

# Table of Contents

<b>Deep Learning and Neural Nets</b>	<b>iii</b>
<b>Preface and Acknowledgments</b>	<b>vi</b>
<b>Part I : Highlights of Linear Algebra</b>	<b>1</b>
I.1 Multiplication $Ax$ Using Columns of $A$ . . . . .	2
I.2 Matrix-Matrix Multiplication $AB$ . . . . .	9
I.3 The Four Fundamental Subspaces . . . . .	14
I.4 Elimination and $A = LU$ . . . . .	21
I.5 Orthogonal Matrices and Subspaces . . . . .	29
I.6 Eigenvalues and Eigenvectors . . . . .	36
I.7 Symmetric Positive Definite Matrices . . . . .	44
I.8 Singular Values and Singular Vectors in the SVD . . . . .	56
I.9 Principal Components and the Best Low Rank Matrix . . . . .	71
I.10 Rayleigh Quotients and Generalized Eigenvalues . . . . .	81
I.11 Norms of Vectors and Functions and Matrices . . . . .	88
I.12 Factoring Matrices and Tensors : Positive and Sparse . . . . .	97
<b>Part II : Computations with Large Matrices</b>	<b>113</b>
II.1 Numerical Linear Algebra . . . . .	115
II.2 Least Squares : Four Ways . . . . .	124
II.3 Three Bases for the Column Space . . . . .	138
II.4 Randomized Linear Algebra . . . . .	146

<b>Part III: Low Rank and Compressed Sensing</b>	<b>159</b>
III.1 Changes in $A^{-1}$ from Changes in $A$ . . . . .	160
III.2 Interlacing Eigenvalues and Low Rank Signals . . . . .	168
III.3 Rapidly Decaying Singular Values . . . . .	178
III.4 Split Algorithms for $\ell^2 + \ell^1$ . . . . .	184
III.5 Compressed Sensing and Matrix Completion . . . . .	195
<b>Part IV: Special Matrices</b>	<b>203</b>
IV.1 Fourier Transforms: Discrete and Continuous . . . . .	204
IV.2 Shift Matrices and Circulant Matrices . . . . .	213
IV.3 The Kronecker Product $A \otimes B$ . . . . .	221
IV.4 Sine and Cosine Transforms from Kronecker Sums . . . . .	228
IV.5 Toeplitz Matrices and Shift Invariant Filters . . . . .	232
IV.6 Graphs and Laplacians and Kirchhoff's Laws . . . . .	239
IV.7 Clustering by Spectral Methods and $k$ -means . . . . .	245
IV.8 Completing Rank One Matrices . . . . .	255
IV.9 The Orthogonal Procrustes Problem . . . . .	257
IV.10 Distance Matrices . . . . .	259
<b>Part V: Probability and Statistics</b>	<b>263</b>
V.1 Mean, Variance, and Probability . . . . .	264
V.2 Probability Distributions . . . . .	275
V.3 Moments, Cumulants, and Inequalities of Statistics . . . . .	284
V.4 Covariance Matrices and Joint Probabilities . . . . .	294
V.5 Multivariate Gaussian and Weighted Least Squares . . . . .	304
V.6 Markov Chains . . . . .	311

<b>Part VI: Optimization</b>	<b>321</b>
VI.1 Minimum Problems: Convexity and Newton's Method . . . . .	324
VI.2 Lagrange Multipliers = Derivatives of the Cost . . . . .	333
VI.3 Linear Programming, Game Theory, and Duality . . . . .	338
VI.4 Gradient Descent Toward the Minimum . . . . .	344
VI.5 Stochastic Gradient Descent and ADAM . . . . .	359
<b>Part VII: Learning from Data</b>	<b>371</b>
VII.1 The Construction of Deep Neural Networks . . . . .	375
VII.2 Convolutional Neural Nets . . . . .	387
VII.3 Backpropagation and the Chain Rule . . . . .	397
VII.4 Hyperparameters: The Fateful Decisions . . . . .	407
VII.5 The World of Machine Learning . . . . .	413
<b>Books on Machine Learning</b>	<b>416</b>
<b>Image Compression by the SVD</b>	<b>417</b>
<b>Codes and Algorithms for Numerical Linear Algebra</b>	<b>418</b>
<b>Counting Parameters in the Basic Factorizations</b>	<b>419</b>
<b>Index of Authors</b>	<b>420</b>
<b>Index</b>	<b>423</b>
<b>Index of Symbols</b>	<b>432</b>