Aapo Hyvärinen, Urs Köster

FastISA: A fast fixed-point algorithm for Independent Subspace Analysis

Helsinki Institute for Information Technology, Department of Computer Science University of Helsinki, Finland {ahyvarin, koster}@cs.helsinki.fi

Abstract

Independent Subspace Analysis (ISA; Hyvarinen & Hoyer, 2000) is an extension of ICA. In ISA, the *components are divided into subspaces*. Components in different subspaces are assumed independent, whereas components in the same subspace have dependencies.

1. **We present a fixed-point algorithm for ISA estimation (formulated in analogy to FastICA).**
2. **We sketch a proof of the quadratic convergence of the algorithm.**
3. **We present simulations that confirm the fast convergence.**
4. **We show that the method is prone to convergence to local minima.**

For details of the proof see the accompanying conference paper!

Introduction

- **Independent Component Analysis (ICA)** has been very successfully in the past, but:
- Because it is a linear model and it requires independent sources, its range is limited.
- **ISA an extension of ICA**, in which certain dependencies between sources can be modeled by including a **pooling stage** and a **nonlinear transformation**.
- The pooling organizes filters into subspaces inside which dependencies are allowed.
- The estimation is very similar to ICA. It follows the assumption that the subspaces are mutually independent.
- It can be performed by gradient descent, which is however quite slow and inefficient.
- This motivates a **fixed-point algorithm for ISA**, which we present in this paper.
- We discuss the **mathematical background** of ISA, sketch a **convergence proof** and show **simulations** of the convergence properties.

Model and Algorithm

- **ICA** models image patches x as a linear superposition of features a, where the activations s are independent, so $x = As$.
- **ISA** extends this with a pooling. The features are grouped into *subspaces* where dependencies are allowed.

Algorithm and Proof

The key difference between ICA and ISA: We introduce **independent feature subspaces**. We do not require independence of individual sources, just between norms of projections on these subspaces. We define one such element as

$$u_i = (\sum_{j \in S_i} s_j^2)^{1/2} = (\sum_{j \in S_i} (\mathbf{w}_j^T \mathbf{x})^2)^{1/2}$$

i.e. we project onto the group S_i of elements which belong to the i-th subspace and compute the norm. This is a *nonlinear mapping*, which makes the method capable of modeling *complicated dependency structures* that linear ICA cannot capture. **For the estimation of the model we define the probability distribution function:**

$$\log p(s_1, ..., s_m) = \sum_{j=1}^{m} \left(-\log Z_j - \frac{G(\sum_{i \in S_j} s_i^2)}{b} \right)$$

The square root has been replaced by the contrast function $G(.)$. Z normalizes the distribution and b ensures unit variance. They can be computed in closed form for some choices of $G(.)$ [3]. We choose $G(x) = \sqrt{x + \gamma}$, where $\gamma = 0.1$ is an arbitrary constant to aid with stability.
We will proof that the following **update step** for the rows $\mathbf{w}$ of the demixing matrix leads to at least quadratic convergence:

$$\begin{aligned} \mathbf{w}_j^+ &= E\{\mathbf{x}(\mathbf{w}_j^T\mathbf{x})g(\sum_{i \in S_j} (\mathbf{w}_i^T\mathbf{x})^2)\} \\ &- E\{g(\sum_{i \in S_j} (\mathbf{w}_i^T\mathbf{x})^2) + 2(\mathbf{w}_j^T\mathbf{x})^2 g'(\sum_{i \in S_j} (\mathbf{w}_i^T\mathbf{x})^2)\}\mathbf{w}_j \end{aligned}$$

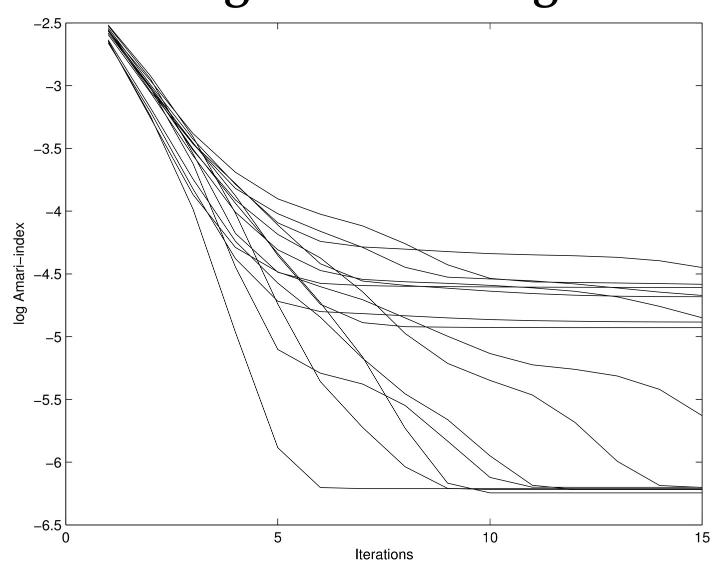
$E\{.\}$ denotes the expectation value, S_j is the set of indices of components belonging to the subspace, $g(.)$ and $g'(.)$ are the first and second derivatives of the nonlinearity $G(.)$. The algorithm requires the data to be whitened. We orthogonalize $\mathbf{W}$ after each step, which is equal to decorrelation since we are in whitened space. Details of the proof cannot be given here, but it proceeds roughly as follows:

- Make a change of variables to $Z = AW$ so at the solution we have $Z = I$
- Assuming convergences close to a solution, Z is identity up to a perturbation ϵ
- Follow the update step
- Use a Taylor expansion in ϵ to expand all the terms
- Evaluating the expectation value of all the terms, one can show that they go to zero or are of the order ϵ^2
- This proofs that convergence is at least quadratic.

