

The EM Algorithm and Extensions

GEOFFREY J. McLACHLAN

THRIYAMBAKAM KRISHNAN



A Wiley-Interscience Publication

JOHN WILEY & SONS, INC.

New York • Chichester • Brisbane • Toronto • Singapore • Weinheim

Contents

Preface

xiii

1. General Introduction

1

- 1.1 Introduction, 1
- 1.2 Maximum Likelihood Estimation, 3
- 1.3 Newton-Type Methods, 5
 - 1.3.1 Introduction, 5
 - 1.3.2 Newton-Raphson Method, 5
 - 1.3.3 Quasi-Newton Methods, 6
 - 1.3.4 Modified Newton Methods, 7
- 1.4 Introductory Examples, 9
 - 1.4.1 Introduction, 9
 - 1.4.2 Example 1.1: A Multinomial Example, 9
 - 1.4.3 Example 1.2: Estimation of Mixing Proportions, 16
- 1.5 Formulation of the EM Algorithm, 21
 - 1.5.1 EM Algorithm, 21
 - 1.5.2 Example 1.3: Censored Exponentially Distributed Survival Times, 23
 - 1.5.3 E- and M-Steps for Regular Exponential Family, 26
 - 1.5.4 Example 1.4: Censored Exponentially Distributed Survival Times (Example 1.3 Continued), 27
 - 1.5.5 Generalized EM Algorithm, 28
 - 1.5.6 GEM Algorithm Based on One Newton–Raphson Step, 28
 - 1.5.7 EM Gradient Algorithm, 30
 - 1.5.8 EM Mapping, 30
- 1.6 EM Algorithm for Maximum *a Posteriori* and Maximum Penalized Likelihood Estimation, 30
 - 1.6.1 Maximum *a Posteriori* Estimation, 30
 - 1.6.2 Example 1.5: A Multinomial Example (Example 1.1 , Continued), 31

vii

1.6.3	Maximum Penalized Likelihood Estimation, 32	
1.7	Brief Summary of the Properties of the EM Algorithm, 32	
1.8	History of the EM Algorithm, 34	
1.8.1	Work Before Dempster, Laird, and Rubin (1977), 34	
1.8.2	EM Examples and Applications Since Dempster, Laird, and Rubin (1977), 36	
1.8.3	Two Interpretations of EM, 37	
1.8.4	Developments in EM Theory, Methodology, and Applications, 38	
1.9	Overview of the Book, 41	
1.10	Notation, 42	
2.	Examples of the EM Algorithm	45
2.1	Introduction, 45	
2.2	Multivariate Data with Missing Values, 45	
2.2.1	Example 2.1: Bivariate Normal Data with Missing Values, 45	
2.2.2	Numerical Illustration, 49	
2.2.3	Multivariate Data: Buck's Method, 50	
2.3	Least Squares with Missing Data, 51	
2.3.1	Healy–Westmacott Procedure, 51	
2.3.2	Example 2.2: Linear Regression with Missing Dependent Values, 52	
2.3.3	Example 2.3: Missing Values in a Latin Square Design, 54	
2.3.4	Healy–Westmacott Procedure as an EM Algorithm, 55	
2.4	Example 2.4: Multinomial with Complex Cell Structure, 57	
2.5	Example 2.5: Analysis of PET and SPECT Data, 60	
2.6	Example 2.6: Multivariate t -Distribution with Known Degrees of Freedom, 63	
2.6.1	ML Estimation of Multivariate t -Distribution, 63	
2.6.2	Numerical Example: Stack Loss Data, 67	
2.7	Finite Normal Mixtures, 68	
2.7.1	Example 2.7: Univariate Component Densities, 68	
2.7.2	Example 2.8: Multivariate Component Densities, 71	
2.7.3	Numerical Example: Red Blood Cell Volume Data, 72	
2.8	Example 2.9: Grouped and Truncated Data, 74	
2.8.1	Introduction, 74	
2.8.2	Specification of Complete Data, 74	
2.8.3	E-Step, 77	
2.8.4	M-Step, 78	
2.8.5	Confirmation of Incomplete-Data Score Statistic, 78	

- 2.8.6 M-Step for Grouped Normal Data, 79
- 2.8.7 Numerical Example: Grouped Log Normal Data, 80

