Geometric Data Analysis
Geometric Data Analysis

An Empirical Approach to Dimensionality Reduction and the Study of Patterns

Michael Kirby
Preface

This text concerns the dimensionality reduction of large data sets. In many instances, the initial high-dimensional setting of a collection of scalar or vector fields, such as digital images or snapshots of the complete state of a fluid flow, is a consequence of a poor coordinate system. Often such data actually afford a significantly reduced representation. This motivates the search for a basis which will hopefully come close to an optimal representation of the data. The search for an optimal, or at least good basis, for the data leads us to the development of empirical transform techniques, extending available analytical tools such as the Fourier or Laplace transform. Here the word empirical reflects the fact that the mappings are to be computed directly from the data of interest. This approach involves more work at the outset given that the transforms and their inverses need to be computed. However, because they are intimately related to the data, these transforms often provide special insight into the structure of the data. Indeed, it is not unusual for an empirical transform to reveal basic facts concerning the data which are not at all apparent from a standard analytical approach.

The title for this text, Geometric Data Analysis, reflects the emphasis on the geometric perspective employed throughout. A pattern is an $n$-tuple, or a point in a vector space, while a collection of patterns is viewed as a geometrical object, i.e., a subspace or a submanifold of a larger vector space. It is the manner in which the data occupies the space where it resides that dictates the

\footnote{We are specifically interested in constructing mappings for which an inverse exists.}
method for representing it. For example, if data densely occupies a region of a two dimensional plane, then it is sensible to employ a subspace description. On the other hand, if the data resides on a closed curve in $\mathbb{R}^3$, then a submanifold description may be more appropriate. These ideas complement the statistical perspective which is traditional in data analysis. Certainly the role of noise in data is important and the geometric perspective cannot stand alone.

Intended for students and researchers whose work involves analyzing patterns in large, high-dimensional data sets, this text has grown out of a course entitled “A Geometric Approach to Pattern Analysis” which has been developed over the last decade at Colorado State University. The typical audience has ranged from first year graduate students to faculty from a broad variety of disciplines including Computer Science, Physics, Engineering, Natural Resources, Atmospheric Sciences and Mathematics. As such, this course has been taught to students with diverse mathematical backgrounds. With this in mind, a significant amount of prerequisite material is scattered throughout the text to make it as self-contained and accessible as possible. As a result, much of the text should be readable by advanced undergraduates who have had a good course in linear algebra. Some knowledge of function spaces is helpful for the sections on the continuous Karhunen-Loève expansion and for the chapter on Wavelets. It is possible, however, to investigate most of the basic ideas within the context of discrete data.

In the brief first chapter the underlying concepts and the associated mathematical framework of dimensionality reduction are presented. Data sources such as numerical simulations of physical models, laboratory experiments and digital imaging systems are discussed as are general issues concerning the nature of data.

The notes are intended to be as self-contained as possible. Thus, in Chapter 2, we review the basic mathematics required in later chapters. Coordinate transformations, change of bases, inner product spaces, subspace operations, the spectral theorem and the singular value decomposition (SVD) are presented. Old questions may be asked in new ways, such as: Does a collection of digital images of human faces form a vector space? The advanced reader is encouraged to skim this chapter and return to it as necessary.

Part II develops the idea of optimal dimensionality reducing mappings in the linear setting. The mappings are also global.

In Chapter 3 we introduce one of the most important tools for dimensionality reduction, namely, the Karhunen-Loève (K-L) transform. We begin the discussion of an example from Pearson’s 1901 paper. After a detailed theoretical presentation which develops the procedure for high-dimensional data sets, we present an application to high-resolution image representation in the context of human face identification. We also present an application to the processing of sequences of time-dependent high-resolution digital images. The main approaches developed i.e., the snapshot method and the direct method, are also discussed in terms of the singular value decomposition.
Chapter 4 continues the discussion of the K-L procedure with more applications, for example, to missing or noisy data. The continuous formulation and the benefits of the exploitation of symmetry are also discussed and a methodology for computing symmetric eigenpictures is presented and applied. The problem of reducing the dimensionality of the dynamical description of physical models, currently an active research area, is also considered.

Part III concerns the time, frequency and scale analysis. Unlike other material in this text, the methods here are all analytical, i.e., the transforms are available in closed form.

Chapter 5 introduces one of the most important tools in applied mathematics and pattern analysis, i.e., the discrete Fourier transform (DFT). The presentation is within the framework of finite orthogonal expansions and the DFT is shown to be a special case of the KL expansion. The short-time Fourier transform is developed to motivate the discussion of an adapted window transform and wavelets.

Chapter 6 provides an introduction to wavelet analysis. The continuous wavelet transform is presented to assist in understanding the discrete wavelet transform and multi-resolution analysis. The basic ideas are presented in the context of the Haar wavelet. The pyramidal algorithm is developed and applied in both one and two dimensional settings, i.e., time-series and image analysis.

Part IV concerns adaptive nonlinear mappings and their application to the construction of empirical dimensionality reducing transformations.

The first two chapters in part IV introduce method for the approximation of non-linear functions which may be applied to high-dimensional data sets. Chapter 7 presents radial basis functions and clustering algorithms. The self-organizing feature map is presented as a clustering routine based on a competitive learning algorithm. Chapter 8 is an introduction to sigmoidal neural networks. We focus primarily on feed-forward networks and the well-known back-propagation training procedure. Again the application is to the approximation of nonlinear functions. The techniques in this chapter are global in nature.

The final chapter presents an overview of dimensionality reducing architectures and relies heavily on material developed in previous chapters. Indeed, the selection of material in the previous chapters was heavily motivated by the needs of this final chapter. The global procedures include the bottleneck neural network, the Whitney reduction network, while the local procedures include a variation of the KL transform and the method of neural charts. This material is the subject of active research and most of the ideas have appeared since 1985; I have been involved with many of them.

The study of empirical mappings, by its very nature, is rooted to computation. Most of the techniques and algorithms presented here require a computer for implementation. To this end, in addition to traditional problem sets, a set of computer experiments has been included at the end of each chapter. These vary in difficulty from easy to project length but none require
a supercomputer. Even the application of the KL procedure for characteri-
zation of human faces was carried out on a PC, in 1985! Students have used
many programming languages to complete these assignments and over the
years MATLAB has appeared as a favorite for speed of implementation and
display.

The topics covered are heavily biased by areas in which I have either worked
directly or I felt were directly connected to my research interests and obviously
many interesting and important topics are omitted. These notes were the re-
result of the fact that there was no textbook available which simultaneously
covered the Karhunen-Loève (KL) expansion with many variations and applica-
tions; Fourier and wavelet analysis; radial basis functions, clustering and
sigmoidal networks as well as a general discussion of dimensionality reduction
transformations. Hopefully this text will fill a void in this area.

Michael Kirby

Fort Collins
The support for this project by the Department of Mathematics at Colorado State University (CSU) is gratefully acknowledged. In addition, many students at CSU have contributed to these notes through their active class participation and hard work on the assignments. Many of the problems were generated by classroom discussions.

I would like to thank the numerous funding sources which have supported my work over the years including the Alexander von Humboldt Foundation; the National Science Foundation programs in neuroengineering, computational mathematics and international travel; the English Scientific and Physical Research Council; the Air Force Office of Sponsored Research and Honeywell Corporation.

Many collaborators have also contributed significantly to the material in these notes. In particular, Larry Sirowich introduced the KL procedure to me and proposed the Rogues Gallery problem in 1985 as a project which became part of my Ph.D. thesis. In addition to providing many figures for the manuscript, Doug Hundley also critically read several drafts and provided valuable input.

M. J. K.