BOOK REVIEW

Independent Component Analysis

By Aapo Hyvarinen, Juha Karhunen, Erkki Oja

A John Wiley-Interscience Publication, 2001

Series on Adaptive and Learning Systems for Signal Processing,

Communications, and Control

Independent Component Analysis (ICA) is one of the most exciting new topics in fields such as neural networks, advanced statistics, and signal processing. This is the first book to provide a comprehensive introduction to this new technique complete with the fundamental mathematical background needed to understand and utilize it. It offers a general overview of the basics of ICA, important solutions and algorithms, and in-depth coverage of new applications in image processing, telecommunications, audio signal processing, and more.

The objective of this book is to cover both the mathematical background and principles, algorithmic solutions, and the practical applications of the present state of the art of ICA. In my opinion a comprehensive and detailed textbook on this subject is absent in literature.

This book provides a comprehensive introduction to ICA as a statistical and computational technique. The emphasis is on fundamental mathematical principles and basic algorithms. Much of the material is based on the original research conducted in the authors' own research group, which is naturally reflected in the weighting of different topics. The authors give a wide coverage especially to those algorithms that are scalable to large problems, that is, work even with a large number of observed variables and data points. These will be increasingly used in the near future when ICA is extensively applied in practical real-world problems instead of the toy problems or small pilot studies that have been predominant until recently. Respectively, somewhat less emphasis is given to more specialized signal processing methods involving convolutive mixtures, delays, and other blind source separation (BSS) techniques than ICA.

The book contains 24 chapters, grouped in 4 parts, references and an index. Chapter 1 of the book "Introduction" provides an overview of the problems discussed in the book and introduces the basic concepts, applications, and estimation principles of ICA.

Part I gives the mathematical preliminaries. It introduces the general mathematical concepts needed in the rest of the book. This part starts with a course on probability theory, statistics, and random process with the emphasis on multivariable statistics and random vectors in Chapter 2, entitled "Random Vectors and Independence". The reader is assumed to be familiar with most of the basic material in this chapter, but also some concepts more specific to ICA are introduced, such as higher-order cumulants and multivariable probability theory. Next, Chapter 3 "Gradients and Optimization Methods" discusses essential concepts in optimization theory and gradient methods, which are needed when developing ICA algorithms. The authors discuss some typical iterative optimization algorithms and their properties. Most of the algorithms are based on the gradients of the cost functions. Therefore, vector and matrix gradients are reviewed first, followed by the most typical ways to solve unconstrained and constrained optimization problems with gradient-type learning algorithms. Estimation theory is reviewed in Chapter 4, "Estimation Theory", which focuses mainly on linear data models, studying the estimation of their parameters. The two cases of deterministic and random parameters are covered, but the parameters are always assumed to be time-invariant. The methods that are widely used in context with independent component analysis are emphasized in this chapter. A complementary theoretical framework for ICA is information theory, covered in Chapter 5, "Information Theory", where the basic concepts of information theory are introduced. The latter half of the chapter deals with a more specialized topic:

approximation of entropy. Part I is concluded by Chapter 6 "Principal Component Analysis and Whitening", which discusses methods related to principal component analysis (PCA), factor analysis, and decorrelation.

In **Part II,** the basic ICA model is covered and solved. This is the linear instantaneous noise-free mixing model that is classic in ICA, and forms the core of the ICA theory. The model is introduced and the question of identifiability of the mixing matrix is treated in Chapter 7 "What is Independent Component Analysis?". The basic concepts of independent component analysis are defined and are discussed in a couple of practical applications. These serve as motivation for the mathematical formulation of ICA, which is given in the form of a statistical estimation problem. The following chapters treat different methods of estimating the model. A central principle is nongaussianity, whose relation to ICA is first discussed in Chapter 8 "ICA by Maximization of Nongaussianity". In this chapter it is introduced a simple and intuitive principle for estimating the model of independent component analysis, based on maximization of nongaussianity. Next, the principles of maximum likelihood, discussed in Chapter 9, "ICA by Maximum Likelihood Estimation", and minimum mutual information, making the object of Chapter 10, "ICA by Minimization of Mutual Information", are reviewed, and connections between these three fundamental principles are shown. Material that is less suitable for an introductory course is covered in Chapter 11, "ICA by Minimization of Mutual Information", which discusses the algebraic approach using higher-order cumulant tensors, and Chapter 12, "ICA by Nonlinear Decorrelation and Nonlinear PCA", which reviews the early work on ICA based on nonlinear decorrelations, as well as the nonlinear principal component approach. Practical algorithms for computing the independent components and the mixing matrix are discussed in connection with each principle. Next, some practical considerations, mainly related to preprocessing and dimension reduction of the data are discussed in Chapter 13, "Practical Considerations", including hints to practitioners on how to really apply ICA to their own problem: preprocessing by time filtering, preprocessing by principal component analysis, how many components should be estimated and choice of the algorithm. An overview and comparison of the various ICA methods is presented in Chapter 14, "Overview and Comparison of Basic ICA Methods", which thus summarizes Part II. The main discussed problems are: objective functions vs. algorithms, connections between ICA estimation principles, statistically optimal nonlinearities and experimental comparison of ICA algorithms.

