ICA using Artificial Neural Networks with Hebbian Learning

Colin Fyfe

Applied Computational Intelligence Research Unit
The University of Paisley
Scotland
colin.fyfe@paisley.ac.uk

Outline

- Hebbian Learning
- The negative feedback network – PCA
- Exploratory Projection Pursuit
  - The Higher Moments Algorithm (1995)
  - The Maximum Likelihood Algorithm (2001)
- Application to Linear ICA
- A Different Problem

Hebb’s Postulate

When an axon of a cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A’s efficiency as one of the cells firing B, is increased.

Hebbian Learning

$$y_i = \sum_j w_{ij} x_j$$

$$\Delta w_{ij} = \eta x_j y_i$$

Therefore

$$\Delta w_{ij} = \eta x_j \sum_k w_{ik} x_k = \eta \sum_k w_{ik} x_k x_j$$

The Negative Feedback Network

- Feedforward: $$y_i = \sum_{j=1}^{N} W_{ij} x_j , \forall_i$$

Feedback $$e_j = x_j - \sum_{i=1}^{M} W_{ij} y_i$$

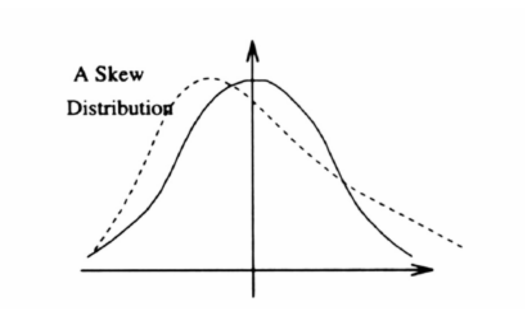
Weight Update: $$\Delta W_{ij} = \eta e_j y_i$$

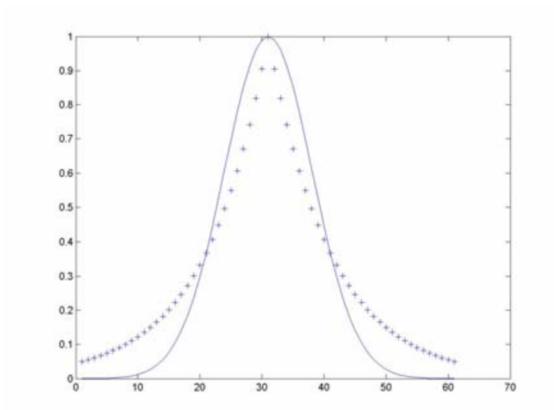
Exploratory Projection Pursuit

Exploratory Projection Pursuit (EPP) is a statistical method for solving the difficult problem of identifying structure in high dimensional data.

Define an “index” that measures the interestingness of a projection.

Interesting directions are those which are as non-Gaussian as possible.





Kurtotic Dis. Vs. Gaussian Dis.

The First EPP network (1995)
The Higher Moments Algorithm

$$s_i = \sum_{j=1}^{N} W_{ij} x_j$$

$$e_j = x_j - \sum_{k=1}^{M} W_{kj} s_k$$

$$r_i = f(s_i)$$

$$\Delta W_{ij} = \eta r_i e_j$$

Which Function ?

- To maximise kurtosis, use $f(y)=y^3$
- To maximise skewness, use $f(y)=y^2$
- To minimise kurtosis, use $f(y)=\tanh(y)$
- To maximise kurtosis, use $f(y)=y-\tanh(y)$

