

Contents

<i>List of contributors</i>	page xiii
<i>Preface</i>	xvii

Part I	Mathematical foundations	1
1	Tensor models: solution methods and applications	3
	Shiqian Ma, Bo Jiang, Xiuzhen Huang, and Shuzhong Zhang	
1.1	Introduction	3
1.2	Tensor models	5
1.2.1	Sparse and low-rank tensor optimization models	5
1.2.2	Tensor principal component analysis	6
1.2.3	The tensor co-clustering problem	8
1.3	Reformulation of tensor models	11
1.3.1	Low- n -rank tensor optimization	11
1.3.2	Equivalent formulation of tensor PCA	13
1.4	Solution methods	16
1.4.1	Directly resorting to some existing solver	16
1.4.2	First-order methods	18
1.4.3	The block optimization technique	22
1.5	Applications	24
1.5.1	Computational results on gene expression data	25
1.6	Conclusions	30
	References	31
2	Sparsity-aware distributed learning	37
	Symeon Chouvardas, Yannis Kopsinis, and Sergios Theodoridis	
2.1	Introduction	37
2.2	Batch distributed sparsity promoting algorithms	39
2.2.1	Problem formulation	39
2.2.2	LASSO and its distributed learning formulation	40
2.2.3	Sparsity-aware learning: the greedy point of view	42
2.2.4	Other distributed sparse recovery algorithms	45

2.3	Online sparsity-aware distributed learning	46
2.3.1	Problem description	46
2.3.2	LMS based sparsity-promoting algorithm	47
2.3.3	The GreeDi LMS algorithm	49
2.3.4	Set-theoretic sparsity-aware distributed learning	51
2.4	Simulation examples	56
2.4.1	Performance evaluation of batch methods	57
2.4.2	Performance evaluation of online methods	58
	References	61
3	Optimization algorithms for big data with application in wireless networks	66
	Mingyi Hong, Wei-Cheng Liao, Ruoyu Sun, and Zhi-Quan Luo	
3.1	Introduction	66
3.1.1	Motivation	66
3.1.2	The organization of the chapter	67
3.2	First-order algorithms for big data	67
3.2.1	The block coordinate descent algorithm	67
3.2.2	The ADMM algorithm	69
3.2.3	The BSUM method	70
3.3	Application to network provisioning problem	72
3.3.1	The setting	72
3.3.2	Network with an uncapacitated backhaul	75
3.3.3	Network with a capacitated backhaul	82
3.4	Numerical results	88
3.4.1	Scenario 1: Performance comparison with heuristic algorithms	89
3.4.2	Scenario 2: The efficiency of N-MaxMin WMMSE algorithm	91
3.4.3	Scenario 3: Multi-commodity routing problem with parallel implementation	92
3.4.4	Scenario 4: Performance evaluation for Algorithm 1 with zones of nodes	94
3.5	Appendix	94
	References	97
4	A unified distributed algorithm for non-cooperative games	101
	Jong-Shi Pang and Meisam Razaviyayn	
4.1	Introduction	101
4.2	The nonsmooth, nonconvex game	104
4.3	The unified algorithm	106
4.3.1	Special cases	108
4.4	Convergence analysis: contraction approach	110
4.4.1	Probabilistic player choices	115

