

# Contents

|                    |     |
|--------------------|-----|
| Contributors ..... | XV  |
| Foreword .....     | XIX |

---

## Part I Temporal ICA Models

---

### 1 Hidden Markov Independent Component Analysis

|   |    |
|---|----|
| William D. Penny, Richard M. Everson and Stephen J. Roberts ..... | 3  |
| 1.1 Introduction .....  | 3  |
| 1.2 Hidden Markov Models .....                                    | 3  |
| 1.3 Independent Component Analysis .....                          | 6  |
| 1.3.1 Generalised Exponential Sources .....                       | 6  |
| 1.3.2 Generalised Autoregressive Sources .....                    | 7  |
| 1.4 Hidden Markov ICA .....                                       | 8  |
| 1.4.1 Generalised Exponential Sources .....                       | 9  |
| 1.4.2 Generalised Autoregressive Sources .....                    | 10 |
| 1.5 Practical Issues .....  | 10 |
| 1.5.1 Initialisation .....  | 10 |
| 1.5.2 Learning .....  | 10 |
| 1.5.3 Model Order Selection .....                                 | 12 |
| 1.6 Results .....   | 12 |
| 1.6.1 Multiple Sinewave Sources .....                             | 12 |
| 1.6.2 Same Sources, Different Mixing .....                        | 14 |
| 1.6.3 Same Mixing, Different Sources .....                        | 16 |
| 1.6.4 EEG Data .....  | 16 |
| 1.7 Conclusion .....  | 19 |
| 1.8 Acknowledgements .....  | 20 |
| 1.9 Appendix .....  | 20 |

### 2 Particle Filters for Non-Stationary ICA

|   |    |
|---|----|
| Richard M. Everson and Stephen J. Roberts .....         | 23 |
| 2.1 Introduction .....                                  | 23 |
| 2.2 Stationary ICA .....                                | 23 |
| 2.3 Non-Stationary Independent Component Analysis ..... | 25 |

|       |  |    |
|-------|--|----|
| 2.3.1 | Source Model   | 27 |
| 2.4   | Particle Filters                                     | 28 |
| 2.4.1 | Source Recovery                                      | 29 |
| 2.5   | Illustration of Non-Stationary ICA                   | 30 |
| 2.6   | Smoothing  | 33 |
| 2.7   | Temporal Correlations                                | 36 |
| 2.8   | Conclusion   | 38 |
| 2.8.1 | Acknowledgement                                      | 38 |
| 2.9   | Appendix: Laplace's Approximation for the Likelihood | 39 |

---

## Part II The Validity of the Independence Assumption

---

### 3 The Independence Assumption: Analyzing the Independence of the Components by Topography

|   |   |    |
|---|---|----|
| Aapo Hyvärinen, Patrik O. Hoyer and Mika Inki | 45  |    |
| 3.1   | Introduction  | 45 |
| 3.2   | Background: Independent Subspace Analysis           | 47 |
| 3.3   | Topographic ICA Model                               | 49 |
| 3.3.1   | Dependence and Topography                           | 49 |
| 3.3.2   | Defining Topographic ICA                            | 50 |
| 3.3.3   | The Generative Model                                | 51 |
| 3.3.4   | Basic Properties of the Topographic ICA Model       | 52 |
| 3.4   | Learning Rule                                       | 53 |
| 3.5   | Comparison with Other Topographic Mappings          | 54 |
| 3.6   | Experiments   | 55 |
| 3.6.1   | Experiments in Feature Extraction of Image Data     | 55 |
| 3.6.2   | Experiments in Feature Extraction of Audio Data     | 57 |
| 3.6.3   | Experiments with Magnetoencephalographic Recordings | 58 |
| 3.7   | Conclusion  | 59 |

### 4 The Independence Assumption: Dependent Component Analysis

|                     |  |    |
|---------------------|--|----|
| Allan Kardec Barros | 63   |    |
| 4.1                 | Introduction                                   | 63 |
| 4.2                 | Blind Source Separation by DCA                 | 64 |
| 4.3                 | The "Cyclone" Algorithm                        | 65 |
| 4.4                 | Experimental Results                           | 67 |
| 4.5                 | Higher-Order Cyclostationary Signal Separation | 68 |
| 4.6                 | Conclusion                                     | 68 |
| 4.7                 | Appendix: Proof of ACF Property 3              | 70 |

---

**Part III Ensemble Learning and Applications**


---

**5 Ensemble Learning**

|  |    |
|--|----|
| Harri Lappalainen and James W. Miskin .....      | 75 |
| 5.1 Introduction .....                           | 75 |
| 5.2 Posterior Averages in Action .....           | 76 |
| 5.3 Approximations of Posterior PDF .....        | 78 |
| 5.4 Ensemble Learning .....                      | 79 |
| 5.4.1 Model Selection in Ensemble Learning ..... | 81 |
| 5.4.2 Connection to Coding .....                 | 82 |
| 5.4.3 EM and MAP .....                           | 83 |
| 5.5 Construction of Probabilistic Models .....   | 83 |
| 5.5.1 Priors and Hyperpriors .....               | 85 |
| 5.6 Examples .....                               | 86 |
| 5.6.1 Fixed Form $Q$ .....                       | 86 |
| 5.6.2 Free Form $Q$ .....                        | 88 |
| 5.7 Conclusion .....                             | 91 |
| References .....                                 | 92 |