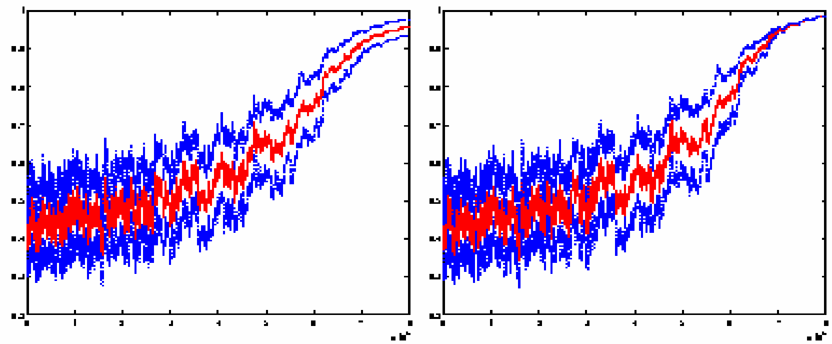


Figure 1: Left: convergence of 67 experiments towards the optimal (kurtotic)direction with $f(y) = y^3$. The graph shows the mean value of the cosine of the angle between the network weights and the optimal weight at each iteration and one standard deviation either side of the mean. Right: convergence of the best 64 experiments.

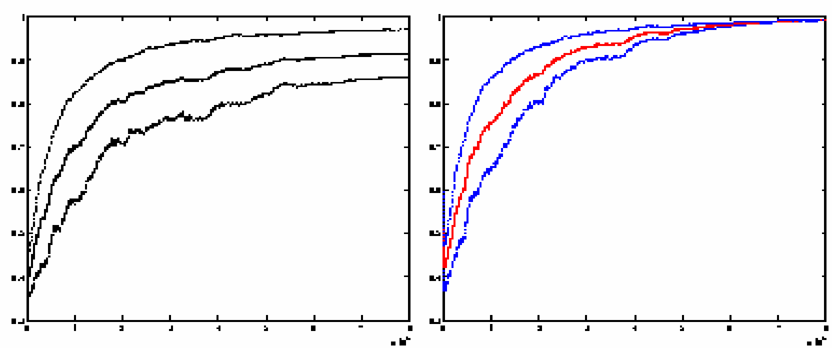


Figure 2: Left: convergence of 100 experiments towards the optimal (kurtotic) direction. The graph shows the mean value of the cosine of the angle between the network weights and the optimal weight at each iteration and one standard deviation either side of the mean. Right: convergence of the best 90 experiments. These graphs show results when using $f(y) = y - tanh(y)$.

Application to Linear ICA

- Most natural signals are kurtotic.
- Mixtures of such signals are more Gaussian than the individual signals.
- Therefore a search for kurtosis is a search for independent components.
- Mark Girolami’s PhD (1998)

The pdf of the residuals

Let the residual after feedback have probability density function:

$$p(\mathbf{e}) = \frac{1}{Z}\exp(-|\mathbf{e}|^{p})$$

A general cost function associated with this network as:

$$J = -\log p(\mathbf{e}) = |\mathbf{e}|^{p} + K$$

Therefore performing gradient descent on J we have:

$$\Delta W \propto -\frac{\partial J}{\partial W} = -\frac{\partial J}{\partial \mathbf{e}}\frac{\partial \mathbf{e}}{\partial W} \approx y(p\,|\mathbf{e}|^{p-1}\; sign(\mathbf{e}))^{T}$$

Maximum Likelihood Hebbian Learning (2001)

- Feedforward: $$y_i = \sum_{j=1}^{N} W_{ij} x_j \qquad \forall i$$
- Feedback: $$e_j = x_j - \sum_{i=1}^{M} W_{ij} y_i$$
- Weight Update: $$\Delta W_{ij} = \eta . y_i . sign(e_j) | e_j |^{p-1}$$
- .

Emilio Corchado's PhD(2002)

Minimum Likelihood Hebbian Learning (anti-hebbian rule)

$$\Delta W \propto \frac{\partial J}{\partial W} = \frac{\partial J}{\partial \mathbf{e}} \frac{\partial \mathbf{e}}{\partial W} \approx -\mathbf{y}(p\,|\,\mathbf{e}\,|^{p-1}\ sign(\mathbf{e}))^T$$

We are aiming to minimise the likelihood of the residual.
We are thus using our learning rules to make the residuals as unlikely as possible (determined by the *p* parameter).

