

Eitan Altman
Bruno Gaujal
Arie Hordijk

Discrete-Event Control of Stochastic Networks: Multimodularity and Regularity



Springer

Table of Contents

Introduction	1
--------------------	---

Part I Theoretical foundations

1	Multimodularity, Convexity and Optimization	11
1.1	Introduction	11
1.1.1	Organization of the chapter	12
1.2	Properties of multimodular functions	12
1.2.1	General properties	13
1.2.2	Multimodularity and convexity	14
1.3	The optimality of bracket policies for a single criterion	21
1.3.1	Upper Bounds	22
1.3.2	Lower Bounds	24
1.3.3	Optimality of the Bracket Sequences	25
1.4	The optimality of bracket policies for multiple criteria	26
1.5	Application of the optimization theorems	29
1.5.1	Applications in high-speed telecommunication systems	29
1.6	Clustering versus Smoothing	34
1.7	Appendix A: proof of Lemma 2	34
1.8	Appendix B: Proof of Lemma 5	36
1.9	Appendix C: average and weighted costs	37
2	Balanced Sequences	39
2.1	Introduction	39
2.1.1	Organization of the chapter	39
2.2	Balanced sequences and bracket sequences	39
2.3	Constant gap sequences	42
2.4	Characterization of balanced sequences	45
2.5	Balanceable rates	46
2.5.1	The case $K = 2$	46
2.5.2	The case $K = 3$	47
2.5.3	The case $K = 4$	48
2.5.4	The general case	49
2.5.5	Extensions of the original problem	52
2.6	Appendix	53

3	Stochastic Event Graphs	55
3.1	Introduction	55
3.1.1	Organization of the chapter	55
3.2	Stochastic Petri Nets	55
3.2.1	Properties	59
3.2.2	Event graphs	59
3.3	Dynamics of Event Graphs	60
3.3.1	State variables: the firing epochs	60
3.3.2	The (max,plus) Semi-ring	62
3.3.3	Evolution Equation in the (max,+) semi-ring	63
3.4	Queuing networks	65
3.4.1	The G/G/1 queue	65
3.4.2	Queues in tandem	66
3.4.3	Kanban systems	68
3.4.4	Window flow control	70
3.4.5	Leaky buckets	72
3.5	Lindley's equation for (max,+) systems	73

Part II Admission and routing control

4	Admission control in stochastic event graphs	79
4.1	Introduction	79
4.1.1	Organization of the chapter	80
4.2	Multimodularity and admission control	80
4.2.1	Admission policy: the time slot approach	80
4.3	The FIFO queue	81
4.3.1	Coupling of the service times with the customers	82
4.3.2	Multimodularity	85
4.4	(max,+) systems with one input: multimodularity	88
4.5	A dual policy: counting variable and waiting time	92
4.5.1	Waiting time	92
4.5.2	Coupling	92
4.5.3	Multimodularity	93
4.6	Optimal admission sequence	94
4.6.1	Dual bracket sequences	95
4.6.2	The time slot approach	96
4.6.3	The counting approach, the bounded case	97
4.6.4	The unbounded case	99
5	Applications in queuing networks	105
5.1	Introduction	105
5.2	Topological assumptions	106
5.3	Stochastic assumptions	107

5.3.1	Independence between service times and inter-arrival times	108
5.3.2	Cross traffic	109
6	Optimal routing	111
6.1	Introduction	111
6.1.1	Organization of the chapter	112
6.2	Routing of customers in multiple queues	112
6.2.1	Presentation of the model	112
6.2.2	Optimal routing sequence	113
6.3	Study of some special cases	117
6.3.1	The case $K = 2$	117
6.3.2	The homogeneous case	117
6.3.3	Two sets of identical servers	118
7	Optimal routing in two deterministic queues	119
7.1	Introduction	119
7.2	Bracket words	120
7.3	Expansion in continued fractions	121
7.4	Average waiting time in a single queue	125
7.4.1	Jumps	125
7.4.2	Formula of the average waiting time	127
7.4.3	Properties	132
7.5	Average waiting time for two queues	134
7.5.1	Presentation of the Model	134
7.5.2	Optimal Policies	134
7.5.3	Optimal Ratio	135
7.5.4	Algorithm and computational issues	138
7.6	Numerical experiments	141
7.7	Appendix: proof of Theorem 29	142
7.8	Appendix: proof of Lemma 37	146
7.9	Appendix 7.9 : proof of Theorem 30	147

