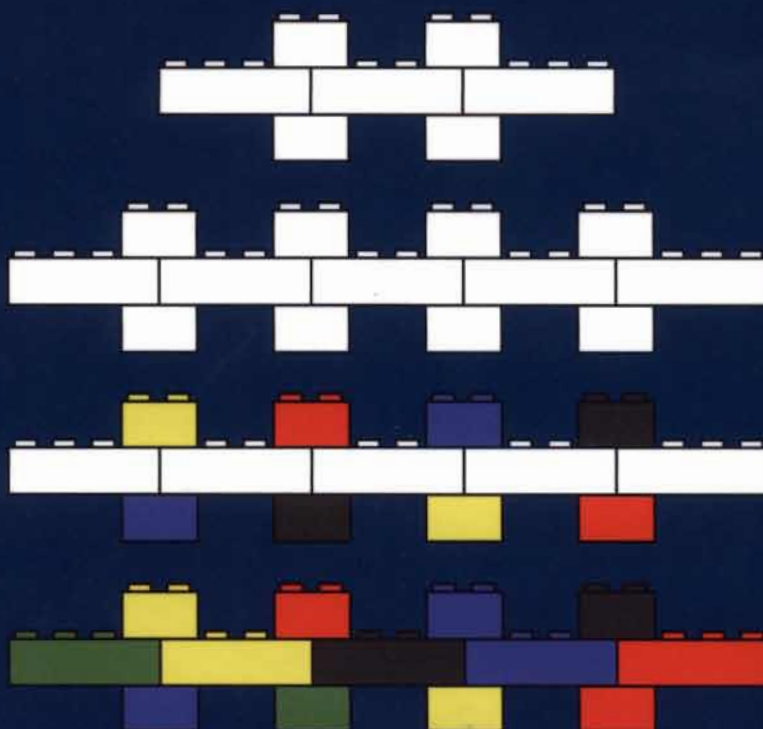


# Mathematics of Big Data

Spreadsheets, Databases, Matrices, and Graphs

Jeremy Kepner and Hayden Jananthan



Foreword by Charles E. Leiserson

## Mathematics of Big Data

## Contents

	<b>Foreword</b>	xi
	<b>Preface</b>	xiii
	<b>About the Authors</b>	xvii
	<b>About the Cover</b>	xix
	<b>Acknowledgments</b>	xxiii
<b>I</b>	<b>APPLICATIONS AND PRACTICE</b>	<b>1</b>
<b>1</b>	<b>Introduction and Overview</b>	<b>3</b>
	1.1 Mathematics of Data	3
	1.2 Data in the World	5
	1.3 Mathematical Foundations	9
	1.4 Making Data Rigorous	14
	1.5 Conclusions, Exercises, and References	16
<b>2</b>	<b>Perspectives on Data</b>	<b>19</b>
	2.1 Interrelations	19
	2.2 Spreadsheets	20
	2.3 Databases	22
	2.4 Matrices	26
	2.5 Graphs	27
	2.6 Map Reduce	29
	2.7 Other Perspectives	30
	2.8 Conclusions, Exercises, and References	31
<b>3</b>	<b>Dynamic Distributed Dimensional Data Model</b>	<b>37</b>
	3.1 Background	37
	3.2 Design	38
	3.3 Matrix Mathematics	39
	3.4 Common SQL, NoSQL, NewSQL Interface	40
	3.5 Key-Value Store Database Schema	41
	3.6 Data-Independent Analytics	44
	3.7 Parallel Performance	49
	3.8 Computing on Masked Data	51
	3.9 Conclusions, Exercises, and References	53

<b>4</b>	<b>Associative Arrays and Musical Metadata</b>	<b>57</b>
	4.1 Data and Metadata	57
	4.2 Dense Data	58
	4.3 Dense Operations	60
	4.4 Sparse Data	62
	4.5 Sparse Operations	63
	4.6 Conclusions, Exercises, and References	65
<b>5</b>	<b>Associative Arrays and Abstract Art</b>	<b>69</b>
	5.1 Visual Abstraction	69
	5.2 Minimal Adjacency Array	71
	5.3 Symmetric Adjacency Array	73
	5.4 Weighted Adjacency Array	75
	5.5 Incidence Array	75
	5.6 Conclusions, Exercises, and References	78
<b>6</b>	<b>Manipulating Graphs with Matrices</b>	<b>81</b>
	6.1 Introduction	81
	6.2 Matrix Indices and Values	86
	6.3 Composable Graph Operations and Linear Systems	89
	6.4 Matrix Graph Operations Overview	96
	6.5 Graph Algorithms and Diverse Semirings	105
	6.6 Conclusions, Exercises, and References	108
<b>7</b>	<b>Graph Analysis and Machine Learning Systems</b>	<b>115</b>
	7.1 Introduction	115
	7.2 Data Representation	116
	7.3 Graph Construction	118
	7.4 Adjacency Array Graph Traversal	120
	7.5 Incidence Array Graph Traversal	122
	7.6 Vertex Degree Centrality	126
	7.7 Edge Degree Centrality	129
	7.8 Eigenvector Centrality	129
	7.9 Singular Value Decomposition	133
	7.10 PageRank	136
	7.11 Deep Neural Networks	138
	7.12 Conclusions, Exercises, and References	140
<b>II</b>	<b>MATHEMATICAL FOUNDATIONS</b>	<b>145</b>
<b>8</b>	<b>Visualizing the Algebra of Associative Arrays</b>	<b>147</b>
	8.1 Associative Array Analogs of Matrix Operations	147
	8.2 Abstract Algebra for Computer Scientists and Engineers	150
	8.3 Depicting Mathematics	152
	8.4 Associative Array Class Diagrams	153

