

---

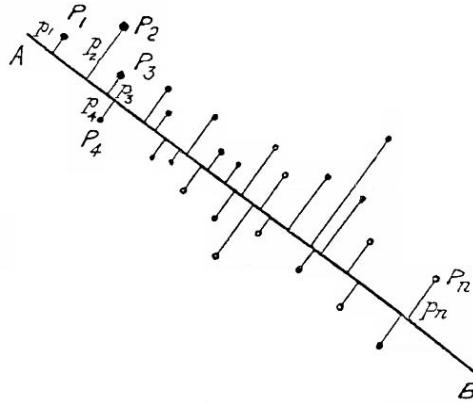
## Preface

In 1901, Karl Pearson [1] explained to the scientific community that the problem of data approximation is (i) important and (ii) nice, and (iii) differs from the regression problem. He demonstrated how to approximate data sets with straight lines and planes. That is, he invented Principal Component Analysis (PCA). Why and when do we need to solve the data approximation problem instead of regression? Let us look at Pearson's explanation:

"(1) In many physical, statistical, and biological investigations it is desirable to represent a system of points in plane, three, or higher dimensioned space by the "best-fitting" straight line or plane. Analytically this consists in taking

$$y = a_0 + a_1x, \text{ or } z = a_0 + a_1x + b_1y,$$
$$\text{or } z = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + \dots + a_nx_n,$$

where  $y, x, z, x_1, x_2, \dots, x_n$  are variables, and determining the "best" values for constants  $a_0, a_1, b_1, a_0, a_1, a_2, \dots, a_n$  in relation to the observed corresponding values of the variables. In nearly all the cases dealt with in the text-books of least squares, the variables on the right of our equations are treated as the independent, those on the left as the dependent variables. The result of this treatment is that we get one straight line or plane if we treat some one variable as independent, and a quite different one if we treat another variable as the independent variable. There is no paradox about this; it is, in fact, an easily understood and most important feature of the theory of a system of correlated variables. The most probable value of  $y$  for a given value of  $x$ , say, is not given by the same relation as the most probable value of  $x$  for the given value of  $y$ . Or, to take a concrete example, the most probable stature of a man with a given length of leg  $l$  being  $s$ , the most probable length of leg for a man with stature  $s$  will not be  $l$ . The "best-fitting" lines and planes ... depend upon



**Fig. 1.** Data approximation by a straight line. The famous illustration from Pearson's paper [1]

a determination of the means, standard-deviations, and correlation-coefficients of the system. In such cases the values of the independent variables are supposed to be accurately known, and the probable value of the dependent variable is ascertained.

(2) In many cases of physics and biology, however, the “independent” variable is subject to just as much deviation or error as the “dependent” variable. We do not, for example, know  $x$  accurately and then proceed to find  $y$ , but both  $x$  and  $y$  are found by experiment or observation. We observe  $x$  and  $y$  and seek for a unique functional relation between them. Men of given stature may have a variety of leg-length; but a point at a given time will have one position only, although our observation of *both* time and position may be in error, and vary from experiment to experiment. In the case we are about to deal with, we suppose the observed variables – all subject to error – to be plotted in plane, three-dimensioned or higher space, and we endeavour to take a line (or plane) which will be the “best fit” to such system of points.

Of course the term “best fit” is really arbitrary; but a good fit will clearly be obtained if we make the sum of the squares of the perpendiculars from the system of points upon the line or plane a minimum.

For example:—Let  $P_1, P_2, \dots, P_n$  be the system of points with coordinates  $x_1, y_1; x_2, y_2; \dots, x_n, y_n$ , and perpendicular distances  $p_1, p_2, \dots, p_n$  from a line  $AB$ . Then we shall make<sup>1</sup>

