
Contents

Preface	xi
Yannis DIMOTIKALIS, Alex KARAGRIGORIOU, Christina PARPOULA and Christos H. SKIADAS	
Part 1. Computational Data Analysis	1
Chapter 1. A Variant of Updating PageRank in Evolving Tree Graphs	3
Benard ABOLA, Pitos Seleka BIGANDA, Christopher ENGSTRÖM, John Magero MANGO, Godwin KAKUBA and Sergei SILVESTROV	
1.1. Introduction	3
1.2. Notations and definitions	5
1.3. Updating the transition matrix	5
1.4. Updating the PageRank of a tree graph	10
1.4.1. Updating the PageRank of tree graph when a batch of edges changes	12
1.4.2. An example of updating the PageRank of a tree	15
1.5. Maintaining the levels of vertices in a changing tree graph	17
1.6. Conclusion	21
1.7. Acknowledgments	21
1.8. References	21
Chapter 2. Nonlinearly Perturbed Markov Chains and Information Networks	23
Benard ABOLA, Pitos Seleka BIGANDA, Sergei SILVESTROV, Dmitrii SILVESTROV, Christopher ENGSTRÖM, John Magero MANGO and Godwin KAKUBA	
2.1. Introduction	23
2.2. Stationary distributions for Markov chains with damping component	26

2.2.1. Stationary distributions for Markov chains with damping component	26
2.2.2. The stationary distribution of the Markov chain $X_{0,n}$	28
2.3. A perturbation analysis for stationary distributions of Markov chains with damping component	29
2.3.1. Continuity property for stationary probabilities	29
2.3.2. Rate of convergence for stationary distributions	29
2.3.3. Asymptotic expansions for stationary distributions	30
2.3.4. Results of numerical experiments	32
2.4. Coupling and ergodic theorems for perturbed Markov chains with damping component	39
2.4.1. Coupling for regularly perturbed Markov chains with damping component	39
2.4.2. Coupling for singularly perturbed Markov chains with damping component	41
2.4.3. Ergodic theorems for perturbed Markov chains with damping component in the triangular array mode	42
2.4.4. Numerical examples	43
2.5. Acknowledgments	51
2.6. References	51
Chapter 3. PageRank and Perturbed Markov Chains	57
Pitos Seleka BIGANDA, Benard ABOLA, Christopher ENGSTRÖM, Sergei SILVESTROV, Godwin KAKUBA and John Magero MANGO	
3.1. Introduction	57
3.2. PageRank of the first-order perturbed Markov chain	59
3.3. PageRank of the second-order perturbed Markov chain	60
3.4. Rates of convergence of PageRanks of first- and second-order perturbed Markov chains	70
3.5. Conclusion	72
3.6. Acknowledgments	72
3.7. References	72
Chapter 4. Doubly Robust Data-driven Distributionally Robust Optimization	75
Jose BLANCHET, Yang KANG, Fan ZHANG, Fei HE and Zhangyi HU	
4.1. Introduction	75
4.2. DD-DRO, optimal transport and supervised machine learning	79
4.2.1. Optimal transport distances and discrepancies	80
4.3. Data-driven selection of optimal transport cost function	81
4.3.1. Data-driven cost functions via metric learning procedures	81
4.4. Robust optimization for metric learning	83

4.4.1. Robust optimization for relative metric learning	83
4.4.2. Robust optimization for absolute metric learning	86
4.5. Numerical experiments	88
4.6. Discussion and conclusion	89
4.7. References	89

**Chapter 5. A Comparison of Graph Centrality Measures
Based on Lazy Random Walks 91**

Collins ANGUZU, Christopher ENGSTRÖM and Sergei SILVESTROV

5.1. Introduction	91
5.1.1. Notations and abbreviations	93
5.1.2. Linear systems and the Neumann series	94
5.2. Review on some centrality measures	95
5.2.1. Degree centrality	95
5.2.2. Katz status and β -centralities	95
5.2.3. Eigenvector and cumulative nomination centralities	96
5.2.4. Alpha centrality	97
5.2.5. PageRank centrality	98
5.2.6. Summary of the centrality measures as steady state, shifted and power series	99
5.3. Generalizations of centrality measures	99
5.3.1. Priors to centrality measures	99
5.3.2. Lazy variants of centrality measures	100
5.3.3. Lazy α -centrality	100
5.3.4. Lazy Katz centrality	102
5.3.5. Lazy cumulative nomination centrality	103
5.4. Experimental results	104
5.5. Discussion	106
5.6. Conclusion	109
5.7. Acknowledgments	109
5.8. References	110