4.5	Convergence analysis: potential approach	116
4.5.1	Generalized potential games	121
	References	122
	Appendix	125
Part II	Big data over cyber networks	135
5	Big data analytics systems	137
	Ganesh Ananthanarayanan and Ishai Menache	
5.1	Introduction	137
5.2	Scheduling	139
5.2.1	Fairness	139
5.2.2	Placement constraints	142
5.2.3	Additional system-wide objectives	145
5.2.4	Stragglers	146
5.3	Storage	148
5.3.1	Distributed file system	148
5.3.2	In-memory storage	151
5.4	Concluding remarks	156
	References	158
6	Distributed big data storage in optical wireless networks	161
	Chen Gong, Zhengyuan Xu, and Xiaodong Wang	
6.1	Introduction	161
6.2	Big data distributed storage in a wireless network	163
6.2.1	Wireless distributed storage network framework	163
6.2.2	Optical wireless framework	165
6.2.3	Rateless coded distributed data storage	167
6.2.4	Network coded system with full downloading	167
6.2.5	Network coded system with partial downloading	168
6.3	Reconstructability condition for partial downloading	169
6.3.1	μ -Reconstructability for MSR point	169
6.3.2	μ -Reconstructability for a practical MBR point coding scheme	170
6.4	Channel and power allocation for partial downloading	173
6.4.1	Wireless resource allocation framework	174
6.4.2	Optimal channel and power allocation for the relaxed problem	174
6.5	Open research topics	176
6.5.1	General research topics for wireless distributed storage networks	176
6.5.2	Research topics for data storage in optical wireless networks	177
6.5.3	Research topics for data storage in named data networks	177
	References	178

7	Big data aware wireless communication: challenges and opportunities	180
	Suzhi Bi, Rui Zhang, Zhi Ding, and Shuguang Cui	
7.1	Introduction	180
7.2	Scalable wireless network architecture for big data	182
7.2.1	Hybrid processing structure	182
7.2.2	Web caching in wireless infrastructure	185
7.2.3	Data aware processing units	187
7.3	Wireless system design in big data era	188
7.3.1	Analog vs. digital backhaul	188
7.3.2	Joint base station and cloud processing with digital backhaul	191
7.3.3	Section summary	198
7.4	Big data aware wireless networking	198
7.4.1	Wireless big data analytics	200
7.4.2	Data-driven mobile cloud computing	205
7.4.3	Software-defined networking design	207
7.4.4	Section summary	209
7.5	Conclusions	210
	Acknowledgement	211
	References	211
8	Big data processing for smart grid security	217
	Lanchao Liu, Zhu Han, H. Vincent Poor, and Shuguang Cui	
8.1	Preliminaries and motivations	217
8.2	Sparse optimization for false data injection detection	219
8.2.1	State estimation and false data injection attacks	219
8.2.2	Nuclear norm minimization	223
8.2.3	Low-rank matrix factorization	225
8.2.4	Numerical results	227
8.3	Distributed approach for security-constrained optimal power flow	232
8.3.1	Security-constrained optimal power flow	232
8.3.2	ADMM method	235
8.3.3	Distributed and parallel approach for SCOPF	236
8.3.4	Numerical results	238
8.4	Concluding remarks	241
	Acknowledgement	241
	References	242
Part III	Big data over social networks	245
9	Big data: a new perspective on cities	247
	Riccardo Gallotti, Thomas Louail, Rémi Louf, and Marc Barthélemy	
9.1	Big data and urban systems	247
9.2	Infrastructure networks	249

9.2.1	Road networks	249
9.2.2	Subway networks	255
9.3	Mobility networks	257
9.3.1	A renewed interest	257
9.3.2	Individual mobility networks	258
9.3.3	From big data to the spatial structure of cities	261
9.4	Scaling in cities	268
9.5	Discussion: towards a new science of cities	272
	Acknowledgments	273
	References	273
10	High-dimensional network analytics: mapping topic networks in Twitter data during the Arab Spring	278
	Kathleen M. Carley, Wei Wei, and Kenneth Joseph	
10.1	Introduction	278
10.2	Arab Spring	280
10.3	General background	280
10.4	Data	281
10.5	The social pulse: geo-temporal trends in Twitter topics and users	284
10.5.1	Methodology	284
10.5.2	Topic overview	285
10.5.3	Over time analysis	287
10.5.4	Characterization of user–topic similarity network	288
10.5.5	Social interaction overview: the reply network	290
10.5.6	Characterization of group structure	291
10.5.7	Key actors	294
10.6	Discussion	295
10.7	Conclusion	297
	Acknowledgements	298
	References	299
11	Social influence analysis in the big data era: a review	301
	Jianping Cao, Dongliang Duan, Liuqing Yang, Qingpeng Zhang, Senzhang Wang, and Feiyue Wang	
11.1	Introduction	301
11.2	Social influence measurement	304
11.2.1	Network-based measures	304
11.2.2	Behavior-based measures	309
11.2.3	Interaction-based measures	312
11.2.4	Topic-based measures	313
11.2.5	Other measures	316
11.3	Influence propagation and maximization	317
11.3.1	Opinion leader identification	317
11.3.2	Influence maximization	319

