

CONTENTS

PREFACE TO THE SECOND EDITION	xix
PREFACE TO THE FIRST EDITION	xxi
LIST OF EXAMPLES	xxv
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	6
1.4 Introductory Examples	8
1.4.1 Introduction	8
1.4.2 Example 1.1: A Multinomial Example	8
	vii

1.4.3	Example 1.2: Estimation of Mixing Proportions	13	2.3.2	Example 2.2: Linear Regression with Missing Dependent Values	47
1.5	Formulation of the EM Algorithm	18	2.3.3	Example 2.3: Missing Values in a Latin Square Design	49
1.5.1	EM Algorithm	18	2.3.4	Healy–Westmacott Procedure as an EM Algorithm	49
1.5.2	Example 1.3: Censored Exponentially Distributed Survival Times	20	2.4	Example 2.4: Multinomial with Complex Cell Structure	51
1.5.3	E- and M-Steps for the Regular Exponential Family	22	2.5	Example 2.5: Analysis of PET and SPECT Data	54
1.5.4	Example 1.4: Censored Exponentially Distributed Survival Times (<i>Example 1.3 Continued</i>)	23	2.6	Example 2.6: Multivariate t -Distribution (Known D.F.)	58
1.5.5	Generalized EM Algorithm	24	2.6.1	ML Estimation of Multivariate t -Distribution	58
1.5.6	GEM Algorithm Based on One Newton-Raphson Step	24	2.6.2	Numerical Example: Stack Loss Data	61
1.5.7	EM Gradient Algorithm	25	2.7	Finite Normal Mixtures	61
1.5.8	EM Mapping	26	2.7.1	Example 2.7: Univariate Component Densities	61
1.6	EM Algorithm for MAP and MPL Estimation	26	2.7.2	Example 2.8: Multivariate Component Densities	64
1.6.1	Maximum <i>a Posteriori</i> Estimation	26	2.7.3	Numerical Example: Red Blood Cell Volume Data	65
1.6.2	Example 1.5: A Multinomial Example (<i>Example 1.1 Continued</i>)	27	2.8	Example 2.9: Grouped and Truncated Data	66
1.6.3	Maximum Penalized Estimation	27	2.8.1	Introduction	66
1.7	Brief Summary of the Properties of the EM Algorithm	28	2.8.2	Specification of Complete Data	66
1.8	History of the EM Algorithm	29	2.8.3	E-Step	69
1.8.1	Early EM History	29	2.8.4	M-Step	70
1.8.2	Work Before Dempster, Laird, and Rubin (1977)	29	2.8.5	Confirmation of Incomplete-Data Score Statistic	70
1.8.3	EM Examples and Applications Since Dempster, Laird, and Rubin (1977)	31	2.8.6	M-Step for Grouped Normal Data	71
1.8.4	Two Interpretations of EM	32	2.8.7	Numerical Example: Grouped Log Normal Data	72
1.8.5	Developments in EM Theory, Methodology, and Applications	33	2.9	Example 2.10: A Hidden Markov AR(1) model	73
1.9	Overview of the Book	36	3 BASIC THEORY OF THE EM ALGORITHM		77
1.10	Notations	37	3.1	Introduction	77
2 EXAMPLES OF THE EM ALGORITHM		41	3.2	Monotonicity of the EM Algorithm	78
2.1	Introduction	41	3.3	Monotonicity of a Generalized EM Algorithm	79
2.2	Multivariate Data with Missing Values	42	3.4	Convergence of an EM Sequence to a Stationary Value	79
2.2.1	Example 2.1: Bivariate Normal Data with Missing Values	42	3.4.1	Introduction	79
2.2.2	Numerical Illustration	45	3.4.2	Regularity Conditions of Wu (1983)	80
2.2.3	Multivariate Data: Buck's Method	45	3.4.3	Main Convergence Theorem for a Generalized EM Sequence	81
2.3	Least Squares with Missing Data	47	3.4.4	A Convergence Theorem for an EM Sequence	82
2.3.1	Healy–Westmacott Procedure	47	3.5	Convergence of an EM Sequence of Iterates	83
			3.5.1	Introduction	83
			3.5.2	Two Convergence Theorems of Wu (1983)	83
			3.5.3	Convergence of an EM Sequence to a Unique Maximum Likelihood Estimate	84