$$U = S(p^2) = \text{a minimum.}''$$

<sup>1</sup>  $S(p^2)$  stands for  $\sum_i p_i^2$

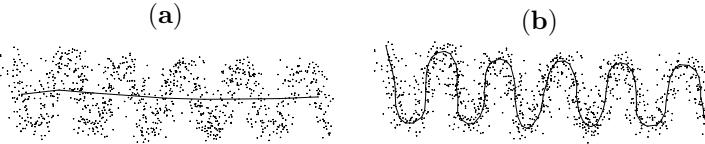
This explanation sounds very modern: in “many cases of physics and biology” there is significant noise in the “independent variables”, and it appears better to approximate data points than the regression functions that transform one set of data coordinates (the “independent variables”) into another. “Of course the term “best fit” is really arbitrary”, but the least squares approach remains the method of choice, if there exist no strong arguments for another choice of metrics. This method was applied to many problems, has been transformed and rediscovered several times, and is now known under several names: mostly as PCA or as *proper orthogonal decomposition*. But the main idea remains the same: we approximate the data set by a point (this is the mean point), then by a line (first principal component), then by a plane, etc.

What was invented in the data approximation during the century? First of all, the approximation by linear manifolds (lines, planes, ...) was supplemented by a rich choice of the approximate objects. The important discovery is the approximation of a data set by a smaller finite set of “centroids”. In the least squares approach to the best fit this gives the famous  $K$ -means algorithm [2]. Usually, this method is discussed as a clustering algorithm, but its application field is much wider. It is useful for adaptive coding and data binning, and is a model reduction method, as well as the PCA: the PCA allows us to substitute a high-dimensional vector by its projection on a best fitted low-dimensional linear manifold, the  $K$ -means approach gives an approximation of a big data set by  $K$  best fitted centroids.

Between the “most rigid” linear manifolds and “most soft” unstructured finite sets there is the whole universe of approximants. If we change the PCA linear manifolds to algebraic curves and manifolds, then a branch of the *algebraic statistics* appears. This field is still relatively new (less than ten years old) [3]. Algebraic curves and manifolds are much more flexible than linear ones, but remain rigid in the following sense: it is impossible to change the algebraic manifold locally, only near a point. Differential manifolds give more freedom, but require specific efforts for regularization.

A step from absolute flexibility of finite sets gives the Self-Organizing Map (SOM) approach [4]. SOM can be formalized either as a manifold learner which represents the manifold as a discretized grid, or a  $K$ -means-like clustering algorithm which adds a topology to the cluster centroids. Although SOM has been slowly replaced by theoretically better founded and better behaving algorithms, its simplicity and computational efficiency makes it one of the most popular data analysis techniques even today. An important improvement of SOM came with the introduction of the Generative Topographic Mapping (GTM) [6], establishing a probabilistic framework and a well-defined objective function. The generative probabilistic model has also become an analytical tool to formalize the faithfulness-conciseness trade-off.

Another big shift of the century is the appearance of the whole framework of machine learning which significantly extends Pearson’s initial “geometrical” approach. It is a common practice in general discussions on machine learning



**Fig. 2.** An ill-defined unsupervised learning problem. Which curve describes the data better, (a) a short curve that is “far” from the data, or (b) a long curve that follows the data closely?

to use the dichotomy of supervised and unsupervised learning to categorize learning methods. Supervised learning algorithms assume that a training set of *(input, output)* observations is given (e.g., digitized images of characters and their class labels). The goal is then to learn a function that predicts the output for previously unseen input patterns. This is a very far generalization of Pearson’s linear regression onto various types of inputs, outputs, and functions.

In unsupervised learning, we only have a set of (input) observations without a desired target, and the goal is to find an efficient representation of the data (for example by reducing the number of attributes or grouping the data into a small number of clusters), or to characterize the data-generating process.

From a conceptual point of view, unsupervised learning is substantially more difficult than supervised learning. Whereas in supervised learning the cost of mis-predicting the output provides a well-defined criteria to optimize, in unsupervised learning we often face a trade-off of representing the data as faithfully as possible while being as concise as possible (Fig. 2). In a certain sense, an unsupervised learner can be considered as a supervised learner where the target is the input itself. In other words, the task is to find a function as close to the identity function as possible. Of course, without restricting the set of admissible predictors this is a trivial problem. These restrictions originate from the other objective of unsupervised learning of finding a mapping which is simple in a certain sense. The trade-off between these two competing objectives depends on the particular problem.

Manifold learning is a sub-domain of unsupervised learning where the goal is to project the input data into a new space which is simpler in a certain sense than the input space, or in which the data distribution is more regular than originally.

