This text is expected to appear in Summer 2018
You could express interest by a note to learningfromdata1@gmail.com
This would not be an order – it would bring updated information
Thank you

LINEAR ALGEBRA AND
LEARNING FROM DATA

GILBERT STRANG
Massachusetts Institute of Technology

WELLESLEY - CAMBRIDGE PRESS
Box 812060 Wellesley MA 02482
# Table of Contents

## Part I: Highlights of Linear Algebra

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.1</td>
<td>Multiplication $Ax$ Using Columns of $A$</td>
<td>2</td>
</tr>
<tr>
<td>I.2</td>
<td>Matrix-Matrix Multiplication $AB$</td>
<td>9</td>
</tr>
<tr>
<td>I.3</td>
<td>The Four Fundamental Subspaces</td>
<td>14</td>
</tr>
<tr>
<td>I.4</td>
<td>Elimination and $A = LU$</td>
<td>20</td>
</tr>
<tr>
<td>I.5</td>
<td>Orthogonal Matrices and Subspaces</td>
<td>27</td>
</tr>
<tr>
<td>I.6</td>
<td>Eigenvalues and Eigenvectors</td>
<td>34</td>
</tr>
<tr>
<td>I.7</td>
<td>Symmetric Positive Definite Matrices</td>
<td>40</td>
</tr>
<tr>
<td>I.8</td>
<td>Singular Values and Singular Vectors in the SVD</td>
<td>48</td>
</tr>
<tr>
<td>I.9</td>
<td>Principal Components and the Best Low Rank Matrix</td>
<td>63</td>
</tr>
<tr>
<td>I.10</td>
<td>Rayleigh Quotients and Generalized Eigenvalues</td>
<td>72</td>
</tr>
<tr>
<td>I.11</td>
<td>Norms of Vectors and Functions and Matrices</td>
<td>79</td>
</tr>
<tr>
<td>I.12</td>
<td>Tensors</td>
<td>87</td>
</tr>
</tbody>
</table>

## Part II: Computations with Large Matrices

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>II.1</td>
<td>Numerical Linear Algebra</td>
<td>94</td>
</tr>
<tr>
<td>II.2</td>
<td>Least Squares: Four Ways</td>
<td>103</td>
</tr>
<tr>
<td>II.3</td>
<td>Three Bases for the Column Space</td>
<td>115</td>
</tr>
<tr>
<td>II.4</td>
<td>Randomized Linear Algebra</td>
<td>121</td>
</tr>
</tbody>
</table>
Part III: Special Matrices

III.1 Fourier Transforms: Discrete and Continuous .......................... 134
III.2 Shift Matrices and Circulant Matrices ................................. 143
III.3 The Kronecker Product $A \otimes B$ ....................................... 151
III.4 Sine and Cosine Transforms from Kronecker Sums ................. 158
III.5 Toeplitz Matrices and Shift Invariant Filters ....................... 163
III.6 Graphs and Kirchhoff’s Laws and $A^TCA$ ............................... 169
III.7 Clustering by Spectral Methods and $k$-means ....................... 174
III.8 The Orthogonal Procrustes Problem .................................. 187
III.9 Distance Matrices .......................................................... 190

Part IV: Low Rank Matrices and Compressed Sensing

IV.1 Perturbations of Small Rank .................................................. 196
IV.2 Interlacing Eigenvalues and Low Rank Signals ..................... 200
IV.3 Rapidly Decaying Singular Values ....................................... 205
IV.4 Sparse and Low Rank: Compressed Sensing in $\ell^1$ ............. 211

Part V: Probability and Statistics

V.1 Mean, Variance, and Probability .......................................... 220
V.2 Probability Distributions ....................................................... 232
V.3 Moments, Cumulants, and Inequalities of Statistics ................ 242
V.4 Covariance Matrices and Joint Probabilities ......................... 248
V.5 Multivariate Gaussian and Weighted Least Squares ............... 258
V.6 Markov Chains ................................................................. 266
V.7 Independent Component Analysis ......................................... 275
Table of Contents

**Part VI: Optimization**

VI.1 Minimum Problems: Convexity and Newton’s Method . . . . . . 282
VI.2 Lagrange Multipliers = Derivatives of the Cost . . . . . . . . 288
VI.3 Linear Programming, Game Theory, and Duality . . . . . . . 294
VI.4 Gradient Descent Toward the Minimum . . . . . . . . . . 299
VI.5 Stochastic Gradient Descent . . . . . . . . . . . . . . . . 315

**Part VII: Learning from Data**

VII.1 The Construction of Deep Neural Networks . . . . . . . . . . 322
VII.2 Backpropagation and the Chain Rule . . . . . . . . . . . . . . 331
VII.3 Hyperparameters: The Fateful Decisions . . . . . . . . . . 337
VII.4 Support Vector Machines and Nonlinear Kernels . . . . . . . 338

Eigenvalues and Singular Values: Rank One

Key Algorithms of Numerical Linear Algebra