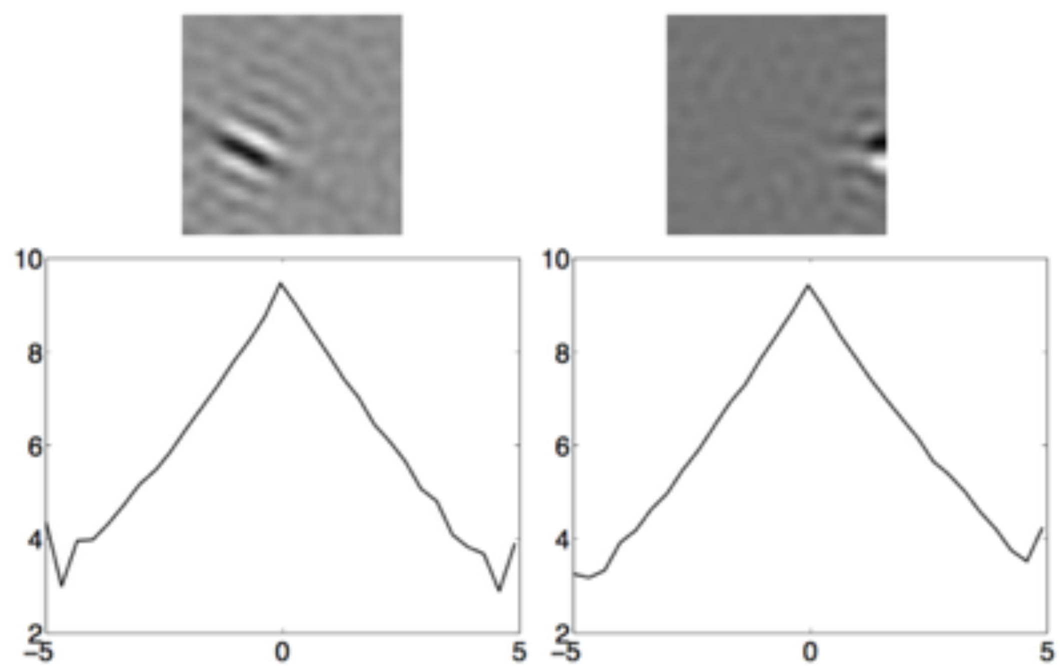


Fig. 7.6: Estimated optimal h_i from natural images. After doing ICA, the histograms of the component with the highest kurtosis and a component with kurtosis in the middle range were computed, and their logarithms taken. The feature corresponding to the highest kurtosis is on the left, and the one corresponding to the mid-range kurtosis is on the right. Top row: feature. Second row: logarithm of pdf. Third row: optimal h_i. Bottom row: the derivative of log-pdf for future reference.

Conclusion

- Want a generative model of signal to do bayesian inference, disentangle independent signals and do sparse coding.
- Objective function: $f(p_s, v) = \frac{1}{T} \sum_i \log(p_s(v_i \cdot z))$
- Results (for natural images) in Gabor-like features and sparse distribution.