

# Contents

<i>Preface</i>	<i>xvii</i>
<i>I Introduction</i>	<i>I</i>
1.1 <i>Linear representation of multivariate data</i>	<i>1</i>
1.1.1 <i>The general statistical setting</i>	<i>1</i>
1.1.2 <i>Dimension reduction methods</i>	<i>2</i>
1.1.3 <i>Independence as a guiding principle</i>	<i>3</i>
1.2 <i>Blind source separation</i>	<i>3</i>
1.2.1 <i>Observing mixtures of unknown signals</i>	<i>4</i>
1.2.2 <i>Source separation based on independence</i>	<i>5</i>
1.3 <i>Independent component analysis</i>	<i>6</i>
1.3.1 <i>Definition</i>	<i>6</i>
1.3.2 <i>Applications</i>	<i>7</i>
1.3.3 <i>How to find the independent components</i>	<i>7</i>
1.4 <i>History of ICA</i>	<i>11</i>

*Part I MATHEMATICAL PRELIMINARIES*

2	<i>Random Vectors and Independence</i>	15
2.1	<i>Probability distributions and densities</i>	15
2.1.1	<i>Distribution <b>of</b> a random variable</i>	15
2.1.2	<i>Distribution <b>of</b> a random vector</i>	17
2.1.3	<i>Joint and marginal distributions</i>	18
2.2	<i>Expectations and moments</i>	19
2.2.1	<i>Definition and general properties</i>	19
2.2.2	<i>Mean vector and correlation matrix</i>	20
2.2.3	<i>Covariances and joint moments</i>	22
2.2.4	<i>Estimation <b>of</b> expectations</i>	24
2.3	<i>Uncorrelatedness and independence</i>	24
2.3.1	<i>Uncorrelatedness and whiteness</i>	24
2.3.2	<i>Statistical independence</i>	27
2.4	<i>Conditional densities and <b>Bayes</b> rule</i>	28
2.5	<i>The multivariate <b>gaussian</b> density</i>	31
2.5.1	<i>Properties <b>of</b> the <b>gaussian</b> density</i>	32
2.5.2	<i>Central limit theorem</i>	34
2.6	<i>Density <b>of</b> a transformation</i>	35
2.7	<i>Higher-order statistics</i>	36
2.7.1	<i>Kurtosis and classification <b>of</b> densities</i>	37
2.7.2	<i>Cumulants, moments, and their properties</i>	40
2.8	<i>Stochastic processes *</i>	43
2.8.1	<i>Introduction and definition</i>	43
2.8.2	<i><b>Stationarity</b>, mean, and autocorrelation</i>	45
2.8.3	<i>Wide-sense stationary processes</i>	46
2.8.4	<i>Time averages and ergodicity</i>	48
2.8.5	<i>Power spectrum</i>	49
2.8.6	<i>Stochastic signal models</i>	50
2.9	<i>Concluding remarks and references</i>	51
	<i>Problems</i>	52
3	<i>Gradients and Optimization Methods</i>	57
3.1	<i>Vector and matrix gradients</i>	57
3.1.1	<i>Vector gradient</i>	57
3.1.2	<i>Matrix gradient</i>	59
3.1.3	<i>Examples <b>of</b> gradients</i>	59

3.1.4	<i>Taylor series expansions</i>	62
3.2	<i>Learning rules for unconstrained optimization</i>	63
3.2.1	<i>-Gradient descent</i>	63
3.2.2	<i>Second-order learning</i>	65
3.2.3	<i>The natural gradient and relative gradient</i>	67
3.2.4	<i>Stochastic gradient descent</i>	68
3.2.5	<i>Convergence <b>of</b> stochastic on-line algorithms *</i>	71
3.3	<i>Learning rules <b>for</b> constrained optimization</i>	73
3.3.1	<i>The Lagrange method</i>	73
3.3.2	<i>Projection methods</i>	73
3.4	<i>Concluding remarks and references</i>	75
	<i>Problems</i>	75
4	<i>Estimation Theory</i>	77
4.1	<i>Basic concepts</i>	78
4.2	<i>Properties <b>of</b> estimators</i>	80
4.3	<i>Method <b>of</b> moments</i>	84
4.4	<i>'Least-squares estimation</i>	86
4.4.1	<i>Linear least-squares method</i>	86
4.4.2	<i>Nonlinear and generalized least squares *</i>	88
4.5	<i>Maximum likelihood method</i>	90
4.6	<i>Bayesian estimation *</i>	94
4.6.1	<i>Minimum mean-square error estimator</i>	94
4.6.2	<i>Wiener filtering</i>	96
4.6.3	<i>Maximum a posteriori (MAP) estimator</i>	97
4.7	<i>Concluding remarks and references</i>	99
	<i>Problems</i>	101
5	<i>Information Theory</i>	105
5.1	<i>Entropy</i>	105
5.1.1	<i>Definition <b>of</b> entropy</i>	105
5.1.2	<i>Entropy and coding length</i>	107
5.1.3	<i>Differential entropy</i>	108
5.1.4	<i>Entropy <b>of</b> a transformation</i>	109
5.2	<i>Mutual information</i>	110
5.2.1	<i>Definition using entropy</i>	110
5.2.2	<i>Definition using Kullback-Leibler divergence</i>	110

