

## Contents

<i>Preface</i>	<i>page</i> xiii
<i>Notation</i>	xvii
<i>List of Contributors</i>	xix
<b>1 Introduction to Information Theory and Data Science</b>	<b>1</b>
Miguel R. D. Rodrigues, Stark C. Draper, Waheed U. Bajwa, and Yonina C. Eldar	
1.1 Classical Information Theory: A Primer	2
1.2 Source Coding: Near-Lossless Compression of Binary Sources	4
1.3 Channel Coding: Transmission over the Binary Symmetric Channel	8
1.4 Linear Channel Coding	12
1.5 Connecting Information Theory to Data Science	15
1.6 Information Theory and Data Acquisition	18
1.7 Information Theory and Data Representation	20
1.8 Information Theory and Data Analysis and Processing	26
1.9 Discussion and Conclusion	35
References	36
<b>2 An Information-Theoretic Approach to Analog-to-Digital Compression</b>	<b>44</b>
Alon Kipnis, Yonina C. Eldar, and Andrea J. Goldsmith	
2.1 Introduction	44
2.2 Lossy Compression of Finite-Dimensional Signals	48
2.3 ADX for Continuous-Time Analog Signals	49
2.4 The Fundamental Distortion Limit	55
2.5 ADX under Uniform Sampling	63
2.6 Conclusion	69
References	70
<b>3 Compressed Sensing via Compression Codes</b>	<b>72</b>
Shirin Jalali and H. Vincent Poor	
3.1 Compressed Sensing	72
3.2 Compression-Based Compressed Sensing	74

3.3	Definitions	76
3.4	Compressible Signal Pursuit	77
3.5	Compression-Based Gradient Descent (C-GD)	81
3.6	Stylized Applications	85
3.7	Extensions	92
3.8	Conclusions and Discussion	100
	References	100
<b>4</b>	<b>Information-Theoretic Bounds on Sketching</b>	104
	Mert Pilancı	
4.1	Introduction	104
4.2	Types of Randomized Sketches	105
4.3	Background on Convex Analysis and Optimization	108
4.4	Sketching Upper Bounds for Regression Problems	110
4.5	Information-Theoretic Lower Bounds	114
4.6	Non-Parametric Problems	121
4.7	Extensions: Privacy and Communication Complexity	124
4.8	Numerical Experiments	126
4.9	Conclusion	127
A4.1	Proof of Theorem 4.3	128
A4.2	Proof of Lemma 4.3	129
	References	130
<b>5</b>	<b>Sample Complexity Bounds for Dictionary Learning from Vector- and Tensor-Valued Data</b>	134
	Zahra Shakeri, Anand D. Sarwate, and Waheed U. Bajwa	
5.1	Introduction	134
5.2	Dictionary Learning for Vector-Valued Data	137
5.3	Dictionary Learning for Tensors	149
5.4	Extensions and Open Problems	157
	References	160
<b>6</b>	<b>Uncertainty Relations and Sparse Signal Recovery</b>	163
	Erwin Riegler and Helmut Bölcskei	
6.1	Introduction	163
6.2	Uncertainty Relations in $(\mathbb{C}^m, \ \cdot\ _2)$	165
6.3	Uncertainty Relations in $(\mathbb{C}^m, \ \cdot\ _1)$	173
6.4	Sparse Signal Separation	181
6.5	The Set-Theoretic Null-Space Property	184
6.6	A Large Sieve Inequality in $(\mathbb{C}^m, \ \cdot\ _2)$	187
6.7	Uncertainty Relations in $L_1$ and $L_2$	189

