

Statistical Inference via Convex Optimization

Anatoli Juditsky

Arkadi Nemirovski

Princeton University Press
Princeton and Oxford

Contents

List of Figures	xi
Preface	xiii
Acknowledgements	xvii
Notational Conventions	xix
About Proofs	xxi
On Computational Tractability	xxi
1 Sparse Recovery via ℓ_1 Minimization	1
1.1 Compressed Sensing: What is it about?	1
1.1.1 Signal Recovery Problem	1
1.1.2 Signal Recovery: Parametric and nonparametric cases	2
1.1.3 Compressed Sensing via ℓ_1 minimization: Motivation	6
1.2 Validity of sparse signal recovery via ℓ_1 minimization	8
1.2.1 Validity of ℓ_1 minimization in the noiseless case	8
1.2.2 Imperfect ℓ_1 minimization	11
1.2.3 Regular ℓ_1 recovery	13
1.2.4 Penalized ℓ_1 recovery	14
1.2.5 Discussion	14
1.3 Verifiability and tractability issues	19
1.3.1 Restricted Isometry Property and s -goodness of random matrices	20
1.3.2 Verifiable sufficient conditions for $\mathbf{Q}_q(s, \kappa)$	20
1.3.3 Tractability of $\mathbf{Q}_\infty(s, \kappa)$	22
1.4 Exercises for Chapter 1	26
1.5 Proofs	30
1.5.1 Proofs of Theorem 1.3, 1.4	30
1.5.2 Proof of Theorem 1.5	32
1.5.3 Proof of Proposition 1.7	33
1.5.4 Proof of Propositions 1.8 and 1.12	36
1.5.5 Proof of Proposition 1.10	37
1.5.6 Proof of Proposition 1.13	39
2 Hypothesis Testing	41
2.1 Preliminaries from Statistics: Hypotheses, Tests, Risks	41
2.1.1 Hypothesis Testing Problem	41
2.1.2 Tests	42
2.1.3 Testing from repeated observations	42
2.1.4 Risk of a simple test	45

2.1.5	Two-point lower risk bound	46
2.2	Hypothesis Testing via Euclidean Separation	49
2.2.1	Situation	49
2.2.2	Pairwise Hypothesis Testing via Euclidean Separation	50
2.2.3	Euclidean Separation, Repeated Observations, and Majority Tests	55
2.2.4	From Pairwise to Multiple Hypotheses Testing	58
2.3	Detectors and Detector-Based Tests	65
2.3.1	Detectors and their risks	65
2.3.2	Detector-based tests	65
2.4	Simple observation schemes	72
2.4.1	Simple observation schemes—Motivation	72
2.4.2	Simple observation schemes—The definition	73
2.4.3	Simple observation schemes—Examples	74
2.4.4	Simple observation schemes—Main result	79
2.4.5	Simple observation schemes—Examples of optimal detectors	83
2.5	Testing multiple hypotheses	87
2.5.1	Testing unions	88
2.5.2	Testing multiple hypotheses “up to closeness”	91
2.5.3	Illustration: Selecting the best among a family of estimates	97
2.6	Sequential Hypothesis Testing	105
2.6.1	Motivation: Election polls	105
2.6.2	Sequential hypothesis testing	108
2.6.3	Concluding remarks	113
2.7	Measurement Design in simple observation schemes	113
2.7.1	Motivation: Opinion polls revisited	113
2.7.2	Measurement Design: Setup	115
2.7.3	Formulating the MD problem	116
2.8	Affine detectors beyond simple observation schemes	123
2.8.1	Situation	124
2.8.2	Main result	132
2.9	Beyond the scope of affine detectors: lifting the observations	139
2.9.1	Motivation	139
2.9.2	Quadratic lifting: Gaussian case	140
2.9.3	Quadratic lifting—Does it help?	142
2.9.4	Quadratic lifting: Sub-Gaussian case	145
2.9.5	Generic application: Quadratically constrained hypotheses	147
2.10	Exercises for Chapter 2	157
2.10.1	Two-point lower risk bound	157
2.10.2	Around Euclidean Separation	157
2.10.3	Hypothesis testing via ℓ_1 -separation	157
2.10.4	Miscellaneous exercises	163
2.11	Proofs	168
2.11.1	Proof of the observation in Remark 2.8	168
2.11.2	Proof of Proposition 2.6 in the case of quasi-stationary K -repeated observations	168
2.11.3	Proof of Theorem 2.23	172
2.11.4	Proof of Proposition 2.37	175
2.11.5	Proof of Proposition 2.43	176
2.11.6	Proof of Proposition 2.46	180

