

## 6.4 Properties of the FastICA Algorithm

The FastICA algorithm and the underlying contrast functions have a number of desirable properties when compared with existing methods for ICA.

1. The convergence is cubic (or at least quadratic), under the assumption of the ICA data model (for a proof, see (Hyvärinen, 1999a)). This is in contrast to ordinary ICA algorithms based on (stochastic) gradient descent methods, where the convergence is only linear. This means a very fast convergence, as has been confirmed by simulations and experiments on real data (see (Giannakopoulos et al., 1998)).
2. Contrary to gradient-based algorithms, there are no step size parameters to choose. This means that the algorithm is easy to use.
3. The algorithm finds directly independent components of (practically) any non-Gaussian distribution using any nonlinearity  $g$ . This is in contrast to many algorithms, where some estimate of the probability distribution function has to be first available, and the nonlinearity must be chosen accordingly.
4. The performance of the method can be optimized by choosing a suitable nonlinearity  $g$ . In particular, one can obtain algorithms that are robust and/or of minimum variance. In fact, the two nonlinearities in (40) have some optimal properties; for details see (Hyvärinen, 1999a).
5. The independent components can be estimated one by one, which is roughly equivalent to doing projection pursuit. This is useful in exploratory data analysis, and decreases the computational load of the method in cases where only some of the independent components need to be estimated.
6. The FastICA has most of the advantages of neural algorithms: It is parallel, distributed, computationally simple, and requires little memory space. Stochastic gradient methods seem to be preferable only if fast adaptivity in a changing environment is required.

A Matlab<sup>TM</sup> implementation of the FastICA algorithm is available on the World Wide Web free of charge.<sup>1</sup>