

Preface

The ability to simplify means to eliminate the unnecessary
so that the necessary may speak.

—Hans Hofmann, 1880–1966

The last few years have seen a great increase in the amount of data available to scientists, engineers, and researchers from many disciplines. Datasets with millions of objects and hundreds, if not thousands, of measurements are now commonplace in areas such as image analysis, computational finance, bio-informatics, and astrophysics. This large volume of data does, however, come at a price, more often than not many computational techniques used to analyze these datasets cannot cope with such large data. Therefore, strategies need to be employed as a pre-processing step to reduce the number of objects, or measurements, whilst retaining important information inherent to the data. One of the key problems with such datasets is how to reduce the number of measurements, often referred to as dimensions, in such a way that the reduced set of measurements captures the main properties of the original data. Spectral dimensionality reduction is one such family of methods that has proven to be an indispensable tool in the data processing pipeline. In recent years, the area has gained much attention; thanks to the development of nonlinear spectral dimensionality reduction methods, often referred to as manifold learning algorithms.

Spectral dimensionality reduction methods can be broadly split into two categories; those that seek to maintain linear properties in the data, and those that seek to maintain nonlinear, manifold, properties. Both linear and nonlinear methods achieve the reduction in dimensionality through the careful construction of a feature matrix, the spectral decomposition of which gives rise to the reduced dimensionality dataset. Ever since the first nonlinear spectral dimensionality reduction methods were proposed over a decade ago, numerous algorithms and improvements have been proposed for the purpose of performing spectral dimensionality reduction. Although these algorithms may improve and extend existing techniques, there is still no gold standard technique. The reasons for this are many; however, one of the core problems with the area is that there are still many obstacles that need to be overcome before spectral dimensionality reduction that can be applied to a specific problem area. These obstacles, referred to herein

as *open problems*, have implications for those without a background in the area who wish to employ spectral dimensionality reduction to their problem domain.

Those wish to use spectral dimensionality reduction without prior knowledge of the field will immediately be confronted with questions that need answering; what parameter values to use? how many dimensions should the data be embedded into? how are new data points incorporated? what about large-scale data? For many, a search of the literature to find answers to these questions is impractical, as such, there is a need for a concise discussion into the problems themselves, how they affect spectral dimensionality reduction and how these problems can be overcome.

This book provides a survey and reference aimed at advanced undergraduate and postgraduate students as well as researchers, scientists, and engineers in a wide range of disciplines. Dimensionality reduction has proven useful in a wide range of problem domains, and so this book will be applicable to anyone with a solid grounding in statistics and computer science seeking to apply spectral dimensionality to their work.

Acknowledgments

We both would like to thank the love and support of Jo, Diana, and our families. As well as to this, we would like to acknowledge both NISCHR and the Aberystwyth APRS funding schemes for partially supporting this work.

Aberystwyth, October 2013

Harry Strange
Reyer Zwiggelaar

Open Problems in Spectral Dimensionality Reduction

Strange, H.; Zwiggelaar, R.

2014, IX, 92 p. 20 illus., 15 illus. in color., Softcover

ISBN: 978-3-319-03942-8