6.8	Proof of Theorem 6.4	189
6.9	Results for $\ \cdot\ _1$ and $\ \cdot\ _2$	192
	References	193
<b>7</b>	<b>Understanding Phase Transitions via Mutual Information and MMSE</b>	197
	Galen Reeves and Henry D. Pfister	
7.1	Introduction	197
7.2	Problem Setup and Characterization	200
7.3	The Role of Mutual Information and MMSE	207
7.4	Proving the Replica-Symmetric Formula	215
7.5	Phase Transitions and Posterior Correlation	219
7.6	Conclusion	223
A7.1	Subset Response for the Standard Linear Model	223
	References	225
<b>8</b>	<b>Computing Choice: Learning Distributions over Permutations</b>	229
	Devavrat Shah	
8.1	Background	229
8.2	Setup	232
8.3	Models	234
8.4	Sparse Model	236
8.5	Random Utility Model (RUM)	247
8.6	Discussion	255
	References	258
<b>9</b>	<b>Universal Clustering</b>	263
	Ravi Kiran Raman and Lav R. Varshney	
9.1	Unsupervised Learning in Machine Learning	263
9.2	Universality in Information Theory	265
9.3	Clustering and Similarity	271
9.4	Distance-Based Clustering	276
9.5	Dependence-Based Clustering	280
9.6	Applications	291
9.7	Conclusion	292
	References	293
<b>10</b>	<b>Information-Theoretic Stability and Generalization</b>	302
	Maxim Raginsky, Alexander Rakhlin, and Aolin Xu	
10.1	Introduction	302
10.2	Preliminaries	305

10.3	Learning Algorithms and Stability	308
10.4	Information Stability and Generalization	311
10.5	Information-Theoretically Stable Learning Algorithms	317
	References	327
<b>11</b>	<b>Information Bottleneck and Representation Learning</b>	330
	Pablo Piantanida and Leonardo Rey Vega	
11.1	Introduction and Overview	330
11.2	Representation and Statistical Learning	333
11.3	Information-Theoretic Principles and Information Bottleneck	339
11.4	The Interplay between Information and Generalization	347
11.5	Summary and Outlook	356
	References	356
<b>12</b>	<b>Fundamental Limits in Model Selection for Modern Data Analysis</b>	359
	Jie Ding, Yuhong Yang, and Vahid Tarokh	
12.1	Introduction	359
12.2	Fundamental Limits in Model Selection	360
12.3	An Illustration on Fitting and the Optimal Model	363
12.4	The AIC, the BIC, and Other Related Criteria	365
12.5	The Bridge Criterion – Bridging the Conflicts between the AIC and the BIC	369
12.6	Modeling-Procedure Selection	375
12.7	Conclusion	379
	References	380
<b>13</b>	<b>Statistical Problems with Planted Structures: Information-Theoretical and Computational Limits</b>	383
	Yihong Wu and Jiaming Xu	
13.1	Introduction	383
13.2	Basic Setup	384
13.3	Information-Theoretic Limits	386
13.4	Computational Limits	401
13.5	Discussion and Open Problems	412
A13.1	Mutual Information Characterization of Correlated Recovery	415
A13.2	Proof of (13.7) $\Rightarrow$ (13.6) and Verification of (13.7) in the Binary Symmetric SBM	418
	References	419

---

<b>14</b>	<b>Distributed Statistical Inference with Compressed Data</b>	425
	Wenwen Zhao and Lifeng Lai	
14.1	Introduction	425
14.2	Basic Model	427
14.3	Interactive Communication	438
14.4	Identity Testing	445
14.5	Conclusion and Extensions	451
	References	452
<b>15</b>	<b>Network Functional Compression</b>	455
	Soheil Feizi and Muriel Médard	
15.1	Introduction	455
15.2	Problem Setup and Prior Work	458
15.3	Network Functional Compression	463
15.4	Discussion and Future Work	473
15.5	Proofs	476
	References	484
<b>16</b>	<b>An Introductory Guide to Fano's Inequality with Applications in Statistical Estimation</b>	487
	Jonathan Scarlett and Volkan Cevher	
16.1	Introduction	487
16.2	Fano's Inequality and Its Variants	492
16.3	Mutual Information Bounds	495
16.4	Applications – Discrete Settings	499
16.5	From Discrete to Continuous	510
16.6	Applications – Continuous Settings	516
16.7	Discussion	523
	References	525
	<i>Index</i>	529