Two distinct groups of methods exist for this purpose that differ in their way of representing the manifold. Thus, Non-linear PCA (NLPCA) extends PCA by replacing the linear encoder and decoder by non-linear functions (for example, feed-forward neural networks [7]), and optimizing them in an auto-encoding setup. The embedded manifold appears only implicitly as the decoded image of the input space, and the geometric notion of projection does not apply.

Principal curves and manifolds [8], on the other hand, extend the geometric interpretation of PCA by explicitly constructing an embedded manifold, and by encoding using standard geometric projection onto the manifold. How to define the “simplicity” of the manifold is problem-dependent, however, it is commonly measured by the intrinsic dimensionality and/or the smoothness of the manifold.

Clustering, another important sub-domain of unsupervised learning, can also be formalized in this framework: the clustering “manifold” is a finite partitioning of the input space, in the simplest case represented as a finite set of singular centroid points. Obviously, in this case simplicity can be measured neither by smoothness nor dimensionality, nevertheless, manifold learning and clustering methodologies are intimately connected both in their theoretical underpinning and on a technical-algorithmic level.

Most of the modern manifold learners find their theoretical and algorithmic roots in one of three basic and well-known data analysis techniques: PCA,  $K$ -means, and Multidimensional Scaling (MDS) [5] also known as Torgerson or Torgerson-Gower scaling. Thus, the basic loop of  $K$ -means that alternates between a projection and an optimization step became the algorithmic skeleton of many non-linear manifold learning algorithms. The SOM algorithm is arguably the torch holder of this batch of nonlinear manifold learners.

The objective of original MDS is somewhat different: find a linear projection that preserves pairwise distances as well as possible. The method does not explicitly construct an embedded manifold, but it has the important role of being the algorithmic forefather of “one-shot” (non-iterative) manifold learners. The most recent representatives of this approach are Local Linear Embedding (LLE) [9] and ISOMAP [10]. Both methods find their origins in MDS in the sense that their goal is to preserve pairwise relationships between data points. LLE conserves local linear patterns whereas ISOMAP applies MDS using the geodesic (manifold) distance approximated by the shortest path on the neighborhood graph (the graph constructed by connecting nearby points). Since the birth of these two methods, several neighborhood-graph-based techniques have emerged, stimulating the development of a common theory around Laplacian eigenmaps and spectral clustering and embedding.

Despite the significant progress made in the last decade, the manifold learning problem is far from being solved. The main drawback of iterative methods is that they are sensitive to initialization, and they can be stuck easily in suboptimal local minima, especially if the manifold is “loopy” or has a complicated topology. Neighborhood-graph-based “one-shot” techniques behave much better in this respect, their disadvantages are computational inefficiency (the complexity of the construction of the neighborhood graph by itself is quadratic in the number of data points) and increased sensitivity to noise around the manifold. One of today’s challenges in manifold learning is to find techniques that combine the advantages of these often incompatible approaches. Another exciting area is non-local manifold learning [11], which abandons two of the implicit premises of manifold learning: that manifolds are

smooth (locally linear) and that we have enough observations in every neighborhood to locally estimate the manifold. A third, very recent but promising, new domain is building deep networks (multiply nested functions) using an unsupervised paradigm (building all the layers except for the last using, for example, an autoassociative objective function [12]). These new areas share the ambitious goal of embedding manifold learning into artificial intelligence in a broad sense.

This book is a collection of reviews and original papers presented partially at the workshop “Principal manifolds for data cartography and dimension reduction” (Leicester, August 24-26, 2006). The problems of Large Data Sets analysis and visualisation, model reduction and the struggle with complexity of data sets are important for many areas of human activity. There exist many scientific and engineering communities that attack these problems from their own sides, and now special efforts are needed to organize communication between these groups, to support exchange of ideas and technology transfer among them. Heuristic algorithms and seminal ideas come from all application fields and from mathematics also, and mathematics has a special responsibility to find a solid basis for heuristics, to transform ideas into exact knowledge, and to transfer the resulting ideal technology to all the participants of the struggle with complexity. The workshop was focused on modern theory and methodology of geometric data analysis and model reduction. Mathematicians, engineers, software developers and advanced users from different areas of applications attended this workshop.