8.5	Set	154
8.6	Semiring	155
8.7	Linear Algebra	158
8.8	Ordered Sets	160
8.9	Boolean Algebra	162
8.10	Associative Array Algebra	164
8.11	Conclusions, Exercises, and References	164
<b>9</b>	<b>Defining the Algebra of Associative Arrays</b>	<b>169</b>
9.1	Operations on Sets	169
9.2	Ordered Sets	175
9.3	Supremum and Infimum	177
9.4	Lattice	181
9.5	The Semirings of Interest	186
9.6	Conclusions, Exercises, and References	189
<b>10</b>	<b>Structural Properties of Associative Arrays</b>	<b>193</b>
10.1	Estimating Structure	193
10.2	Associative Array Formal Definition	194
10.3	Padding Associative Arrays with Zeros	197
10.4	Zero, Null, Zero-Sum-Free	198
10.5	Properties of Matrices and Associative Arrays	199
10.6	Properties of Zero Padding	201
10.7	Support and Size	207
10.8	Image and Rank	208
10.9	Example: Music	209
10.10	Example: Art	211
10.11	Properties of Element-Wise Addition	213
10.12	Properties of Element-Wise Multiplication	217
10.13	Array Multiplication	221
10.14	Closure of Operations between Arrays	228
10.15	Conclusions, Exercises, and References	229
<b>11</b>	<b>Graph Construction and Graphical Patterns</b>	<b>235</b>
11.1	Introduction	235
11.2	Adjacency and Incidence Array Definitions	236
11.3	Adjacency Array Construction	242
11.4	Graph Construction with Different Semirings	250
11.5	Special Arrays and Graphs	255
11.6	Key Ordering	258
11.7	Algebraic Properties	263
11.8	Subobject Properties	264
11.9	Conclusions, Exercises, and References	266

III	<b>LINEAR SYSTEMS</b>	<b>269</b>
12	<b>Survey of Common Transformations</b>	<b>271</b>
	12.1 Array Transformations	271
	12.2 Identity	274
	12.3 Contraction	290
	12.4 Stretching	293
	12.5 Rotation	297
	12.6 Conclusions, Exercises, and References	299
13	<b>Maps and Bases</b>	<b>303</b>
	13.1 Semimodules	303
	13.2 Linear Maps	307
	13.3 Linear Independence and Bases	309
	13.4 Existence of Bases	312
	13.5 Size of Bases	313
	13.6 Semialgebras and the Algebra of Arrays	317
	13.7 Conclusions, Exercises, and References	320
14	<b>Linearity of Associative Arrays</b>	<b>323</b>
	14.1 The Null Space of Linear Maps	323
	14.2 Supremum-Blank Algebras	326
	14.3 Max-Blank Structure Theorem	334
	14.4 Examples of Supremum-Blank Algebras	338
	14.5 Explicit Computations of $x(\mathbf{A}, \mathbf{w})$ for Supremum-Blank Algebras	342
	14.6 Conclusions, Exercises, and References	348
15	<b>Eigenvalues and Eigenvectors</b>	<b>351</b>
	15.1 Introduction	351
	15.2 Quasi-Inverses	353
	15.3 Existence of Eigenvalues for Idempotent Multiplication	359
	15.4 Strong Dependence and Characteristic Bipolynomial	360
	15.5 Eigenanalysis for Irreducible Matrices for Invertible Multiplication	367
	15.6 Eigen-Semimodules	373
	15.7 Singular Value Decomposition	378
	15.8 Conclusions, Exercises, and References	385
16	<b>Higher Dimensions</b>	<b>389</b>
	16.1 $d$ -Dimensional Associative Arrays	389
	16.2 Key Ordering and Two-Dimensional Projections	392
	16.3 Algebraic Properties	398
	16.4 Sub-Array Properties	400
	16.5 Conclusions, Exercises, and References	402
	<b>Appendix: Notation</b>	<b>405</b>
	<b>Index</b>	<b>413</b>