5.3	<i>Maximum entropy</i>	111
5.3.1	<i>Maximum entropy distributions</i>	111
5.3.2	<i>Maximality property <b>of gaussian</b> distribution</i>	112
5.4	<i>Negen tropy</i>	112
5.5	<i>Approximation <b>of</b> entropy by cumulants</i>	113
5.5.1	<i>Polynomial density expansions</i>	113
5.5.2	<i>Using expansions <b>for</b> entropy approximation</i>	114
5.6	<i>Approximation <b>of</b> entropy by nonpolynomial functions</i>	115
5.6.1	<i>Approximating the maximum entropy</i>	116
5.6.2	<i>Choosing the nonpolynomial functions</i>	117
5.6.3	<i>Simple special cases</i>	118
5.6.4	<i>Illustration</i>	119
5.7	<i>Concluding remarks and references</i>	120
	<i>Problems</i>	121
	<i>Appendix proofs</i>	122
6	<i>Principal Component Analysis and Whitening</i>	125
6.1	<i>Principal components</i>	125
6.1.1	<i>PCA by variance maximization</i>	127
6.1.2	<i>PCA by minimum <b>MSE</b> compression</i>	128
6.1.3	<i>Choosing the number <b>of</b> principal components</i>	129
6.1.4	<i>Closed-form computation <b>of</b> PCA</i>	131
6.2	<i>PCA by on-line learning</i>	132
6.2.1	<i>The stochastic gradient ascent algorithm</i>	133
6.2.2	<i>The <b>subspace</b> learning algorithm</i>	134
6.2.3	<i>The <b>PAST</b> algorithm *</i>	135
6.2.4	<i>PCA and back-propagation learning *</i>	136
6.2.5	<i>Extensions <b>of</b> PCA to nonquadratic criteria *</i>	137
6.3	<i>Factor analysis</i>	138
6.4	<i>Whitening</i>	140
6.5	<i>Orthogonalization</i>	141
6.6	<i>Concluding remarks and references</i>	143
	<i>Problems</i>	144

## Part II BASIC INDEPENDENT COMPONENT ANALYSIS

7	<i>What is Independent Component Analysis?</i>	147
7.1	<i>Motivation</i>	147
7.2	<i>Definition <b>of</b> independent component analysis</i>	151
7.2.1	<i>ICA as estimation <b>of</b> a generative model</i>	151
7.2.2	<i>Restrictions in ICA</i>	152
7.2.3	<i>Ambiguities <b>of</b> ICA</i>	154
7.2.4	<i>Centering the variables</i>	154
7.3	<i>Illustration <b>of</b> ICA</i>	155
7.4	<i>ICA is stronger than whitening</i>	158
7.4.1	<i>Uncorrelatedness and whitening</i>	158
7.4.2	<i>Whitening is only half ICA</i>	160
7.5	<i>Why <b>gaussian</b> variables are forbidden</i>	161
7.6	<i>Concluding remarks and references</i>	163
	<i>Problems</i>	164
8	<i>ICA by Maximization <b>of</b> Nongaussianity</i>	165
8.1	<i>“Nongaussian is independent,”</i>	166
8.2	<i>Measuring nongaussianity by kurtosis</i>	171
8.2.1	<i>Extrema give independent components</i>	171
8.2.2	<i>Gradient algorithm using kurtosis</i>	175
8.2.3	<i>A fast fixed-point algorithm using kurtosis</i>	178
8.2.4	<i>Examples</i>	179
8.3	<i>Measuring nongaussianity by negentropy</i>	182
8.3.1	<i>Critique <b>of</b> kurtosis</i>	182
8.3.2	<i>Negentropy as nongaussianity measure</i>	182
8.3.3	<i>Approximating negentropy</i>	183
8.3.4	<i>Gradient algorithm using negentropy</i>	185
8.3.5	<i>A fast fixed-point algorithm using negentropy</i>	188
8.4	<i>Estimating several independent components</i>	192
8.4.1	<i>Constraint <b>of</b> uncorrelatedness</i>	192
8.4.2	<i>Deflationary orthogonalization</i>	194
8.4.3	<i>Symmetric orthogonalization</i>	194
8.5	<i>ICA and projection pursuit</i>	197
8.5.1	<i>Searching <b>for</b> interesting directions</i>	197
8.5.2	<i>Nongaussian is interesting</i>	197
8.6	<i>Concluding remarks and references</i>	198