Chapter 6. Error Detection in Sequential Laser Sensor Input 113

Gwenael GATTO and Olympia HADJILIADIS

6.1. Introduction	113
6.2. Data description	114
6.3. Algorithms	116
6.3.1. Algorithm for consecutive changes in mean	118
6.3.2. Algorithm for burst detection	120
6.4. Results	125
6.5. Acknowledgments	127
6.6. References	127

Chapter 7. Diagnostics and Visualization of Point Process Models for Event Times on a Social Network	129
Jing WU, Anna L. SMITH and Tian ZHENG	
7.1. Introduction	129
7.2. Background	131
7.2.1. Univariate point processes	131
7.2.2. Network point processes	132
7.3. Model checking for time heterogeneity	134
7.3.1. Time rescaling theorem	134
7.3.2. Residual process	136
7.4. Model checking for network heterogeneity and structure	138
7.4.1. Kolmogorov–Smirnov test	138
7.4.2. Structure score based on the Pearson residual matrix	141
7.5. Summary	143
7.6. Acknowledgments	144
7.7. References	144
Part 2. Data Analysis Methods and Tools	147
Chapter 8. Exploring the Distribution of Conditional Quantile Estimates: An Application to Specific Costs of Pig Production in the European Union	149
Dominique DESBOIS	
8.1. Introduction	150
8.2. Conceptual framework and methodological aspects	150
8.2.1. The empirical model for estimating the specific production costs	151
8.2.2. The procedures for estimating and testing conditional quantiles	152
8.2.3. Symbolic PCA of the specific cost distributions	154
8.2.4. Symbolic clustering analysis of the specific cost distributions	162
8.3. Results	165
8.3.1. The SO-PCA of specific cost estimates	167
8.3.2. The divisive hierarchy of specific cost estimates	170
8.4. Conclusion	171
8.5. References	172
Chapter 9. Maximization Problem Subject to Constraint of Availability in Semi-Markov Model of Operation	175
Franciszek GRABSKI	
9.1. Introduction	175
9.2. Semi-Markov decision process	176
9.3. Semi-Markov decision model of operation	177

9.3.1. Description and assumptions	177
9.3.2. Model construction	177
9.4. Optimization problem	178
9.4.1. Linear programming method	179
9.5. Numerical example	182
9.6. Conclusion	184
9.7. References	185

**Chapter 10. The Impact of Multicollinearity on Big Data
Multivariate Analysis Modeling** 187

Kimon NTOTSIS and Alex KARAGRIGORIOU

10.1. Introduction	187
10.2. Multicollinearity	188
10.3. Dimension reduction techniques	191
10.3.1. Beale <i>et al.</i>	192
10.3.2. Principal component analysis	192
10.4. Application	194
10.4.1. The modeling of PPE	194
10.4.2. Concluding remarks	200
10.5. Acknowledgments	200
10.6. References	200

**Chapter 11. Weak Signals in High-Dimensional Poisson
Regression Models** 203

Orawan REANGSEPHET, Supanee LISAWADI and Syed Ejaz AHMED

11.1. Introduction	203
11.2. Statistical background	204
11.3. Methodologies	205
11.3.1. Predictor screening methods	205
11.3.2. Post-screening parameter estimation methods	206
11.4. Numerical studies	208
11.4.1. Simulation settings and performance criteria	208
11.4.2. Results	209
11.5. Conclusion	217
11.6. Acknowledgments	218
11.7. References	218

**Chapter 12. Groundwater Level Forecasting for Water Resource
Management** 221

Andrea ZIRULIA, Alessio BARBAGLI and Enrico GUASTALDI

12.1. Introduction	221
12.2. Materials and methods	222

12.2.1. Study area	222
12.2.2. Forecast method	222
12.3. Results	224
12.4. Conclusion	230
12.5. References	230
Chapter 13. Phase I Non-parametric Control Charts for Individual Observations: A Selective Review and Some Results	233
Christina PARPOULA	
13.1. Introduction	234
13.1.1. Background	234
13.1.2. Univariate non-parametric process monitoring	235
13.2. Problem formulation	237
13.3. A comparative study	239
13.3.1. The existing methodologies	239
13.3.2. Simulation settings	240
13.3.3. Simulation-study results	242
13.4. Concluding remarks	247
13.5. References	247
Chapter 14. On Divergence and Dissimilarity Measures for Multiple Time Series	249
Konstantinos MAKRIS, Alex KARAGRIGORIOU and Ilia VONTA	
14.1. Introduction	249
14.2. Classical measures	250
14.3. Divergence measures	252
14.4. Dissimilarity measures for ordered data	254
14.4.1. Standard dissimilarity measures	254
14.4.2. Advanced dissimilarity measures	256
14.5. Conclusion	259
14.6. References	259
List of Authors	261
Index	265