3. Basic Theory of the EM Algorithm 82

- 3.1 Introduction, 82
- 3.2 Monotonicity of the EM Algorithm, 82
- 3.3 Monotonicity of a Generalized EM Algorithm, 84
- 3.4 Convergence of an EM Sequence to a Stationary Value, 84
 - 3.4.1 Introduction, 84
 - 3.4.2 Regularity Conditions of Wu, 86
 - 3.4.3 Main Convergence Theorem for a Generalized EM Sequence, 87
 - 3.4.4 A Convergence Theorem for an EM Sequence, 88
- 3.5. Convergence of an EM Sequence of Iterates, 89
 - 3.5.1 Introduction, 89
 - 3.5.2 Two Convergence Theorems of Wu, 89
 - 3.5.3 Convergence of an EM Sequence to a Unique Maximum Likelihood Estimate, 91
- 3.6 Examples of Nontypical Behavior of an EM (GEM) Sequence, 91
 - 3.6.1 Example 3.1: Convergence to a Saddle Point, 91
 - 3.6.2 Example 3.2: Convergence to a Local Minimum, 94
 - 3.6.3 Example 3.3: Nonconvergence of a Generalized EM Sequence, 97
- 3.7 Score Statistic, 99
- 3.8 Missing Information, 100
 - 3.8.1 Missing Information Principle, 100
 - 3.8.2 Example 3.4 (Example 1.3 Continued), 102
- 3.9 Rate of Convergence of the EM Algorithm, 105
 - 3.9.1 Rate Matrix for Linear Convergence, 105
 - 3.9.2 Measuring the Linear Rate of Convergence, 106
 - 3.9.3 Rate Matrix in Terms of Information Matrices, 106
 - 3.9.4 Rate Matrix for Maximum *a Posteriori* Estimation, 108
 - 3.9.5 Derivation of Rate Matrix in Terms of Information Matrices, 108
 - 3.9.6 Example 3.5 (Example 3.4 Continued), 109

4. Standard Errors and Speeding Up Convergence 110

- 4.1 Introduction, 110
- 4.2 Observed Information Matrix, 111
 - 4.2.1 Direct Evaluation, 111
 - 4.2.2 Extraction of Observed Information Matrix in Terms of the Complete-Data Log Likelihood, 111
 - 4.2.3 Regular Case, 113

- 4.2.4 Evaluation of the Conditional Expected Complete-Data Information Matrix, 114
- 4.2.5 Examples, 115
- 4.3 Approximations to the Observed Information Matrix: i.i.d. Case, 120
- 4.4 Observed Information Matrix for Grouped Data, 123
 - 4.4.1 Approximation Based on Empirical Information, 123
 - 4.4.2 Example 4.3: Grouped Data from an Exponential Distribution, 125
- 4.5 Supplemented EM Algorithm, 128
 - 4.5.1 Definition, 128
 - 4.5.2 Calculation of $\mathbf{J}(\hat{\Psi})$ via Numerical Differentiation, 130
 - 4.5.3 Stability, 131
 - 4.5.4 Monitoring Convergence, 131
 - 4.5.5 Example 4.4: Univariate Contaminated Normal Data, 132
 - 4.5.6 Example 4.5: Bivariate Normal Data with Missing Values, 136
- 4.6 Bootstrap Approach to Standard Error Approximation, 139
- 4.7 Acceleration of the EM Algorithm via Aitken's Method, 141
 - 4.7.1 Aitken's Acceleration Method, 141
 - 4.7.2 Louis' Method, 141
 - 4.7.3 Example 4.6: Multinomial Data, 143
 - 4.7.4 Example 4.7: Geometric Mixture, 143
 - 4.7.5 Grouped and Truncated Data (Example 2.8 Continued), 147
- 4.8 An Aitken Acceleration-Based Stopping Criterion, 148
- 4.9 Conjugate Gradient Acceleration of the EM Algorithm, 149
 - 4.9.1 Conjugate Gradient Method, 149
 - 4.9.2 A Generalized Conjugate Gradient Algorithm, 150
 - 4.9.3 Accelerating the EM Algorithm, 151
- 4.10 Hybrid Methods for Finding Maximum Likelihood Estimate, 152
 - 4.10.1 Introduction, 152
 - 4.10.2 Combined EM and Modified Newton-Raphson Algorithm, 152
- 4.11 A Generalized EM Algorithm Based on One Newton-Raphson Step, 154
 - 4.11.1 Derivation of a Condition to be a Generalized EM Sequence, 154
 - 4.11.2 Simulation Experiment, 155
- 4.12 EM Gradient Algorithm, 156
- 4.13 A Quasi-Newton Acceleration of the EM Algorithm, 158
 - 4.13.1 The Method, 158
 - 4.13.2 Example 4.8: Dirichlet Distribution, 161