Experiments Using an Astronomical Data set

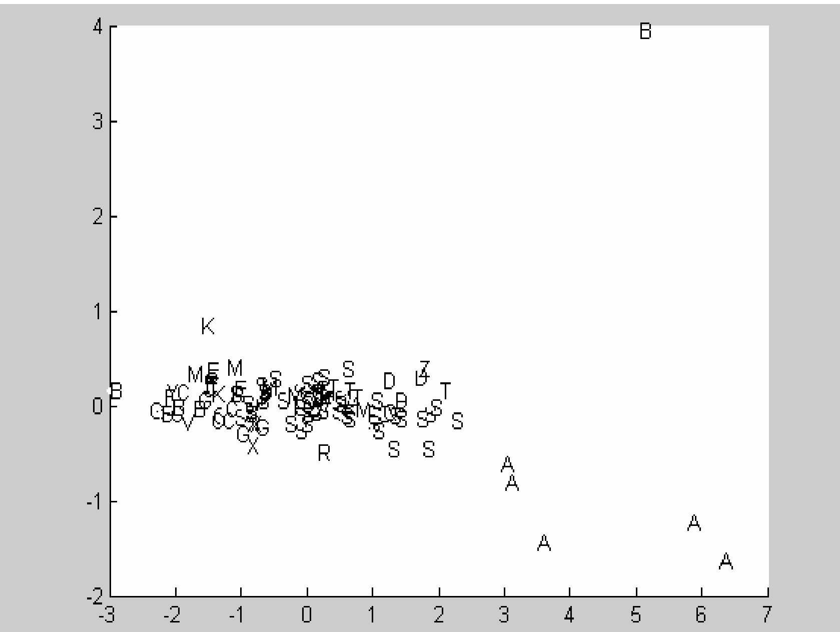


Fig. Projection of the asteroid data set onto the first two Principal Components.

Emilio Corchado's PhD (2002)

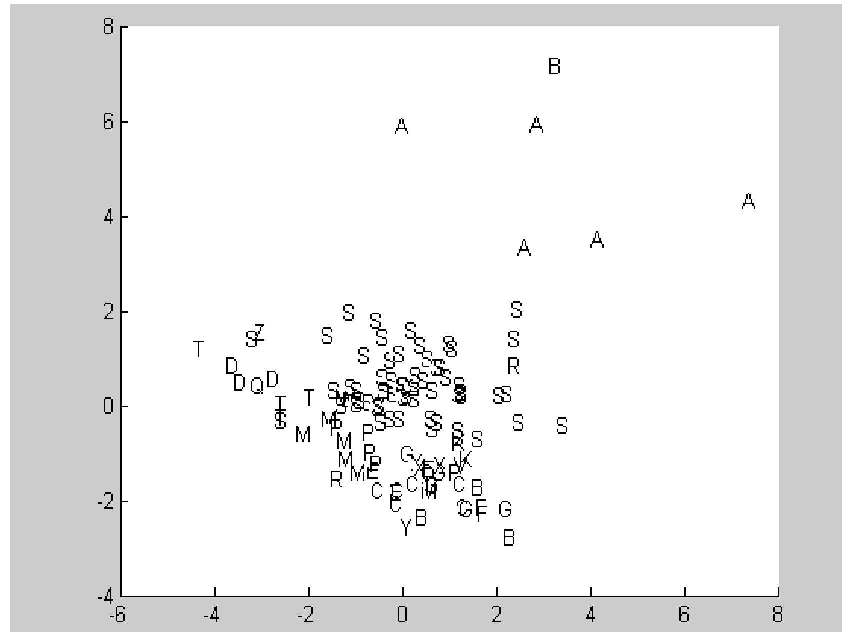


Fig. Projection of the asteroid data onto the filters found by Maximum Likelihood Hebbian Learning with p=0.5

Independent Component Analysis

Extraction of a signal from a mixture

- 5 subgaussian artificially generated signals randomly mixed. Their kurtosis values were -0.9845, -0.9638, -0.9769, -0.9795 and -0.9673 respectively. We used 40,000 samples, a learning rate of 0.0001 with p=4.