3	From Hypothesis Testing to Estimating Functionals	185
3.1	Estimating linear forms on unions of convex sets	185
3.1.1	The problem	186
3.1.2	The estimate	187
3.1.3	Main result	189
3.1.4	Near-optimality	190
3.1.5	Illustration	191
3.2	Estimating N -convex functions on unions of convex sets	193
3.2.1	Outline	194
3.2.2	Estimating N -convex functions: Problem setting	197
3.2.3	Bisection estimate: Construction	199
3.2.4	Building Bisection estimate	201
3.2.5	Bisection estimate: Main result	202
3.2.6	Illustration	203
3.2.7	Estimating N -convex functions: An alternative	205
3.3	Estimating linear forms beyond simple observation schemes	211
3.3.1	Situation and goal	212
3.3.2	Construction and main results	213
3.3.3	Estimation from repeated observations	216
3.3.4	Application: Estimating linear forms of sub-Gaussianity parameters	218
3.4	Estimating quadratic forms via quadratic lifting	222
3.4.1	Estimating quadratic forms, Gaussian case	222
3.4.2	Estimating quadratic form, sub-Gaussian case	228
3.5	Exercises for Chapter 3	238
3.6	Proofs	250
3.6.1	Proof of Proposition 3.3	250
3.6.2	Verifying 1-convexity of the conditional quantile	253
3.6.3	Proof of Proposition 3.4	254
3.6.4	Proof of Proposition 3.14	258
4	Signal Recovery by Linear Estimation	260
	Overview	260
4.1	Preliminaries: Executive summary on Conic Programming	262
4.1.1	Cones	262
4.1.2	Conic problems and their duals	263
4.1.3	Schur Complement Lemma	265
4.2	Near-optimal linear estimation from Gaussian observations	265
4.2.1	Situation and goal	265
4.2.2	Building a linear estimate	267
4.2.3	Byproduct on semidefinite relaxation	274
4.3	From ellitopes to spectratopes	275
4.3.1	Spectratopes: Definition and examples	275
4.3.2	Semidefinite relaxation on spectratopes	277
4.3.3	Linear estimates beyond ellitopic signal sets and $\ \cdot\ _2$ -risk	278
4.4	Linear estimates of stochastic signals	291
4.4.1	Minimizing Euclidean risk	293
4.4.2	Minimizing $\ \cdot\ $ -risk	294
4.5	Linear estimation under uncertain-but-bounded noise	295
4.5.1	Uncertain-but-bounded noise	295

4.5.2	Mixed noise	299
4.6	Calculus of ellitopes/spectratopes	300
4.7	Exercises for Chapter 4	302
4.7.1	Linear estimates vs. Maximum Likelihood	302
4.7.2	Measurement Design in Signal Recovery	303
4.7.3	Around semidefinite relaxation	306
4.7.4	Around Propositions 4.4 and 4.14	317
4.7.5	Signal recovery in Discrete and Poisson observation schemes	335
4.7.6	Numerical lower-bounding minimax risk	347
4.7.7	Around \mathcal{S} -Lemma	359
4.7.8	Miscellaneous exercises	360
4.8	Proofs	361
4.8.1	Preliminaries	361
4.8.2	Proof of Proposition 4.6	364
4.8.3	Proof of Proposition 4.8	366
4.8.4	Proof of Lemma 4.17	368
4.8.5	Proofs of Propositions 4.5, 4.16 and 4.19	371
4.8.6	Proofs of Propositions 4.18 and 4.19, and justification of Remark 4.20	383
5	Signal Recovery Beyond Linear Estimates	386
	Overview	386
5.1	Polyhedral estimation	386
5.1.1	Motivation	386
5.1.2	Generic polyhedral estimate	388
5.1.3	Specifying sets \mathcal{H}_δ for basic observation schemes	390
5.1.4	Efficient upper-bounding of $\mathfrak{R}[H]$ and contrast design, I.	392
5.1.5	Efficient upper-bounding of $\mathfrak{R}[H]$ and contrast design, II.	399
5.1.6	Assembling estimates: Contrast aggregation	411
5.1.7	Numerical illustration	413
5.1.8	Calculus of compatibility	413
5.2	Recovering signals from nonlinear observations by Stochastic Optimization	415
5.2.1	Problem setting	415
5.2.2	Assumptions	417
5.2.3	Estimating via Sample Average Approximation	420
5.2.4	Stochastic Approximation estimate	423
5.2.5	Numerical illustration	425
5.2.6	“Single-observation” case	428
5.3	Exercises for Chapter 5	431
5.3.1	Estimation by Stochastic Optimization	431
5.4	Proofs	440
5.4.1	Proof of (5.8)	440
5.4.2	Proof of Lemma 5.6	441
5.4.3	Verification of (5.44)	442
5.4.4	Proof of Proposition 5.10	443
	Solutions to Selected Exercises	447
6.1	Solutions for Chapter 1	447
6.2	Solutions for Chapter 2	454

6.2.1	Two-point lower risk bound	454
6.2.2	Around Euclidean Separation	455
6.2.3	Hypothesis testing via ℓ_1 separation	457
6.2.4	Miscellaneous exercises	465
6.3	Solutions for Chapter 3	477
6.4	Solutions for Chapter 4	495
6.4.1	Linear Estimates vs. Maximum Likelihood	495
6.4.2	Measurement Design in Signal Recovery	497
6.4.3	Around semidefinite relaxation	502
6.4.4	Around Propositions 4.4 and 4.14	518
6.4.5	Numerical lower-bounding minimax risk	572
6.4.6	Around \mathcal{S} -Lemma	586
6.4.7	Miscellaneous exercises	589
6.5	Solutions for Chapter 5	592
6.5.1	Estimation by Stochastic Optimization	592
Appendix: Executive Summary on Efficient Solvability of Convex Optimization Problems		609
Bibliography		613
Index		629