3.5.4	Constrained Parameter Spaces	84	4.5.1	Definition	120
3.6	Examples of Nontypical Behavior of an EM (GEM) Sequence	85	4.5.2	Calculation of $J(\hat{\Psi})$ via Numerical Differentiation	122
3.6.1	Example 3.1: Convergence to a Saddle Point	85	4.5.3	Stability	123
3.6.2	Example 3.2: Convergence to a Local Minimum	88	4.5.4	Monitoring Convergence	124
3.6.3	Example 3.3: Nonconvergence of a Generalized EM Sequence	90	4.5.5	Difficulties of the SEM Algorithm	124
3.6.4	Example 3.4: Some E-Step Pathologies	93	4.5.6	Example 4.4: Univariate Contaminated Normal Data	125
3.7	Score Statistic	95	4.5.7	Example 4.5: Bivariate Normal Data with Missing Values	128
3.8	Missing Information	95	4.6	Bootstrap Approach to Standard Error Approximation	130
3.8.1	Missing Information Principle	95	4.7	Baker's, Louis', and Oakes' Methods for Standard Error Computation	131
3.8.2	Example 3.5: Censored Exponentially Distributed Survival Times (<i>Example 1.3 Continued</i>)	96	4.7.1	Baker's Method for Standard Error Computation	131
3.9	Rate of Convergence of the EM Algorithm	99	4.7.2	Louis' Method of Standard Error Computation	132
3.9.1	Rate Matrix for Linear Convergence	99	4.7.3	Oakes' Formula for Standard Error Computation	133
3.9.2	Measuring the Linear Rate of Convergence	100	4.7.4	Example 4.6: Oakes' Standard Error for Example 1.1	134
3.9.3	Rate Matrix in Terms of Information Matrices	101	4.7.5	Example 4.7: Louis' Method for Example 2.4	134
3.9.4	Rate Matrix for Maximum <i>a Posteriori</i> Estimation	102	4.7.6	Baker's Method for Standard Error for Categorical Data	135
3.9.5	Derivation of Rate Matrix in Terms of Information Matrices	102	4.7.7	Example 4.8: Baker's Method for Example 2.4	136
3.9.6	Example 3.6: Censored Exponentially Distributed Survival Times (<i>Example 1.3 Continued</i>)	103	4.8	Acceleration of the EM Algorithm via Aitken's Method	137
4	STANDARD ERRORS AND SPEEDING UP CONVERGENCE	105	4.8.1	Aitken's Acceleration Method	137
4.1	Introduction	105	4.8.2	Louis' Method	137
4.2	Observed Information Matrix	106	4.8.3	Example 4.9: Multinomial Data	138
4.2.1	Direct Evaluation	106	4.8.4	Example 4.10: Geometric Mixture	139
4.2.2	Extraction of Observed Information Matrix in Terms of the Complete-Data Log Likelihood	106	4.8.5	Example 4.11: Grouped and Truncated Data. (<i>Example 2.8 Continued</i>)	142
4.2.3	Regular Case	108	4.9	An Aitken Acceleration-Based Stopping Criterion	142
4.2.4	Evaluation of the Conditional Expected Complete-Data Information Matrix	108	4.10	Conjugate Gradient Acceleration of EM Algorithm	144
4.2.5	Examples	109	4.10.1	Conjugate Gradient Method	144
4.3	Approximations to Observed Information Matrix: i.i.d. Case	114	4.10.2	A Generalized Conjugate Gradient Algorithm	144
4.4	Observed Information Matrix for Grouped Data	116	4.10.3	Accelerating the EM Algorithm	145
4.4.1	Approximation Based on Empirical Information	116	4.11	Hybrid Methods for Finding the MLE	146
4.4.2	Example 4.3: Grouped Data from an Exponential Distribution	117	4.11.1	Introduction	146
4.5	Supplemented EM Algorithm	120	4.11.2	Combined EM and Modified Newton-Raphson Algorithm	146
			4.12	A GEM Algorithm Based on One Newton-Raphson Step	148
			4.12.1	Derivation of a Condition to be a Generalized EM Sequence	148
			4.12.2	Simulation Experiment	149
			4.13	EM Gradient Algorithm	149
			4.14	A Quasi-Newton Acceleration of the EM Algorithm	151
			4.14.1	The Method	151