In **Part III**, different extensions of the basic ICA model are given. This part is by its nature more speculative than Part II, since most of the extensions have been introduced very recently, and many open problems remain. In an introductory course on ICA, only selected chapters from this part may be covered. First, in Chapter 15, "Noisy ICA", it is treated the problem of introducing explicit observational noise in the ICA model. Estimation of the mixing matrix seems to be quite difficult when noise is present. It could be argued that in practice, the problem would be to reduce noise in the data before performing ICA. In noisy ICA, the authors encounter a new problem: estimation of the noise-free realizations of the independent components. The noisy model is not invertible, and therefore estimation of the noise-free components requires new methods. This problem leads to some interesting forms of denoising.. The situation where there are more independent components than observed mixtures is treated in Chapter 16, "ICA with Overcomplete Bases", under the form of estimation of the independent components and estimation of the mixing matrix. Chapter 17, "Nonlinear ICA", deals with independent component analysis for nonlinear mixing models. It is also addressed the nonlinear blind source separation problem. After considering these matters, some methods introduced for solving ICA and BSS problems are discussed in detail. Special emphasis is given to a Bayesian approach that applies ensemble learning to a flexible multilayer perceptron model for finding the sources and nonlinear mixing mapping that have most probably given rise to the observed mixed data. The efficiency of the method is demonstrated using both artificial and real-world data. Chapter 18, "Methods using Time Structure", discusses methods that estimate a linear mixing model similar to that of ICA, but with quite different assumptions: the components are not nongaussian but have some time dependencies instead. Chapter 19, "Convolutive Mixtures and Blind Deconvolution", deals with blind deconvolution and blind separation of convolutive mixtures. The used techniques for convolutive mixtures have in fact been developed by extending methods designed originally for either the blind deconvolution or standard ICA/BSS problems. In the appendix of this chapter, certain basic concepts of discrete-time filters needed in this chapter are briefly reviewed. Chapter 20, "Other Extensions", presents some additional extensions of the basic independent component analysis model. First, it is discussed the use of prior information on the mixing matrix, especially on its sparseness. Second, are presented models that somewhat relax the assumption of the independence of the components. Finally, it is shown how to adapt some of the basic ICA algorithms to the case where the data are complex-valued instead of real-valued.

Part IV treats some applications of ICA methods. Chapter 21, "Feature Extraction by ICA", considers a certain class of widely used signals, called natural images. The working hypotheses here is that this class is sufficiently homogenous, so that it can build a statistical model using observations of those signals, and then later use this model for processing the signals, for example, to compress or denoise them. It is proved that ICA provides a model that is very similar to most sophisticated low-level image representations used in image processing and vision research. Chapter 22, "Brain Imaging Applications", concentrates on measurements of the electrical and magnetic activity of the human brain. This is especially the case with electroencephalograms (EEG) and magnetoencephalograms (MEG), which are recordings of electric and magnetic fields of signals emerging from neural currents within the brain. In this chapter, some of these brain-imaging applications, concentrating on EEG and MEG, are reviewed. Chapter 23, "Telecommunications", deals with applications of independent component analysis (ICA) and blind source separation (BSS) methods in telecommunications. The authors concentrate on code division multiple access (CDMA) techniques, because this specific branch of telecommunications provides several possibilities for applying ICA and BSS in a meaningful way. The ultimate goal of these applications is to detect the desired user's symbols, but for achieving this, intermediate quantities such as fading channels or delays must usually be estimated first. Finally, references to other communications applications of ICA and related blind techniques used in communications are given. Some econometric (finding hidden factors in financial data, time series prediction by ICA) and audio signal processing applications, together with pointers to miscellaneous other applications are treated in Chapter 24, "Other Applications".

The material of the book is well organized, presented with clarity, and many examples developed in sufficient detail facilitate the reader's understanding. Also, all arguments are clear and logically correct. It is worth mentioning the fluency of the style, the absence of the typographical errors and the accuracy of graphical material that make the book pleasant to read.

The chapters are logically ordered, and contain plenty of examples covering important points that illustrate the given techniques. Each chapter ends with concluding remarks and references that bring the material of the chapter into context with existing literature, and an array of problems.

Concerning the audience of the book, the readership could be from a variety of disciplines, such as statistics, signal processing, neural networks, applied mathematics, neural and cognitive sciences, information theory, artificial intelligence, and engineering. It is not a book for the uninitiated; some mathematical maturity is essential to study the material. An understanding of the book requires from the reader some prior knowledge on probability, random process, linear algebra, and classical filter theory.

In conclusion, this is a comprehensive and well-written book that provides the involved reader with the sophistication and depth of independent component analysis and applications in different fields. I heartily recommend it to the readers.

Theodor Dan Popescu

National Institute for Research and Development in Informatics