	<i>Problems</i>	199
	<i>Appendix proofs</i>	201
9	<i>ICA by Maximum Likelihood Estimation</i>	203
9.1	<i>The likelihood <b>of</b> the ICA model</i>	203
9.1.1	<i>Deriving the likelihood</i>	203
9.1.2	<i>Estimation <b>of</b> the densities</i>	204
9.2	<i>Algorithms <b>for</b> maximum likelihood estimation</i>	207
9.2.1	<i>Gradient algorithms</i>	207
9.2.2	<i>A fast fixed-point algorithm</i>	209
9.3	<i>The infomax principle</i>	211
9.4	<i>Examples</i>	213
9.5	<i>Concluding remarks and references</i>	214
	<i>Problems</i>	218
	<i>Appendix proofs</i>	219
10	<i>ICA by Minimization <b>of</b> Mutual Information</i>	221
10.1	<i>Defining ICA by mutual information</i>	221
10.1.1	<i>Information-theoretic concepts</i>	221
10.1.2	<i>Mutual information as measure <b>of</b> dependence</i>	222
10.2	<i>Mutual information and nongaussianity</i>	223
10.3	<i>Mutual information and likelihood</i>	224
10.4	<i>Algorithms <b>for</b> minimization <b>of</b> mutual information</i>	224
10.5	<i>Examples</i>	225
10.6	<i>Concluding remarks and references</i>	225
	<i>Problems</i>	227
11	<i>ICA by Tensorial Methods</i>	229
11.1	<i>Definition <b>of</b> cumulant tensor</i>	229
11.2	<i>Tensor eigenvalues give independent components</i>	230
11.3	<i>Tensor decomposition by a power method</i>	232
11.4	<i>Joint approximate diagonalization <b>of</b> eigenmatrices</i>	234
11.5	<i>Weighted correlation matrix approach</i>	235
11.5.1	<i>The <b>FOBI</b> algorithm</i>	235
11.5.2	<i>From <b>FOBI</b> to <b>JADE</b></i>	235
11.6	<i>Concluding remarks and references.</i>	236
	<i>Problems</i>	237

12	<i>ICA by Nonlinear Decorrelation and Nonlinear PCA</i>	239
12.1	<i>Nonlinear correlations and independence</i>	240
12.2	<i>The Héroult-Jutten algorithm</i>	242
12.3	<i>The Cichocki-Unbehauen algorithm</i>	243
12.4	<i>The estimating functions approach *</i>	245
12.5	<i>Equivariant adaptive separation via independence</i>	247
12.6	<i>Nonlinear principal components</i>	249
12.7	<i>The nonlinear PCA criterion and ICA</i>	251
12.8	<i>Learning rules for the nonlinear PCA criterion</i>	254
12.8.1	<i>The nonlinear subspace rule</i>	254
12.8.2	<i>Convergence of the nonlinear subspace rule *</i>	255
12.8.3	<i>Nonlinear recursive least-squares rule</i>	258
12.9	<i>Concluding remarks and references</i>	261
	<i>Problems</i>	262
13	<i>Practical Considerations</i>	263
13.1	<i>Preprocessing by time filtering</i>	263
13.1.1	<i>Why time filtering is possible</i>	264
13.1.2	<i>Low-pass filtering</i>	265
13.1.3	<i>High-pass filtering and innovations</i>	265
13.1.4	<i>Optimal filtering</i>	266
13.2	<i>Preprocessing by PCA</i>	267
13.2.1	<i>Making the mixing matrix square</i>	267
13.2.2	<i>Reducing noise and preventing overlearning</i>	268,
13.3	<i>How many components should be estimated?</i>	269
13.4	<i>Choice of algorithm</i>	271
13.5	<i>Concluding remarks and references</i>	272
	<i>Problems</i>	272
14	<i>Overview and Comparison of Basic ICA Methods</i>	273
14.1	<i>Objective functions vs. algorithms</i>	273
14.2	<i>Connections between ICA estimation principles</i>	274
14.2.1	<i>Similarities between estimation principles</i>	274
14.2.2	<i>Differences between estimation principles</i>	275
14.3	<i>Statistically optimal nonlinearities</i>	276
14.3.1	<i>Comparison of asymptotic variance *</i>	276
14.3.2	<i>Comparison of robustness *</i>	277
14.3.3	<i>Practical choice of nonlinearity</i>	279