Simulations

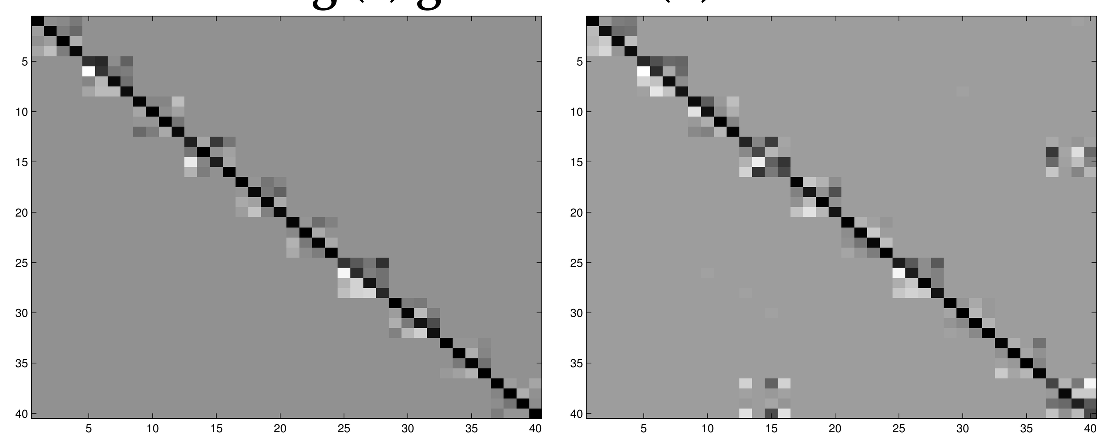
We generated mixtures of supergaussian data with an embedded subspace structure by first taking 50,000 samples from a 40-dimensional white Gaussian distribution with zero mean and unit variance. We then divided this into subspaces of dimensionality four, and multiplied each element of a subspace by a random variable drawn from a uniform distribution. This produces the required supergaussian distribution, and introduces dependencies within subspaces. We randomly generated a mixing matrix to obtain the "observations", which were then whitened. Then the FastISA algorithm was applied. To investigate whether the algorithm converged to a local minimum instead of the global solution, we computed the matrix $\mathbf{Z} = \mathbf{A}^T\mathbf{W}$.

1. Convergence of the algorithm



The convergence is fast, the algorithm usually converges to a minimum in 5-15 steps. The algorithm was initialized with the correct solution perturbed by white noise of unit norm. Under these conditions, convergence is to the global minimum for 6 of the 15 random trials.

2. Z reaching (a) global and (b) local minimum



(a) *The product* $\mathbf{Z}$ *of the mixing and filter matrices is plotted, which gives a block-diagonal matrix for the global optimum. The residual log Amari-index is* -6.2 *which corresponds to an residual error of the order* 10^{-3}*. This is mainly due to the assumption of infinite expectations.* **(b)** *It cannot be guaranteed in general that the global minimum of the error surface is found. A local minimum is indicated by multiple blocks in the bottom and leftmost position. The log Amari-index [4] here is* -5.

Simulations

Initializing $\mathbf{W}$ randomly, we observed that convergence was always to a local minimum. Therefore we validated the convergence properties by starting the optimization not on a random point on the error surface, but close to the optimal solution, which is known in this case, since the mixtures are artificially generated. Under these conditions we get convergence to the global minimum, given that the starting point was close enough. This is depicted in Fig. 2a for data with a dimensionality of 40. $\mathbf{Z}$ should converge to a permuted block-diagonal matrix, since rotations inside subspaces do not affect the likelihood. Here, the Amari-index[4] for subspaces was computed by adding up the absolute values of all elements of the blocks of the main diagonal.

Conclusion

We presented a fixed-point algorithm for ISA, analogous to the ones presented in [1, 5]. The convergence of the algorithm was proven to be quadratic. Simulations show that the convergence is fast, but they also point out the problem of local minima. The problem of local minima is probably more related to the model specification itself because it was already encountered in [6], and not due to our particular algorithm.

References

[1] A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. on Neural Networks*, 10(3):626–634, 1999.

[2] P. Comon. Independent component analysis – a new concept? *Signal Processing*, 36:287–314, 1994.

[3] U. Köster A. Hyvärinen. Complex cell pooling and the statistics of natural images. *submitted to Network: Computation in Neural Systems*, available online: cs.helsinki.fi/u/koster/koster05.pdf, 2005.

[4] S. Cichocki, A.; Amari. *Adaptive Blind Signal and Image Processing. Learning Algorithms and Applications*. Wiley, 2002.

[5] A. Hyvärinen E. Bingham. A fast fixed-point algorithm for independent component analysis of complex valued signals. *Journal of Neural Systems*, 10:1–8, 2000.

[6] A. Hyvärinen and P. O. Hoyer. Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces. *Neural Computation*, 12(7):1705–1720, 2000.

Acknowledgments: Urs Köster is supported by the Alfried Krupp von Bohlen und Halbach-Foundation.