WVA

-0.0330	0.0219	**1.0011**	-0.0458	0.0134
-0.0211	-0.0351	0.0552	**0.9977**	-0.0288
0.0386	0.0228	- 0.0010	-0.0122	**-0.9713**
-0.0062	**-0.9966**	0.0275	-0.0399	-0.0999
-0.9986	0.0125	-0.0363	-0.0041	-0.0214

Real Speech Signals

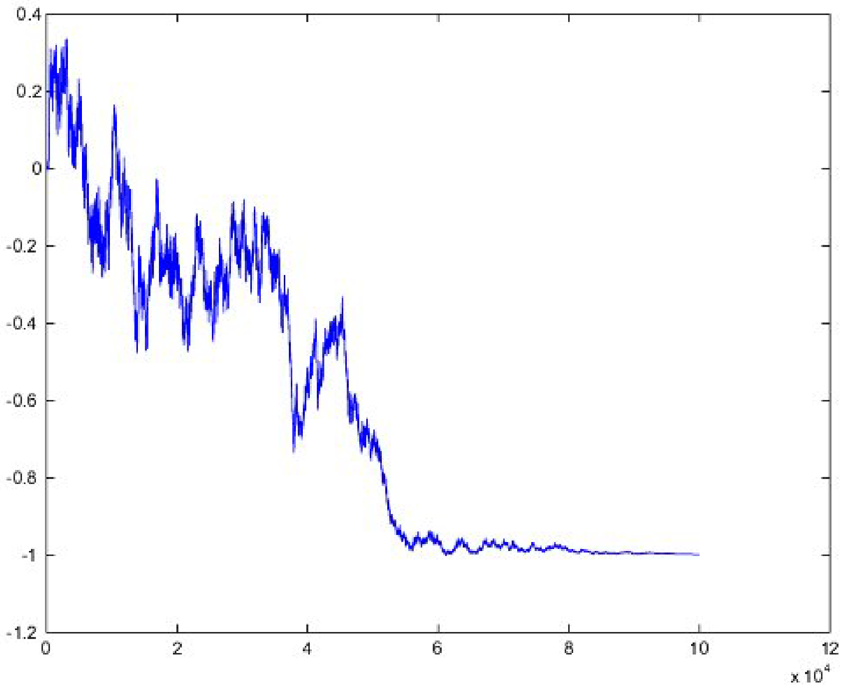
- 3 speakers, linearly mixed.

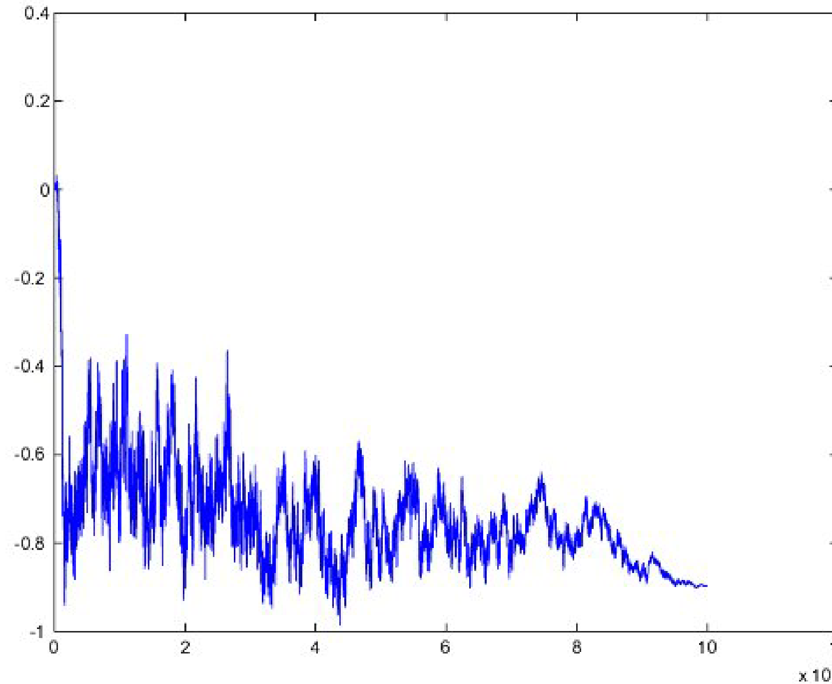
$$\Delta W_{ij} = \eta . y_i . sign\left(e_j\right)$$

- W*V*A=

W*V*A=	0.017	-0.005	**-1.000**
	1.001	-0.009	0.016
	0.012	**1.000**	-0.002

Comparing and Mixing the two EPP methods





The left figure shows the convergence of the higher moments EPP algorithm.