14.4	<i>Experimental comparison of ICA algorithms</i>	280
14.4.1	<i>Experimental set-up and algorithms</i>	281
14.4.2	<i>Results for simulated data</i>	282
14.4.3	<i>Comparisons with real-world data</i>	286
14.5	<i>References</i>	287
14.6	<i>Summary of basic ICA</i>	287
	<i>Appendix Proofs</i>	289

### Part III EXTENSIONS AND RELATED METHODS

15	<i>Noisy ICA</i>	293
15.1	<i>Definition</i>	293
15.2	<i>Sensor noise vs. source noise</i>	294
15.3	<i>Few noise sources</i>	295
15.4	<i>Estimation of the mixing matrix</i>	295
15.4.1	<i>Bias removal techniques</i>	296
15.4.2	<i>Higher-order cumulant methods</i>	298
15.4.3	<i>Maximum likelihood methods</i>	299
15.5	<i>Estimation of the noise-free independent components</i>	299
15.5.1	<i>Maximum a posteriori estimation</i>	299
15.5.2	<i>Special case of shrinkage estimation</i>	300
15.6	<i>Denoising by sparse code shrinkage</i>	303
15.7	<i>Concluding remarks</i>	304
16	<i>ICA with Overcomplete Bases</i>	305
16.1	<i>Estimation of the independent components</i>	306
16.1.1	<i>Maximum likelihood estimation</i>	306
16.1.2	<i>The case of supergaussian components</i>	307
16.2	<i>Estimation of the mixing matrix</i>	307
16.2.1	<i>Maximizing joint likelihood</i>	307
16.2.2	<i>Maximizing likelihood approximations</i>	308
16.2.3	<i>Approximate estimation by quasiorthogonality</i>	309
16.2.4	<i>Other approaches</i>	311
16.3	<i>Concluding remarks</i>	313



<i>17 Nonlinear ICA</i>	315
17.1 <i>Nonlinear ICA and BSS</i>	315
17.1.1 <i>The nonlinear ICA and BSS problems</i>	315
17.1.2 <i>Existence and uniqueness of nonlinear ICA</i>	317
17.2 <i>Separation of post-nonlinear mixtures</i>	319
17.3 <i>Nonlinear BSS using self-organizing maps</i>	320
17.4 <i>A generative topographic mapping approach *</i>	322
17.4.1 <i>Background</i>	322
17.4.2 <i>The modified GTM method</i>	323
17.4.3 <i>An experiment</i>	326
17.5 <i>An ensemble learning approach to nonlinear BSS</i>	328
17.5.1 <i>Ensemble learning</i>	328
17.5.2 <i>Model structure</i>	329
17.5.3 <i>Computing Kullback-Leibler cost function *</i>	330
17.5.4 <i>Learning procedure *</i>	332
17.5.5 <i>Experimental results</i>	333
17.6 <i>Other approaches</i>	337
17.7 <i>Concluding remarks</i>	339
<i>18 Methods using Time Structure</i>	341
18.1 <i>Separation by autocovariances</i>	342
18.1.1 <i>An alternative to nongaussianity</i>	342
18.1.2 <i>Using one time lag</i>	343
18.1.3 <i>Extension to several time lags</i>	344
18.2 <i>Separation by nonstationarity of variances</i>	346
18.2.1 <i>Using local autocorrelations</i>	347
18.2.2 <i>Using cross-cumulants</i>	349
18.3 <i>Separation principles unified</i>	351
18.3.1 <i>Comparison of separation principles</i>	351
18.3.2 <i>Kolmogoroff complexity as unifying framework</i>	352
18.4 <i>Concluding remarks</i>	354

