Towards an Agent–Based Independent Component Analysis

Roman V. Belavkin (R.Belavkin@mdx.ac.uk)

School of Computing Science,
Middlesex University, London NW4 4BT, UK

September 22, 2005

THE ICA PROBLEM

Let $\boldsymbol{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^T$ be the observable mixture of sources $\boldsymbol{S} = (\mathbf{s}_1, \ldots \mathbf{s}_m)^T$

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S}\,, \quad \text{where } \boldsymbol{A} = (a_{ij}) \text{ is an } m \times n \text{ matrix}$$

Assuming that

1. $P(\mathbf{s}_1, \ldots, \mathbf{s}_m) = P(\mathbf{s}_1) \cdots P(\mathbf{s}_m)$ (independence)
2. $\forall \mathbf{s}_i$ but one are non–Gaussian

Find demixing matrix $\boldsymbol{W} \approx \boldsymbol{A}^{-1}$ such that

$$\boldsymbol{Y} = \boldsymbol{W}\boldsymbol{X} \approx \boldsymbol{A}^{-1}\boldsymbol{X} = \boldsymbol{S}$$

MEASURE OF INDEPENDENCE

We seek $\boldsymbol{W}$ to minimise mutual information in $\boldsymbol{Y}$

$$
\begin{aligned}
I(\boldsymbol{Y}) &= \sum_{i=1}^{n} H(\mathbf{y}_i) - H(\boldsymbol{Y}) \\
&= \sum_{i=1}^{n} H(\mathbf{y}_i) - H(\boldsymbol{X}) - \ln|\boldsymbol{W}| \to 0
\end{aligned}
$$

For pre–whitened $\boldsymbol{X}$, $\ln|\boldsymbol{W}| = 0$, and therefore

$$\boldsymbol{W} = \arg\min_{\boldsymbol{W}} H(\mathbf{y}_1) + \cdots + H(\mathbf{y}_n)$$

DIRECT ENTROPY ESTIMATION

To estimate $H(\mathbf{y}_i)$, we use the direct approximation due to Vasicek (1976):

$$H(z^1, \dots, z^n) \approx \frac{1}{n} \sum_{i=1}^{n-m} \ln \left(\frac{n}{m} (z^{(i+m)} - z^{(i)}) \right)$$

where $z^1, \dots, z^n$ is a sample of random variable Z, and $z^{(i)}$ is a non–decreasing ordering $z^{(1)} \leq \dots \leq z^{(n)}$.

We shall minimise $\sum_{i=1}^{n} H(\mathbf{y}_i)$ by rotating $\boldsymbol{W}$ by angle θ as in Learned-Miller and Fisher (2003).

JACOBI ROTATIONS

Used to rotate $\boldsymbol{W}$ in i, j plane by angle θ

$$\boldsymbol{J}(i,j,\theta) = \begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & \cos\theta & \cdots & -\sin\theta & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & \sin\theta & \cdots & \cos\theta & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}$$

In two dimensions $\boldsymbol{J}(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$

$$\boldsymbol{W}^{\text{new}} = \boldsymbol{J}(i,j,\theta)\boldsymbol{W}$$

THE AGENT ARCHITECTURE

The following decision–theoretic agents architecture is used (Belavkin, in press)

$$X = \{x_1, \dots, x_m\} \quad \text{percepts}$$
$$Y = \{y_1, \dots, y_n\} \quad \text{preferences (e.g. } Y = \{\text{success, failure}\}\text{)}$$
$$Z = \{z_1, \dots, z_k\} \quad \text{actions}$$

The Markov transition model $P(X, Y, Z) = (p_{ij}^k)$, where $p_{ij}^k = P(x_i, y_j, z_k)$, is used as the associative memory and can be used for Bayesian inference

$$P(Y \mid X, Z) = \alpha P(X, Y, Z)\,, \quad \text{where } \alpha = \frac{1}{\|P(Y|X,Z)\|}$$