**6 Bayesian Non-Linear Independent Component Analysis by Multi-Layer Perceptrons**

|  |     |
|--|-----|
| Harri Lappalainen and Antti Honkela .....        | 93  |
| 6.1 Introduction .....                           | 93  |
| 6.2 Choosing Among Competing Explanations .....  | 95  |
| 6.3 Non-Linear Factor Analysis .....             | 97  |
| 6.3.1 Definition of the Model .....              | 97  |
| 6.3.2 Cost Function .....                        | 99  |
| 6.3.3 Update Rules .....                         | 102 |
| 6.4 Non-Linear Independent Factor Analysis ..... | 106 |
| 6.5 Experiments .....                            | 107 |
| 6.5.1 Learning Scheme .....                      | 107 |
| 6.5.2 Helix .....                                | 108 |
| 6.5.3 Non-Linear Artificial Data .....           | 109 |
| 6.5.4 Process Data .....                         | 115 |
| 6.6 Comparison with Existing Methods .....       | 116 |
| 6.6.1 SOM and GTM .....                          | 116 |
| 6.6.2 Auto-Associative MLPs .....                | 117 |
| 6.6.3 Generative Learning with MLPs .....        | 118 |
| 6.7 Conclusion .....                             | 118 |
| 6.7.1 Validity of the Approximations .....       | 118 |
| 6.7.2 Initial Inversion by Auxiliary MLP .....   | 119 |
| 6.7.3 Future Directions .....                    | 120 |
| 6.8 Acknowledgements .....                       | 120 |

## 7 Ensemble Learning for Blind Image Separation and Deconvolution

|   |     |
|---|-----|
| James Miskin and David J. C. MacKay .....   | 123 |
| 7.1 Introduction .....                      | 123 |
| 7.2 Separation of Images .....              | 124 |
| 7.2.1 Learning the Ensemble .....           | 126 |
| 7.2.2 Learning the Model .....              | 129 |
| 7.2.3 Example .....                         | 129 |
| 7.2.4 Parts-Based Image Decomposition ..... | 132 |
| 7.3 Deconvolution of Images .....           | 134 |
| 7.4 Conclusion .....                        | 140 |
| 7.5 Acknowledgements .....                  | 141 |
| References .....                            | 141 |

---

## Part IV Data Analysis and Applications

---

### 8 Multi-Class Independent Component Analysis (MUCICA) for Rank-Deficient Distributions

|  |     |
|--|-----|
| Francesco Palmieri and Alessandra Budillon ..... | 145 |
| 8.1 Introduction .....                           | 145 |
| 8.2 The Rank-Deficient One Class Problem .....   | 146 |
| 8.2.1 Method I: Three Blocks .....               | 148 |
| 8.2.2 Method II : Two Blocks .....               | 149 |
| 8.2.3 Method III : One Block .....               | 150 |
| 8.3 The Rank-Deficient Multi-Class Problem ..... | 151 |
| 8.4 Simulations .....                            | 154 |
| 8.5 Conclusion .....                             | 158 |
| References .....                                 | 159 |

### 9 Blind Separation of Noisy Image Mixtures

|   |     |
|---|-----|
| Lars Kai Hansen .....   | 161 |
| 9.1 Introduction .....  | 161 |
| 9.2 The Likelihood .....  | 162 |
| 9.3 Estimation of Sources for the Case of Known Parameters .....      | 163 |
| 9.4 Joint Estimation of Sources, Mixing Matrix, and Noise Level ..... | 164 |
| 9.5 Simulation Example .....  | 166 |
| 9.6 Generalization and the Bias-Variance Dilemma .....                | 167 |
| 9.7 Application to Neuroimaging .....                                 | 170 |
| 9.8 Conclusion .....  | 175 |
| 9.9 Acknowledgments .....   | 178 |
| 9.10 Appendix: The Generalized Boltzmann Learning Rule .....          | 179 |

## 10 Searching for Independence in Electromagnetic Brain Waves

|  |     |
|--|-----|
| Ricardo Vigário, Jaakko Särelä and Erkki Oja .....         | 183 |
| 10.1 Introduction .....                                    | 183 |
| 10.2 Independent Component Analysis .....                  | 184 |
| 10.2.1 The Model .....                                     | 184 |
| 10.2.2 The FastICA Algorithm .....                         | 184 |
| 10.3 Electro- and Magnetoencephalography .....             | 186 |
| 10.4 On the Validity of the Linear ICA Model .....         | 188 |
| 10.5 The Analysis of EEG and MEG Data .....                | 189 |
| 10.5.1 Artifact Identification and Removal from EEG/MEG .. | 189 |
| 10.5.2 Analysis of Multimodal Evoked Fields .....          | 191 |
| 10.5.3 Segmenting Auditory Evoked Fields .....             | 193 |
| 10.6 Conclusion .....                                      | 194 |