5. Extensions of the EM Algorithm

166

- 5.1 Introduction, 166

- 5.2 ECM Algorithm, 167
 - 5.2.1 Motivation, 167
 - 5.2.2 Formal Definition, 167
 - 5.2.3 Convergence Properties, 169
 - 5.2.4 Speed of Convergence, 170
 - 5.2.5 Discussion, 171
- 5.3 Multicycle ECM Algorithm, 171
- 5.4 Example 5.1: Normal Mixtures with Equal Correlations, 173
 - 5.4.1 Normal Components with Equal Correlations, 173
 - 5.4.2 Application of ECM Algorithm, 173
 - 5.4.3 Fisher's *Iris* Data, 175
- 5.5 Example 5.2: Mixture Models for Survival Data, 175
 - 5.5.1 Competing Risks in Survival Analysis, 175
 - 5.5.2 A Two-Component Mixture Regression Model, 176
 - 5.5.3 Observed Data, 177
 - 5.5.4 Application of EM Algorithm, 178
 - 5.5.5 M-Step for Gompertz Components, 179
 - 5.5.6 Application of a Multicycle ECM Algorithm, 180
- 5.6 Example 5.3: Contingency Tables with Incomplete Data, 181
- 5.7 ECME Algorithm, 183
- 5.8 Example 5.4: Maximum Likelihood Estimation of t -Distribution with Unknown Degrees of Freedom, 184
 - 5.8.1 Application of EM Algorithm, 184,
 - 5.8.2 M-Step, 185
 - 5.8.3 Application of the ECM Algorithm, 185
 - 5.8.4 Application of ECME Algorithm, 186
 - 5.8.5 Some Standard Results, 187
 - 5.8.6 Missing Data, 188
 - 5.8.7 Numerical Examples, 190
 - 5.8.8 Theoretical Results on the Rate of Convergence, 191
- 5.9 Example 5.5: Variance Components, 191
 - 5.9.1 A Variance Components Model, 191
 - 5.9.2 E-Step, 192,
 - 5.9.3 M-Step, 194
 - 5.9.4 Application of Two Versions of ECME Algorithm, 195
 - 5.9.5 Numerical Example, 195
- 5.10 Example 5.6: Factor Analysis, 197
 - 5.10.1 ECM Algorithm for Factor Analysis, 197
 - 5.10.2 Numerical Example, 199
- 5.11 Efficient Data Augmentation, 200
 - 5.11.1 Motivation, 200
 - 5.11.2 Maximum Likelihood Estimation of t -Distribution, 201
 - 5.11.3 Variance Components Model, 206

5.12	Alternating ECM Algorithm, 206	
5.13	EMS Algorithm, 208	
5.14	One-Step-Late Algorithm, 209	
5.15	Variance Estimation for Penalized EM and OSL Algorithms, 210	
5.15.1	Penalized EM Algorithm, 210	
5.15.2	OSL Algorithm, 211	
5.15.3	Example 5.7 (Example 4.1 Continued), 211	
5.16	Linear Inverse Problems, 213	
6.	Miscellaneous Topics	214
6.1	Iterative Simulation Algorithms, 214	
6.2	Monte Carlo E-Step, 214	
6.3	Stochastic EM Algorithm, 216	
6.4	Data Augmentation Algorithm, 218	
6.5	Multiple Imputation, 220	
6.6	Sampling-Importance Resampling, 221	
6.7	Gibbs Sampler, 222	
6.7.1	The Algorithm, 222	
6.7.2	Example 6.1: Gibbs Sampler for the Mixture Problem, 224	
6.7.3	Gibbs Sampler Analogs of ECM and ECME Algorithms, 225	
6.7.4	Some Other Analogs of the Gibbs Sampler, 226	
6.8	Missing-Data Mechanism and Ignorability, 227	
6.9	Competing Methods and Some Comparisons with the EM Algorithm, 228	
6.9.1	Introduction, 228	
6.9.2	Simulated Annealing, 229	
6.9.3	Comparison of SA and EM Algorithm for Normal Mixtures, 230	
6.10	The Delta Algorithm, 230	
6.11	Image Space Reconstruction Algorithm, 232	
6.12	Further Applications of the EM Algorithm, 232	
6.12.1	Introduction, 232	
6.12.2	Hidden Markov Models, 233	
6.12.3	AIDS Epidemiology, 237	
6.12.4	Neural Networks, 239	
References		243
Author Index		261
Subject Index		265