Preface

In recent years, computer vision, robotics, machine learning, and data science have been some of the key areas that have contributed to major advances in technology. Anyone who looks at papers or books in the above areas will be baffled by a strange jargon involving exotic terms such as kernel PCA, ridge regression, lasso regression, support vector machines (SVM), Lagrange multipliers, KKT conditions, *etc.* Do support vector machines chase cattle to catch them with some kind of super lasso? No! But one will quickly discover that behind the jargon which always comes with a new field (perhaps to keep the outsiders out of the club), lies a lot of "classical" linear algebra and techniques from optimization theory. And there comes the main challenge: in order to understand and use tools from machine learning, computer vision, and so on, one needs to have a firm background in linear algebra and optimization theory. To be honest, some probablity theory and statistics should also be included, but we already have enough to contend with.

Many books on machine learning struggle with the above problem. How can one understand what are the dual variables of a ridge regression problem if one doesn't know about the Lagrangian duality framework? Similarly, how is it possible to discuss the dual formulation of SVM without a firm understanding of the Lagrangian framework?

The easy way out is to sweep these difficulties under the rug. If one is just a consumer of the techniques we mentioned above, the cookbook recipe approach is probably adequate. But this approach doesn't work for someone who really wants to do serious research and make significant contributions. To do so, we believe that one must have a solid background in linear algebra and optimization theory.

This is a problem because it means investing a great deal of time and energy studying these fields, but we believe that perseverance will be amply rewarded.

Our main goal is to present fundamentals of linear algebra and optimization theory, keeping in mind applications to machine learning, robotics, and computer vision. This work consists of two volumes, the first one being linear algebra, the second one optimization theory and applications, especially to machine learning.

This first volume covers "classical" linear algebra, up to and including the primary decomposition and the Jordan form. Besides covering the standard topics, we discuss a few topics that are important for applications. These include:

1. Haar bases and the corresponding Haar wavelets.

2. Hadamard matrices.

3. Affine maps (see Section 5.4).

4. Norms and matrix norms (Chapter 8).

5. Convergence of sequences and series in a normed vector space. The matrix exponential e^A and its basic properties (see Section 8.8).

6. The group of unit quaternions, $\mathbf{SU}(2)$, and the representation of rotations in $\mathbf{SO}(3)$ by unit quaternions (Chapter 15).

7. An introduction to algebraic and spectral graph theory.

8. Applications of SVD and pseudo-inverses, in particular, principal component analysis, for short PCA (Chapter 21).

9. Methods for computing eigenvalues and eigenvectors, with a main focus on the QR algorithm (Chapter 17).

Four topics are covered in more detail than usual. These are

1. Duality (Chapter 10).

2. Dual norms (Section 13.7).

3. The geometry of the orthogonal groups $\mathbf{O}(n)$ and $\mathbf{SO}(n)$, and of the unitary groups $\mathbf{U}(n)$ and $\mathbf{SU}(n)$.

4. The spectral theorems (Chapter 16).

Except for a few exceptions we provide complete proofs. We did so to make this book self-contained, but also because we believe that no deep knowledge of this material can be acquired without working out some proofs. However, our advice is to skip some of the proofs upon first reading, especially if they are long and intricate.

The chapters or sections marked with the symbol ⊛ contain material that is typically more specialized or more advanced, and they can be omitted upon first (or second) reading.