## 11 ICA on Noisy Data: A Factor Analysis Approach

|   |     |
|---|-----|
| Shiro Ikeda.....                                    | 201 |
| 11.1 Introduction .....                             | 201 |
| 11.2 Factor Analysis and ICA .....                  | 202 |
| 11.2.1 Factor Analysis .....                        | 202 |
| 11.2.2 Factor Analysis in Preprocessing .....       | 204 |
| 11.2.3 ICA as Determining the Rotation Matrix ..... | 204 |
| 11.3 Experiment with Synthesized Data .....         | 205 |
| 11.4 MEG Data Analysis .....                        | 208 |
| 11.4.1 Experiment with Phantom Data .....           | 209 |
| 11.4.2 Experiment with Real Brain Data .....        | 211 |
| 11.5 Conclusion .....                               | 213 |
| 11.6 Acknowledgements .....                         | 215 |

## 12 Analysis of Optical Imaging Data Using Weak Models and ICA

|  |     |
|--|-----|
| John Porrill, James V. Stone, Jason Berwick, John Mayhew<br>and Peter Coffey ..... | 217 |
| 12.1 Introduction .....  | 217 |
| 12.2 Linear Component Analysis.....  | 218 |
| 12.3 Singular Value Decomposition .....  | 219 |
| 12.3.1 SVD Applied to OI Data Set .....  | 220 |
| 12.4 Independent Component Analysis .....  | 221 |
| 12.4.1 Minimisation Routines.....  | 223 |
| 12.4.2 Application of sICA to OI Data.....   | 223 |
| 12.5 The Weak Causal Model .....   | 225 |
| 12.5.1 Weak Causal Model Applied to the OI Data Set .....                          | 226 |
| 12.5.2 Some Remarks on Significance Testing .....                                  | 227 |
| 12.6 The Weak Periodic Model .....   | 227 |

|   |     |
|---|-----|
| 12.7 Regularised Weak Models .....  | 228 |
| 12.8 Regularised Weak Causal Model Applied to OI Data .....               | 229 |
| 12.9 Image Goodness and Multiple Models .....                             | 230 |
| 12.10 A Last Look at the OI Data Set .....                                | 231 |
| 12.11 Conclusion .....  | 232 |
| References .....  | 233 |
| <b>13 Independent Components in Text</b>                                  |     |
| Thomas Kolenda, Lars Kai Hansen and Sigurdur Sigurdsson .....             | 235 |
| 13.1 Introduction .....   | 235 |
| 13.1.1 Vector Space Representations .....                                 | 235 |
| 13.1.2 Latent Semantic Indexing .....                                     | 237 |
| 13.2 Independent Component Analysis .....                                 | 238 |
| 13.2.1 Noisy Separation of Linear Mixtures .....                          | 239 |
| 13.2.2 Learning ICA Text Representations on the LSI Space ..              | 242 |
| 13.2.3 Document Classification Based on Independent<br>Components .....   | 243 |
| 13.2.4 Keywords from Context Vectors .....                                | 244 |
| 13.2.5 Generalisation and the Bias-Variance Dilemma .....                 | 244 |
| 13.3 Examples .....   | 246 |
| 13.3.1 MED Data Set .....   | 248 |
| 13.3.2 CRAN Data Set .....  | 249 |
| 13.4 Conclusion .....   | 251 |
| <b>14 Seeking Independence Using Biologically-Inspired ANN's</b>          |     |
| Pei Ling Lai, Darryl Charles and Colin Fyfe .....                         | 257 |
| 14.1 Introduction .....   | 257 |
| 14.2 The Negative Feedback Network .....                                  | 258 |
| 14.3 Independence in Unions of Sources .....                              | 259 |
| 14.3.1 Factor Analysis .....  | 261 |
| 14.3.2 Minimal Overcomplete Bases .....                                   | 261 |
| 14.4 Canonical Correlation Analysis .....                                 | 264 |
| 14.4.1 Extracting Multiple Correlations .....                             | 266 |
| 14.4.2 Using Minimum Correlations to Extract Independent<br>Sources ..... | 267 |
| 14.4.3 Experiments .....  | 268 |
| 14.5 $\epsilon$ -Insensitive Hebbian Learning .....                       | 269 |
| 14.5.1 Is this a Hebbian Rule? .....                                      | 270 |
| 14.5.2 Extraction of Sinusoids .....                                      | 271 |
| 14.5.3 Noise Reduction .....  | 273 |
| 14.6 Conclusion .....   | 275 |
| References .....  | 275 |
| <b>Index</b> .....  | 277 |