11.3.3	Diffusion network inference	324
11.3.4	Challenges of IP&M	327
11.4	Challenges in big data	327
11.5	Summary	328
	Acknowledgement	329
	References	329
Part IV	Big data over biological networks	335
12	Inference of gene regulatory networks: validation and uncertainty	337
	Xiaoning Qian, Byung-Jun Yoon, and Edward R. Dougherty	
12.1	Introduction	337
12.2	Background	338
12.2.1	Markov chains	339
12.2.2	Logical regulatory networks	340
12.2.3	Control policy for maximal steady-state alteration	341
12.2.4	Inference algorithms	341
12.3	Network distance functions	343
12.3.1	Semi-metrics	343
12.3.2	Rule-based distance	343
12.3.3	Topology-based distance	344
12.3.4	Transition-probability-based distance	344
12.3.5	Steady-state distance	345
12.3.6	Control-based distance	345
12.4	Inference performance	346
12.4.1	Measuring inference performance using distance functions	346
12.4.2	Analytic example	347
12.4.3	Synthetic examples	348
12.5	Consistency	352
12.6	Approximation	352
12.7	Validation from experimental data	353
12.7.1	Metastatic melanoma network inference	354
12.8	Uncertainty quantification	354
12.8.1	Mean objective cost of uncertainty	356
12.8.2	Intervention in yeast cell cycle network with uncertainty	358
	References	360
13	Inference of gene networks associated with the host response to infectious disease	365
	Zhe Gan, Xin Yuan, Ricardo Henao, Ephraim L. Tsalik, and Lawrence Carin	
13.1	Background	365
13.2	Factor models in gene expression analysis	366

13.3	Factor models	367
13.3.1	Shrinkage prior	368
13.3.2	Multiplicative gamma process	369
13.4	Discriminative models	370
13.4.1	Bayesian log-loss	370
13.4.2	Bayesian hinge-loss	372
13.5	Discriminative factor model	372
13.5.1	Multi-task learning	374
13.6	Inference	374
13.7	Experiments	376
13.7.1	Performance measures	377
13.7.2	Experimental setup	377
13.7.3	Classification results	378
13.7.4	Interpretation	382
13.8	Closing remarks	384
13.9	Inference details	385
	Acknowledgements	387
	References	388
14	Gene-set-based inference of biological network topologies from big molecular profiling data	391
	Lipi Acharya and Dongxiao Zhu	
14.1	Introduction	391
14.2	Big data to network components	393
14.3	Gene sets related to network components	394
14.4	Reconstructing biological network topologies using gene sets	395
14.4.1	A general setting	395
14.4.2	Gene set Gibbs sampling	397
14.4.3	Gene set simulated annealing	397
14.5	Discussion and future work	403
	References	406
15	Large-scale correlation mining for biomolecular network discovery	409
	Alfred Hero and Bala Rajaratnam	
15.1	Introduction	409
15.2	Illustrative example	414
15.2.1	Pairwise correlation	416
15.2.2	From pairwise correlation to networks of correlations	417
15.3	Principles of correlation mining for big data	419
15.3.1	Correlation mining for correlation flips between two populations	424
15.3.2	Large-scale implementation of correlation mining	426

15.4	Perspectives and future challenges	427
15.4.1	State-of-the-art in correlation mining	427
15.4.2	Future challenges in correlation mining biomolecular networks	429
15.5	Conclusion	431
	Acknowledgements	431
	References	432
	<i>Index</i>	437