4.14.2	Example 4.12: Dirichlet Distribution	153	5.8.8	Theoretical Results on the Rate of Convergence	181
4.15	Ikeda Acceleration	157	5.9	Example 5.6: Variance Components	182
5	EXTENSIONS OF THE EM ALGORITHM	159	5.9.1	A Variance Components Model	182
5.1	Introduction	159	5.9.2	E-Step	183
5.2	ECM Algorithm	160	5.9.3	M-Step	184
5.2.1	Motivation	160	5.9.4	Application of Two Versions of ECME Algorithm	185
5.2.2	Formal Definition	160	5.9.5	Numerical Example	185
5.2.3	Convergence Properties	162	5.10	Linear Mixed Models	186
5.2.4	Speed of Convergence	162	5.10.1	Introduction	186
5.2.5	Convergence Rates of EM and ECM	163	5.10.2	General Form of Linear Mixed Model	187
5.2.6	Example 5.1: ECM Algorithm for Hidden Markov AR(1) Model	164	5.10.3	REML Estimation	188
5.2.7	Discussion	164	5.10.4	Example 5.7: REML Estimation in a Hierarchical Random Effects Model	188
5.3	Multicycle ECM Algorithm	165	5.10.5	Some Other EM-Related Approaches to Mixed Model Estimation	191
5.4	Example 5.2: Normal Mixtures with Equal Correlations	166	5.10.6	Generalized Linear Mixed Models	191
5.4.1	Normal Components with Equal Correlations	166	5.11	Example 5.8: Factor Analysis	193
5.4.2	Application of ECM Algorithm	166	5.11.1	EM Algorithm for Factor Analysis	193
5.4.3	Fisher's <i>Iris</i> Data	168	5.11.2	ECME Algorithm for Factor Analysis	196
5.5	Example 5.3: Mixture Models for Survival Data	168	5.11.3	Numerical Example	196
5.5.1	Competing Risks in Survival Analysis	168	5.11.4	EM Algorithm in Principal Component Analysis	196
5.5.2	A Two-Component Mixture Regression Model	169	5.12	Efficient Data Augmentation	198
5.5.3	Observed Data	169	5.12.1	Motivation	198
5.5.4	Application of EM Algorithm	170	5.12.2	Maximum Likelihood Estimation of t -Distribution	198
5.5.5	M-Step for Gompertz Components	171	5.12.3	Variance Components Model	202
5.5.6	Application of a Multicycle ECM Algorithm	172	5.13	Alternating ECM Algorithm	202
5.5.7	Other Examples of EM Algorithm in Survival Analysis	173	5.14	Example 5.9: Mixtures of Factor Analyzers	204
5.6	Example 5.4: Contingency Tables with Incomplete Data	174	5.14.1	Normal Component Factor Analyzers	205
5.7	ECME Algorithm	175	5.14.2	E-step	205
5.8	Example 5.5: MLE of t -Distribution with Unknown D.F.	176	5.14.3	CM-steps	206
5.8.1	Application of the EM Algorithm	176	5.14.4	t -Component Factor Analyzers	207
5.8.2	M-Step	177	5.14.5	E-step	210
5.8.3	Application of ECM Algorithm	177	5.14.6	CM-steps	211
5.8.4	Application of ECME Algorithm	178	5.15	Parameter-Expanded EM (PX-EM) Algorithm	212
5.8.5	Some Standard Results	178	5.16	EMS Algorithm	213
5.8.6	Missing Data	179	5.17	One-Step-Late Algorithm	213
5.8.7	Numerical Examples	181	5.18	Variance Estimation for Penalized EM and OSL Algorithms	214

5.18.1	Penalized EM Algorithm	214
5.18.2	OSL Algorithm	215
5.18.3	Example 5.9: Variance of MPLE for the Multinomial (<i>Examples 1.1 and 4.1 Continued</i>)	215
5.19	Incremental EM	216
5.20	Linear Inverse Problems	217

6 MONTE CARLO VERSIONS OF THE EM ALGORITHM 219

6.1	Introduction	219
6.2	Monte Carlo Techniques	220
6.2.1	Integration and Optimization	220
6.2.2	Example 6.1: Monte Carlo Integration	221
6.3	Monte Carlo EM	221
6.3.1	Introduction	221
6.3.2	Example 6.2: Monte Carlo EM for Censored Data from Normal	223
6.3.3	Example 6.3: MCEM for a Two-Parameter Multinomial (<i>Example 2.4 Continued</i>)	224
6.3.4	MCEM in Generalized Linear Mixed Models	224
6.3.5	Estimation of Standard Error with MCEM	225
6.3.6	Example 6.4: MCEM Estimate of Standard Error for One-Parameter Multinomial (<i>Example 1.1 Continued</i>)	226
6.3.7	Stochastic EM Algorithm	227
6.4	Data Augmentation	228
6.4.1	The Algorithm	228
6.4.2	Example 6.5: Data Augmentation in the Multinomial (<i>Examples 1.1, 1.5 Continued</i>)	229
6.5	Bayesian EM	230
6.5.1	Posterior Mode by EM	230
6.5.2	Example 6.6: Bayesian EM for Normal with Semi-Conjugate Prior	231
6.6	I.I.D. Monte Carlo Algorithms	232
6.6.1	Introduction	232
6.6.2	Rejection Sampling Methods	233
6.6.3	Importance Sampling	234
6.7	Markov Chain Monte Carlo Algorithms	236
6.7.1	Introduction	236