The first chapter of the book presents a general review of existing NLPCA algorithms (U. Kruger, J. Zhang, and L. Xie). Next, M. Scholz, M. Fraunholz, and J. Selbig focus attention on autoassociative neural network approach for NLPCA with applications to metabolite data analysis and gene expression analysis. H. Yin provides an overview on the SOM in the context of manifold learning. Its variant, the visualisation induced SOM (ViSOM) proposed for preserving local metric on the map, is introduced and reviewed for data visualisation. The relationships among the SOM, ViSOM, multidimensional scaling, and principal curves are analysed and discussed. A. Gorban and A. Zinovyev developed a general geometric framework for constructing “principal objects” of various dimensions and topologies with the simple quadratic form of the smoothness penalty. The approach was proposed in the middle of 1990s. It is based on mechanical analogy between principal manifolds and elastic membranes and plates.

M. Peña, W. Barbakh, and C. Fyfe present a family of topology preserving mappings similar to SOM and GTM. These techniques can be considered as a non-linear projection from input or data space to the output or latent space. B. Mirkin develops the iterative extraction approach to clustering and describes additive models for clustering entity-to-feature and similarity. This approach emerged within the PCA framework by extending the bilinear Singular Value Decomposition model to that of clustering.

In their contribution, J. Einbeck, L. Evers, and C. Bailer-Jones give a short review of localized versions of PCA, focusing on local principal curves and local partitioning algorithms. These methods can work with branched and disconnected principal components. S. Girard and S. Iovleff introduce auto-associative models, a new tool for building NLP PCA methods, and compare it to other modern methods. A. Gorban, N. Sumner, and A. Zinovyev propose new type of low-dimensional “principal object”: *principal cubic complex*, the product of one-dimensional branching principal components. This complex is a generalization of linear and non-linear principal manifolds and includes them as a particular case. To construct such an object, they combine the method of *topological grammars* with the minimization of elastic energy defined for its embedding into multidimensional data space.

B. Nadler, S. Lafon, R. Coifman, and I. G. Kevrekidis provide a diffusion based probabilistic analysis of embedding and clustering algorithms that use the normalized graph Laplacian. They define a random walk on the graph of points and a diffusion distance between any two points. The characteristic relaxation times and processes of the random walk on the graph govern the properties of spectral clustering and spectral embedding algorithms. Specifically, for spectral clustering to succeed, a necessary condition is that the mean exit times from each cluster need to be significantly larger than the largest (slowest) of all relaxation times inside all of the individual clusters. Diffusion metrics is studied also by S. Damelin in the context of the optimal discretization problem. He shows that a general notion of extremal energy defines a diffusion metric on  $X$  which is equivalent to a discrepancy on  $X$ . The diffusion metric is used to learn  $X$  via normalized graph Laplacian dimension reduction and the discrepancy is used to discretize  $X$ .

Modern biological applications inspire development of new approaches to data approximation. In many chapters biological applications play central role. For the comparison of various algorithms, several test datasets were selected and presented to the workshop participants. These datasets contain results of a high-throughput experimental technology application in molecular biology (microarray data). Principal component analysis and principal manifolds are useful methods for analysis of this kind of data, where the “curse of dimensionality” is an important issue. Because of it some variant of dimension reduction is absolutely required, for example, for regularization of classification problems that simply can not be solved otherwise. An interesting and underexplored question is: can non-linear principal manifolds serve better for this purpose as compared to the linear PCA or feature preselection?

M. Journée, A. E. Teschendorff, P.-A. Absil, S. Tavaré, and R. Sepulchre present an overview of the most popular algorithms to perform ICA. These algorithms are then applied on a microarray breast-cancer data set. D. Elizondo, B. N. Passow, R. Birkenhead, and A. Huemer present a comparison study of the performance of the linear principal component analysis and the non linear local tangent space alignment principal manifold methods to the problem of dimensionality reduction of microarray data.

The volume ends with a tutorial “PCA and  $K$ -Means decipher genome”. This exercise on principal component analysis and  $K$ -Means clustering can be used for courses of statistical methods and bioinformatics. By means of PCA students “discover” that the information in the genome is encoded by non-overlapping triplets. Next, they learn to find gene positions. In Appendix the MatLab program listings are presented.