The right one shows the convergence of the Maximum Likelihood EPP algorithm in terms of the dot product to the ideal solution

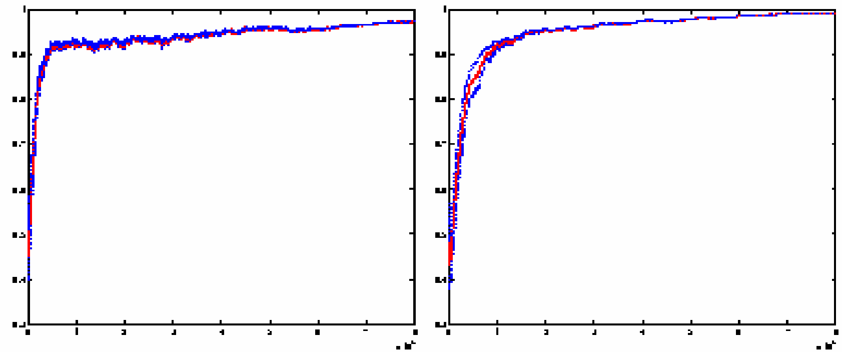


Figure 5: Convergence of the algorithm using the combined learning rule. Left:bimodal data. Right: leptokurtotic data

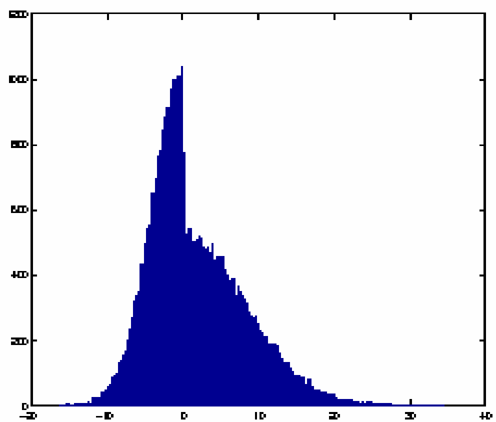
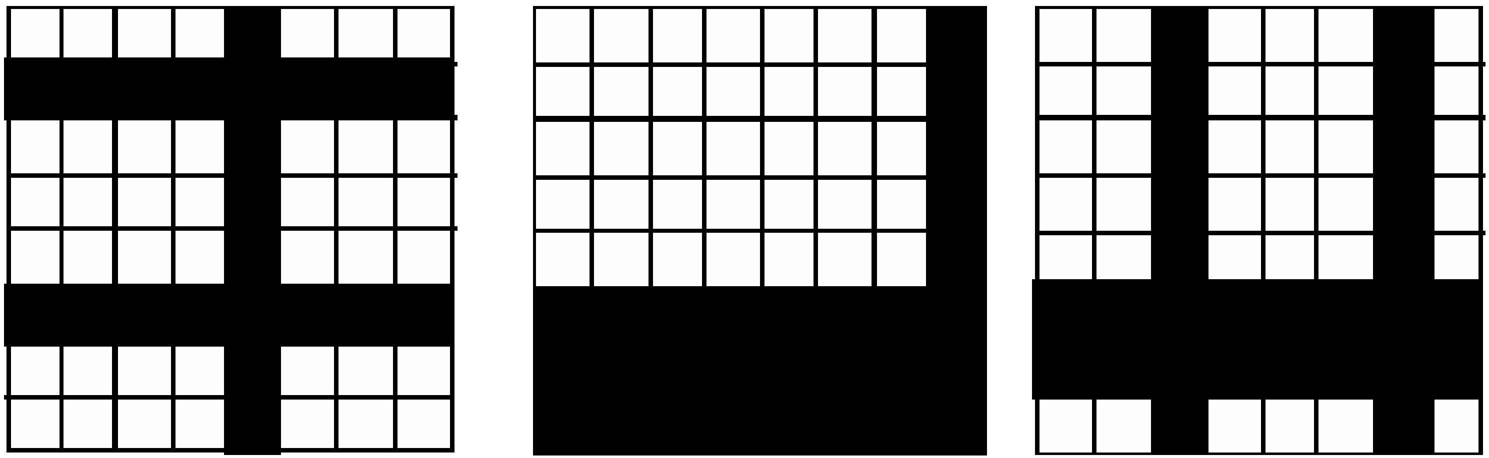