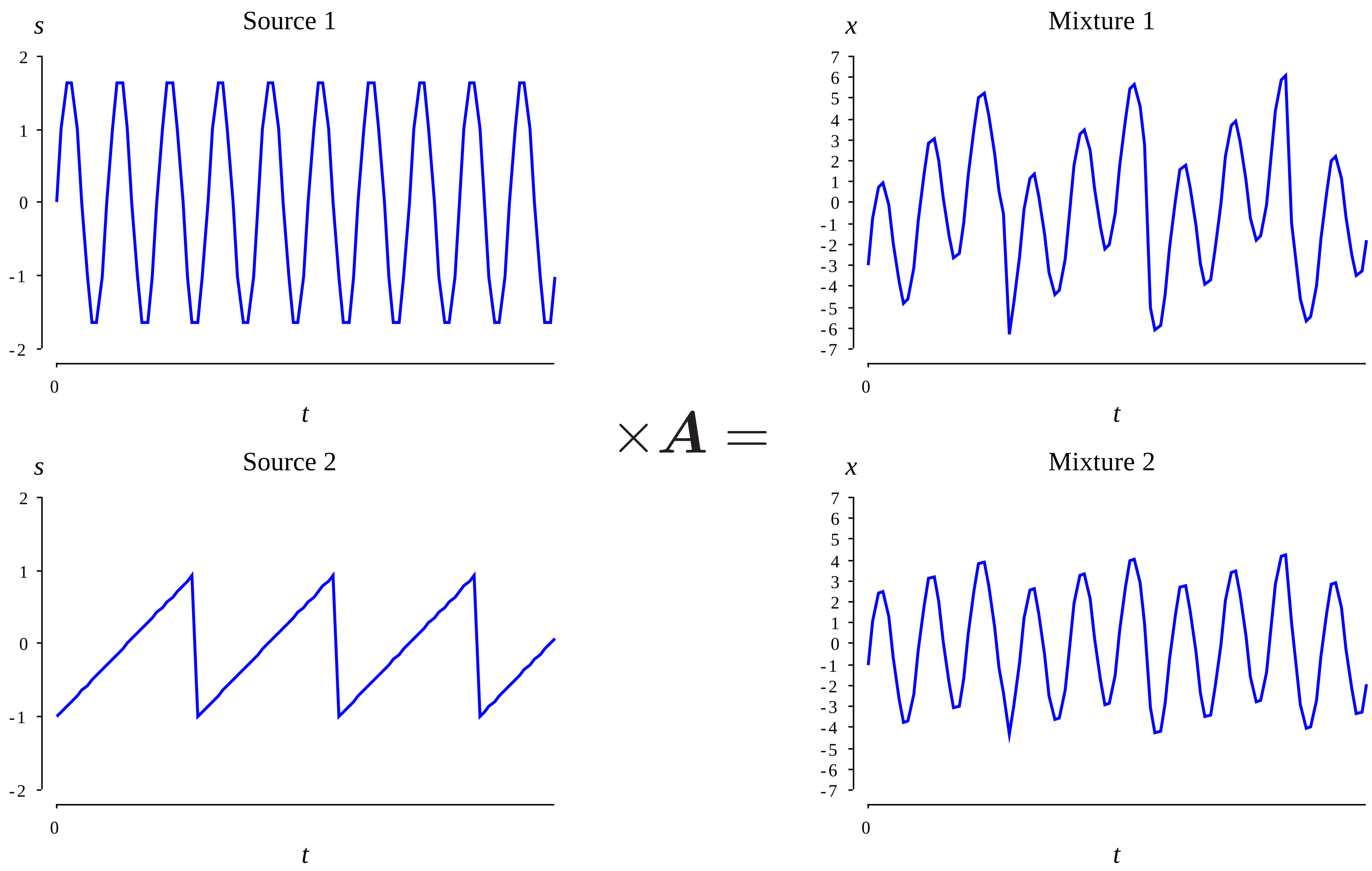
SETTING UP THE AGENT FOR ICA

- Let angles $\theta \in [0, \pi/2]$ be the percepts of the agent
- Changes of angle $\Delta\theta \in [\theta^-, \theta^+]$ be the actions
- Changes of entropy $\Delta H = \Delta \sum_{i=1}^{n} H(\mathbf{y}_i)$ be related to preferences (i.e. negative change Δ is a success)

With this setup, the agent learns which rotations $\Delta\theta$ minimise the entropy faster