The methods of data approximation, data visualization and model reduction developed during last century form an important part of the modern intellectual technology of data analysis and modeling. In this book we present some slices of this interdisciplinary technology and aim at eliminating some of the traditional language barriers that, unnecessarily sometimes, impede scientific cooperation and interaction of researchers across disciplines.

## References

1. Pearson, K.: On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, Ser. VI 2, 559–572 (1901)
2. MacQueen, J. B.: Some methods for classification and analysis of multivariate observations. In: *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1. University of California Press, Berkeley, 281–297 (1967)
3. Pachter, L., Sturmfels, B. (eds): *Algebraic Statistics for Computational Biology*, Cambridge University Press, Cambridge, United Kingdom (2005)
4. Kohonen, T.: *The Self-Organizing Map*. Springer, Berlin Heidelberg New York (1997)
5. Torgerson, W. S.: *Theory and Methods of Scaling*. Wiley, New York (1958)
6. Bishop, C. M., Svensén, M., and Williams, C. K. I.: The generative topographic mapping. *Neural Computation*, **10**, 215–235 (1998)
7. Kramer, M.A.: Nonlinear principal component analysis using autoassociative neural networks. *AIChE Journal*, **37**, 233–243 (1991)
8. Hastie, T. and Stuetzle, W.: Principal curves. *Journal of the American Statistical Association*, **84**, 502–516 (1989)
9. Roweis, S. and Saul L. K.: Nonlinear dimensionality reduction by locally linear embedding. *Science*, **290**, 2323–2326 (2000)
10. Tenenbaum, J. B., de Silva, V., and Langford J. C.: A global geometric framework for nonlinear dimensionality reduction. *Science*, **290**, 2319–2323 (2000)
11. Bengio, Y., Monperrus, M., and Larochelle, H.: Nonlocal estimation of manifold structure. *Neural Computation*, **18** 2509–2528 (2006)
12. Hinton, G. E. and Salakhutdinov, R.: Reducing the dimensionality of data with neural networks. *Science*, **313**, 504–507 (2006)

Leicester, UK  
 Orsay, France  
 Rolla, MO, USA  
 Paris, France

May, 2007

*Alexander N. Gorban  
 Balázs Kégl  
 Donald C. Wunsch  
 Andrei Y. Zinov'yev*

---

## Contents

### 1 Developments and Applications of Nonlinear Principal Component Analysis – a Review

<i>Uwe Kruger, Junping Zhang, Lei Xie</i> .....	1
1.1 Introduction .....	1
1.2 PCA Preliminaries .....	3
1.3 Nonlinearity Test for PCA Models .....	6
1.3.1 Assumptions .....	7
1.3.2 Disjunct Regions .....	7
1.3.3 Confidence Limits for Correlation Matrix .....	8
1.3.4 Accuracy Bounds .....	10
1.3.5 Summary of the Nonlinearity Test .....	11
1.3.6 Example Studies .....	12
1.4 Nonlinear PCA Extensions .....	16
1.4.1 Principal Curves and Manifolds .....	16
1.4.2 Neural Network Approaches .....	24
1.4.3 Kernel PCA .....	29
1.5 Analysis of Existing Work .....	31
1.5.1 Computational Issues .....	32
1.5.2 Generalization of Linear PCA? .....	34
1.5.3 Roadmap for Future Developments (Basics and Beyond) ..	37
1.6 Concluding Summary .....	38
References .....	39

### 2 Nonlinear Principal Component Analysis: Neural Network Models and Applications

<i>Matthias Scholz, Martin Fraunholz, Joachim Selbig</i> .....	45
2.1 Introduction .....	45
2.2 Standard Nonlinear PCA .....	48
2.3 Hierarchical Nonlinear PCA .....	49
2.3.1 The Hierarchical Error Function .....	51
2.4 Circular PCA .....	52

2.5	Inverse Model of Nonlinear PCA .....	53
2.5.1	The Inverse Network Model .....	55
2.5.2	NLPCA Models Applied to Circular Data .....	56
2.5.3	Inverse NLPCA for Missing Data .....	57
2.5.4	Missing Data Estimation .....	58
2.6	Applications .....	59
2.6.1	Application of Hierarchical NLPCA.....	60
2.6.2	Metabolite Data Analysis .....	61
2.6.3	Gene Expression Analysis .....	63
2.7	Summary.....	65
	References .....	66