Part III Several extensions

8	Networks with no buffers	155
8.1	Introduction	155
8.1.1	Organization of the chapter	156
8.2	The admission into a single server	156
8.3	Properties of the cost and of policies	158
8.4	Time averages	161
8.5	Routing to several servers	162
8.6	The service assignment problem	163
8.7	The multimodularity	164

8.8	MDP formulation and new structural results	166
8.9	Multimodularity of the global cost: two servers	172
8.10	Examples of arrival processes	174
8.11	Robot scheduling for web search engines	179
9	Vacancies, service allocation and polling	183
9.1	Introduction	183
9.1.1	Organization of the chapter	183
9.2	The generic control models and main results	184
9.3	A single queue with service driven vacations	186
9.4	An arrival-driven vacation model	191
9.5	Arrival-driven polling model	195
9.6	The potential vacation times are a renewal process	196
9.6.1	A single queue	196
9.6.2	The polling control problem	199
9.7	1-gated service	199
9.7.1	A single queue	200
9.7.2	The case of several queues	201
9.7.3	Application to an ATM switch	202
10	Monotonicity of feedback control	205
10.1	Introduction	205
10.1.1	Organization of the chapter	206
10.2	Monotonicity in initial actions	206
10.3	Exogenous random variables and information patterns	208
10.4	Monotonicity: full information case	211
10.5	Monotonicity: general case and delayed information	212
10.6	State representation	213
10.6.1	General full information case	214
10.6.2	Monotonicity of the switching curves	216
10.6.3	State representation with delayed information structure	216
10.7	Relation between multimodularity, superconvexity and submodularity	218
10.8	Appendix: Key Lemma	220

Part IV Comparisons

11	Comparison of queues with discrete-time arrival processes	229
11.1	Introduction	229
11.2	On the Ross conjecture in discrete-time	231
11.3	A comparison lemma and its applications	232
11.3.1	Application 1: two i.i.d. MAP-sources perform better than 2 completely coupled Map-sources	235

11.3.2 Application 2: a fixed batch size is better than random batch sizes	235
11.3.3 Application 3: Fluid scaling improves the performance	236
11.4 A second comparison lemma	237
11.5 A stochastic lower bound on the traveling times	238
12 Simplex convexity	243
12.1 Introduction	243
12.1.1 Organization of the chapter	243
12.2 Multimodular Triangulations	244
12.3 Multimodular Functions	247
12.4 Sub-meshes	251
12.5 Cones	253
12.5.1 Minimization	254
12.5.2 Cone ordering and monotonicity	255
12.6 Application: periodic admission sequences in $G/G/1/\infty$ tandem queues	257
12.6.1 An example	259
13 Orders and bounds for multimodular functions	261
13.1 Introduction	261
13.2 The multimodular order and the cone order	262
13.2.1 Shift invariant counterparts	266
13.3 The graph order and the unbalance	268
13.4 Relations and counterexamples	272
13.4.1 The shift invariant cone order does not imply the graph order	272
13.4.2 The shift invariant multimodular order does not imply the shift invariant cone order	273
13.4.3 The graph order does not imply the shift invariant multimodular order	274
13.5 Bounding the Difference in Waiting Times for One Queue	275
13.6 Routing to Parallel Queues	278
14 Regular Ordering	283
14.1 Introduction	283
14.2 Preliminaries	285
14.2.1 Gap Sequences	285
14.2.2 Balanced Sequences	285
14.2.3 Regularity and Schur Convexity	287
14.3 Application 1: Maximal Waiting Times in Networks	289
14.3.1 The D/D/1 Model	289
14.3.2 Characterization using regular preserving functions	290
14.3.3 The average waiting time is not regular preserving	291
14.3.4 The G/G/1 Model	291

14.3.5 The Event Graph Model	294
14.3.6 Routing Problem	297
14.3.7 Computational Problems	298
14.3.8 A routing example	299
14.4 Application 2: Assignment to queues with no buffer with redundancy	299
14.5 Appendix: properties of the gap sequences	301
14.6 Appendix: relations between regularity and multimodularity ..	303
References	305
Index	311