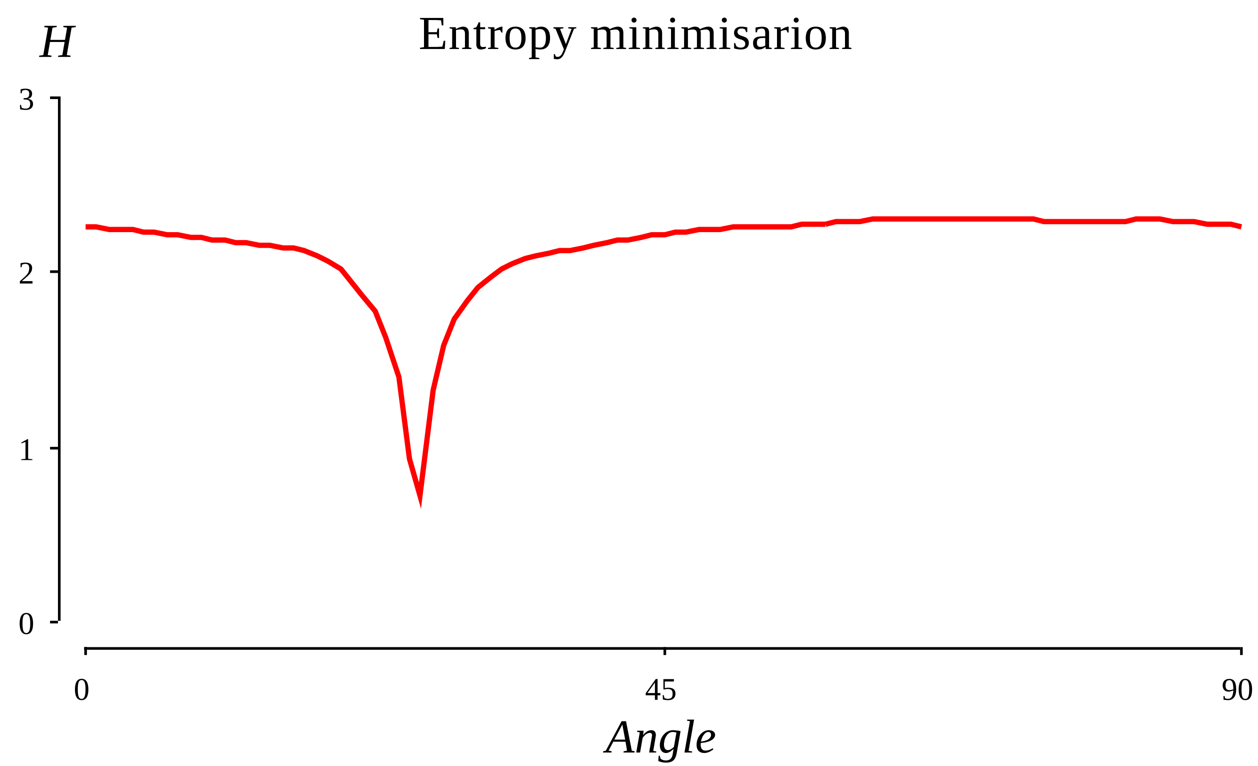
$$P(\Delta H \mid \theta, \Delta\theta)$$

MIXING TWO SOURCES

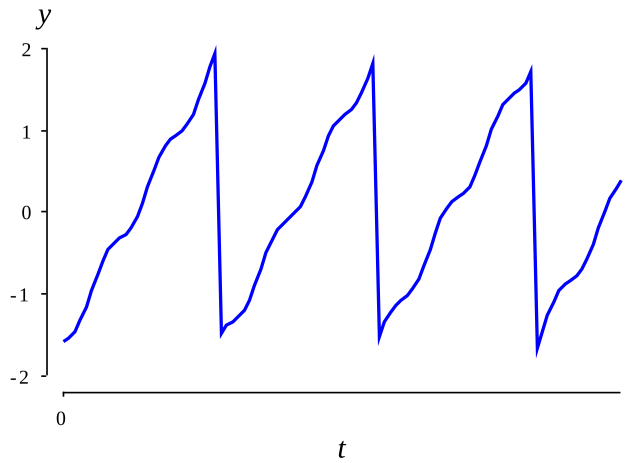


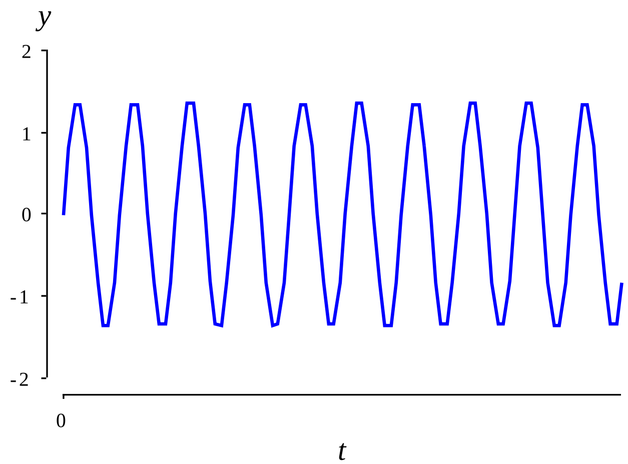
MARGINAL ENTROPIES

By changing the angle, the agent searches the angle space $\theta \in [0, \pi/2]$ in order to minimise marginal entropy



DEMIXING THE SIGNALS





The agent's output is $J(\theta)$ and demixing matrix $\boldsymbol{J}(\theta)\boldsymbol{W}$ such that

$$\boldsymbol{J}(\theta)\boldsymbol{W}\boldsymbol{Y} \approx \boldsymbol{S}$$

FUTURE WORK

- Use the estimation of entropy as a feedback parameter to control the precision of rotations $\Delta\theta$.
- Use communities of agents each minimising individual component $H(\mathbf{y}_i)$.
- Investigate the possibility of a **non–linear** ICA using the same agent–based approach. This may be done by assigning different (non–linear) transformations to actions of agents.

References

Belavkin, R. V. (in press). *Acting irrationally to improve performance in stochastic worlds.* (In *Proceedings pf the 25th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence*)

Learned-Miller, E., & Fisher, J. (2003). ICA using spacings estimates of entropy. *Journal of Machine Learning Research*, *4*, 1271–1295.

Vasicek, O. (1976). A test for normality based on sample entropy. *Journal of the Royal Statistical Society, Series B*, *38*(1), 54–59.