### 3 Learning Nonlinear Principal Manifolds by Self-Organising Maps

<i>Hujun Yin</i> .....	69	
3.1	Introduction .....	69
3.2	Biological Background .....	70
3.2.1	Lateral Inhibition and Hebbian Learning .....	70
3.2.2	From Von Marsburg and Willshaw's Model to Kohonen's SOM .....	73
3.2.3	The SOM Algorithm.....	76
3.3	Theories .....	77
3.3.1	Convergence and Cost Functions .....	77
3.3.2	Topological Ordering Measures .....	80
3.4	SOMs, Multidimensional Scaling and Principal Manifolds .....	81
3.4.1	Multidimensional Scaling.....	81
3.4.2	Principal Manifolds.....	83
3.4.3	Visualisation Induced SOM (ViSOM) .....	85
3.5	Examples.....	87
3.5.1	Data Visualisation.....	88
3.5.2	Document Organisation and Content Management.....	88
	References .....	92

### 4 Elastic Maps and Nets for Approximating Principal Manifolds and Their Application to Microarray Data Visualization

<i>Alexander N. Gorban, Andrei Y. Zinovyev</i> .....	97	
4.1	Introduction and Overview .....	97
4.1.1	Fréchet Mean and Principal Objects: K-Means, PCA, what else? .....	97
4.1.2	Principal Manifolds.....	99
4.1.3	Elastic Functional and Elastic Nets .....	101
4.2	Optimization of Elastic Nets for Data Approximation .....	104
4.2.1	Basic Optimization Algorithm .....	104

4.2.2	Missing Data Values . . . . .	106
4.2.3	Adaptive Strategies . . . . .	107
4.3	Elastic Maps . . . . .	110
4.3.1	Piecewise Linear Manifolds and Data Projectors . . . . .	110
4.3.2	Iterative Data Approximation . . . . .	110
4.4	Principal Manifold as Elastic Membrane . . . . .	111
4.5	Method Implementation . . . . .	113
4.6	Examples . . . . .	113
4.6.1	Test Examples . . . . .	113
4.6.2	Modeling Molecular Surfaces . . . . .	114
4.6.3	Visualization of Microarray Data . . . . .	115
4.7	Discussion . . . . .	125
	References . . . . .	128

## 5 Topology-Preserving Mappings for Data Visualisation

<i>Marian Peña, Wesam Barbakh, Colin Fyfe</i>	132	
5.1	Introduction . . . . .	132
5.2	Clustering Techniques . . . . .	133
5.2.1	$K$ -Means . . . . .	133
5.2.2	K-Harmonic Means . . . . .	134
5.2.3	Neural Gas . . . . .	136
5.2.4	Weighted $K$ -Means . . . . .	137
5.2.5	The Inverse Weighted $K$ -Means . . . . .	138
5.3	Topology Preserving Mappings . . . . .	139
5.3.1	Generative Topographic Map . . . . .	139
5.3.2	Topographic Product of Experts ToPoE . . . . .	141
5.3.3	The Harmonic Topographic Map . . . . .	142
5.3.4	Topographic Neural Gas . . . . .	144
5.3.5	Inverse-Weighted $K$ -Means Topology-Preserving Map . . . . .	144
5.4	Experiments . . . . .	144
5.4.1	Projections in Latent Space . . . . .	145
5.4.2	Responsibilities . . . . .	145
5.4.3	U-matrix, Hit Histograms and Distance Matrix . . . . .	145
5.4.4	The Quality of The Map . . . . .	148
5.5	Conclusions . . . . .	150
	References . . . . .	151

## 6 The Iterative Extraction Approach to Clustering

<i>Boris Mirkin</i>	153	
6.1	Introduction . . . . .	153
6.2	Clustering Entity-to-feature Data . . . . .	154
6.2.1	Principal Component Analysis . . . . .	154
6.2.2	Additive Clustering Model and ITEX . . . . .	156
6.2.3	Overlapping and Fuzzy Clustering Case . . . . .	158
6.2.4	$K$ -Means and iK-Means Clustering . . . . .	159