19	<i>Convolutional Mixtures and Blind Deconvolution</i>	355
19.1	<i>Blind deconvolution</i>	356
19.1.1	<i>Problem definition</i>	356
19.1.2	<i>Bussgang methods</i>	357
19.1.3	<i>Cumulant-based methods</i>	358
19.1.4	<i>Blind deconvolution using linear ICA</i>	360
19.2	<i>Blind separation of convolutional mixtures</i>	361
19.2.1	<i>The convolutional BSS problem</i>	361
19.2.2	<i>Reformulation as ordinary ICA</i>	363
19.2.3	<i>Natural gradient methods</i>	364
19.2.4	<i>Fourier transform methods</i>	365
19.2.5	<i>Spatiotemporal decorrelation methods</i>	367
19.2.6	<i>Other methods for convolutional mixtures</i>	367
19.3	<i>Concluding remarks</i>	368
	<i>Appendix Discrete-time filters and the z-transform</i>	369
20	<i>Other Extensions</i>	371
20.1	<i>Priors on the mixing matrix</i>	371
20.1.1	<i>Motivation for prior information</i>	371
20.1.2	<i>Classic priors</i>	372
20.1.3	<i>Sparse priors</i>	374
20.1.4	<i>Spatiotemporal ICA</i>	377
20.2	<i>Relaxing the independence assumption</i>	378
20.2.1	<i>Multidimensional ICA</i>	379
20.2.2	<i>Independent subspace analysis</i>	380
20.2.3	<i>Topographic ICA</i>	382
20.3	<i>Complex-valued data</i>	383
20.3.1	<i>Basic concepts of complex random variables</i>	383
20.3.2	<i>Indeterminacy of the independent components</i>	384
20.3.3	<i>Choice of the non-Gaussianity measure</i>	385
20.3.4	<i>Consistency of estimator</i>	386
20.3.5	<i>Fixed-point algorithm</i>	386
20.3.6	<i>Relation to independent subspaces</i>	387
20.4	<i>Concluding remarks</i>	387

## Part IV APPLICATIONS OF ICA

2.1	<i>Feature Extraction by ICA</i>	391
21.1	<i>Linear representations</i>	392
21.1.1	<i>Definition</i>	392
21.1.2	<i>Gabor analysis</i>	392
21.1.3	<i>Wavelets</i>	394
21.2	<i>ICA and Sparse Coding</i>	396
21.3	<i>Estimating ICA bases from images</i>	398
21.4	<i>Image denoising by sparse code shrinkage</i>	398
21.4.1	<i>Component statistics</i>	399
21.4.2	<i>Remarks on windowing</i>	400
21.4.3	<i>Denoising results</i>	401
21.5	<i>Independent subspaces and topographic ICA</i>	401
21.6	<i>Neurophysiological connections</i>	403
21.7	<i>Concluding remarks</i>	405
2.2	<i>Brain Imaging Applications</i>	407
22.1	<i>Electro- and magnetoencephalography</i>	407
22.1.1	<i>Classes of brain imaging techniques</i>	407
22.1.2	<i>Measuring electric activity in the brain</i>	408
22.1.3	<i>Validity of the basic ICA model</i>	409
22.2	<i>Artifact identification from EEG and MEG</i>	410
22.3	<i>Analysis of evoked magnetic fields</i>	411
22.4	<i>ICA applied on other measurement techniques</i>	413
22.5	<i>Concluding remarks</i>	414
2.3	<i>Telecommunications</i>	417
23.1	<i>Multiuser detection and CDMA communications</i>	417
23.2	<i>CDMA signal model and ICA</i>	422
23.3	<i>Estimating fading channels</i>	424
23.3.1	<i>Minimization of complexity</i>	424
23.3.2	<i>Channel estimation *</i>	426
23.3.3	<i>Comparisons and discussion</i>	428
23.4	<i>Blind separation of convolved CDMA mixtures *</i>	430
23.4.1	<i>Feedback architecture</i>	430
23.4.2	<i>Semiblind separation method</i>	431
23.4.3	<i>Simulations and discussion</i>	432

23.5	<i>Improving multiuser detection using complex ICA *</i>	434
23.5.1	<i>Data model</i>	435
23.5.2	<i>ICA based receivers</i>	436
23.5.3	<i>Simulation results</i>	438
23.6	<i>Concluding remarks and references</i>	439
24	<i>Other Applications</i>	441
24.1	<i>Financial applications</i>	441
24.1.1	<i>Finding hidden factors in financial data</i>	441
24.1.2	<i>Time series prediction by ICA</i>	443
24.2	<i>Audio separation</i>	446
24.3	<i>Further applications</i>	448
	<i>References</i>	449
	<i>Index</i>	476