Figure 6: Histogram of samples of skewed distribution.

A Different ICA Problem



- Darryl Charles PhD (1999)

Rectification

- At the weight change stage, if a weight becomes negative, set it to 0.

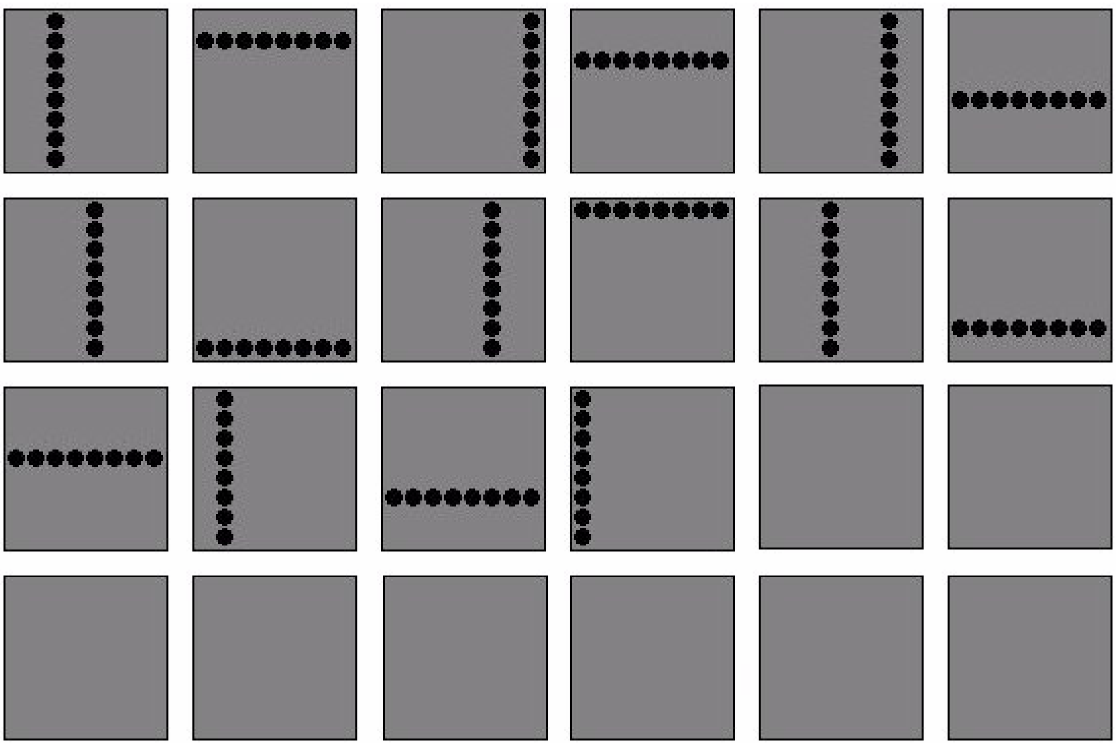
OR

- At the feedforward stage, if an output is negative, set it to zero.

- Related to factor analysis

The Minimum Overcomplete Basis

- Addition of noise to the outputs forces identification of MOB.



Conclusion

•Feedforward: $$y_i = \sum_{j=1}^{N} W_{ij} x_j , \forall_i$$

Feedback $$e_j = x_j - \sum_{i=1}^{M} W_{ij} y_i$$

Weight Update: $$\Delta W_{ij} = \eta e_j y_i$$