6.7.2	Essence of MCMC	238
6.7.3	Metropolis–Hastings Algorithms	239
6.8	Gibbs Sampling	241
6.8.1	Introduction	241
6.8.2	Rao–Blackwellized Estimates with Gibbs Samples	242
6.8.3	Example 6.7: Why Does Gibbs Sampling Work?	243
6.9	Examples of MCMC Algorithms	245
6.9.1	Example 6.8: M-H Algorithm for Bayesian Probit Regression	245
6.9.2	Monte Carlo EM with MCMC	246
6.9.3	Example 6.9: Gibbs Sampling for the Mixture Problem	249
6.9.4	Example 6.10: Bayesian Probit Analysis with Data Augmentation	250
6.9.5	Example 6.11: Gibbs Sampling for Censored Normal	251
6.10	Relationship of EM to Gibbs Sampling	254
6.10.1	EM–Gibbs Sampling Connection	254
6.10.2	Example 6.12: EM–Gibbs Connection for Censored Data from Normal (<i>Example 6.11 Continued</i>)	256
6.10.3	Example 6.13: EM–Gibbs Connection for Normal Mixtures	257
6.10.4	Rate of Convergence of Gibbs Sampling and EM	257
6.11	Data Augmentation and Gibbs Sampling	258
6.11.1	Introduction	258
6.11.2	Example 6.14: Data Augmentation and Gibbs Sampling for Censored Normal (<i>Example 6.12 Continued</i>)	259
6.11.3	Example 6.15: Gibbs Sampling for a Complex Multinomial (<i>Example 2.4 Continued</i>)	260
6.11.4	Gibbs Sampling Analogs of ECM and ECME Algorithms	261
6.12	Empirical Bayes and EM	263
6.13	Multiple Imputation	264
6.14	Missing-Data Mechanism, Ignorability, and EM Algorithm	265

7 SOME GENERALIZATIONS OF THE EM ALGORITHM 269

7.1	Introduction	269
7.2	Estimating Equations and Estimating Functions	270
7.3	Quasi-Score and the Projection-Solution Algorithm	270
7.4	Expectation-Solution (ES) Algorithm	273
7.4.1	Introduction	273

7.4.2	Computational and Asymptotic Properties of the ES Algorithm	274	8.4.9	Hierarchical Mixture of Experts	307
7.4.3	Example 7.1: Multinomial Example by ES Algorithm (<i>Example 1.1 Continued</i>)	274	8.4.10	Boltzmann Machine	308
7.5	Other Generalizations	275	8.5	Data Mining	309
7.6	Variational Bayesian EM Algorithm	276	8.6	Bioinformatics	310
7.7	MM Algorithm	278	REFERENCES		311
7.7.1	Introduction	278	AUTHOR INDEX		339
7.7.2	Methods for Constructing Majorizing/Minorizing Functions	279	SUBJECT INDEX		347
7.7.3	Example 7.2: MM Algorithm for the Complex Multinomial (<i>Example 1.1 Continued</i>)	280			
7.8	Lower Bound Maximization	281			
7.9	Interval EM Algorithm	283			
7.9.1	The Algorithm	283			
7.9.2	Example 7.3: Interval-EM Algorithm for the Complex Multinomial (<i>Example 2.4 Continued</i>)	283			
7.10	Competing Methods and Some Comparisons with EM	284			
7.10.1	Introduction	284			
7.10.2	Simulated Annealing	284			
7.10.3	Comparison of SA and EM Algorithm for Normal Mixtures	285			
7.11	The Delta Algorithm	286			
7.12	Image Space Reconstruction Algorithm	287			
8	FURTHER APPLICATIONS OF THE EM ALGORITHM	289			
8.1	Introduction	289			
8.2	Hidden Markov Models	290			
8.3	AIDS Epidemiology	293			
8.4	Neural Networks	295			
8.4.1	Introduction	295			
8.4.2	EM Framework for NNs	296			
8.4.3	Training Multi-Layer Perceptron Networks	297			
8.4.4	Intractibility of the Exact E-Step for MLPs	300			
8.4.5	An Integration of the Methodology Related to EM Training of RBF Networks	300			
8.4.6	Mixture of Experts	301			
8.4.7	Simulation Experiment	305			
8.4.8	Normalized Mixtures of Experts	306			