6.3	ITEX Structuring and Clustering for Similarity Data .....	163
6.3.1	Similarity Clustering: a Review .....	163
6.3.2	The Additive Structuring Model and ITEX .....	165
6.3.3	Additive Clustering Model.....	167
6.3.4	Approximate Partitioning .....	168
6.3.5	One Cluster Clustering.....	170
6.3.6	Some Applications.....	172
	References .....	176

## **7 Representing Complex Data Using Localized Principal Components with Application to Astronomical Data**

<i>Jochen Einbeck, Ludger Evers, Coryn Bailer-Jones</i> .....	180	
7.1	Introduction .....	180
7.2	Localized Principal Component Analysis .....	183
7.2.1	Cluster-wise PCA .....	183
7.2.2	Principal Curves .....	187
7.2.3	Further Approaches .....	190
7.3	Combining Principal Curves and Regression .....	191
7.3.1	Principal Component Regression and its Shortcomings .....	191
7.3.2	The Generalization to Principal Curves .....	192
7.3.3	Using Directions Other than the Local Principal Components .....	194
7.3.4	A Simple Example .....	195
7.4	Application to the Gaia Survey Mission .....	197
7.4.1	The Astrophysical Data .....	197
7.4.2	Principal Manifold Based Approach .....	198
7.5	Conclusion .....	200
	References .....	201

## **8 Auto-Associative Models, Nonlinear Principal Component Analysis, Manifolds and Projection Pursuit**

<i>Stéphane Girard, Serge Iovleff</i> .....	205	
8.1	Introduction .....	205
8.2	Auto-Associative Models .....	206
8.2.1	Approximation by Manifolds.....	206
8.2.2	A Projection Pursuit Algorithm .....	208
8.2.3	Theoretical Results .....	209
8.3	Examples.....	210
8.3.1	Linear Auto-Associative Models and PCA .....	210
8.3.2	Additive Auto-Associative Models and Neural Networks .....	211
8.4	Implementation Aspects.....	212
8.4.1	Estimation of the Regression Functions .....	212
8.4.2	Computation of Principal Directions .....	214
8.5	Illustration on Real and Simulated Data .....	216
	References .....	220

<b>9 Beyond The Concept of Manifolds: Principal Trees, Metro Maps, and Elastic Cubic Complexes</b>	
<i>Alexander N. Gorban, Neil R. Sumner, Andrei Y. Zinov'yev</i>	223
9.1 Introduction and Overview	223
9.1.1 Elastic Principal Graphs	225
9.2 Optimization of Elastic Graphs for Data Approximation	226
9.2.1 Elastic Functional Optimization	226
9.2.2 Optimal Application of Graph Grammars	227
9.2.3 Factorization and Transformation of Factors	228
9.3 Principal Trees (Branching Principal Curves)	229
9.3.1 Simple Graph Grammar ("Add a Node", "Bisect an Edge")	229
9.3.2 Visualization of Data Using "Metro Map" Two-Dimensional Tree Layout	229
9.3.3 Example of Principal Cubic Complex: Product of Principal Trees	231
9.4 Analysis of the Universal 7-Cluster Structure of Bacterial Genomes	233
9.4.1 Brief Introduction	234
9.4.2 Visualization of the 7-Cluster Structure	236
9.5 Visualization of Microarray Data	238
9.5.1 Dataset Used	238
9.5.2 Principal Tree of Human Tissues	238
9.6 Discussion	238
References	240
<b>10 Diffusion Maps - a Probabilistic Interpretation for Spectral Embedding and Clustering Algorithms</b>	
<i>Boaz Nadler, Stephane Lafon, Ronald Coifman, Ioannis G. Kevrekidis</i>	242
10.1 Introduction	242
10.2 Diffusion Distances and Diffusion Maps	244
10.2.1 Asymptotics of the Diffusion Map	249
10.3 Spectral Embedding of Low Dimensional Manifolds	250
10.4 Spectral Clustering of a Mixture of Gaussians	254
10.5 Summary and Discussion	262
References	262
<b>11 On Bounds for Diffusion, Discrepancy and Fill Distance Metrics</b>	
<i>Steven B. Damelin</i>	265
11.1 Introduction	265
11.2 Energy, Discrepancy, Distance and Integration on Measurable Sets in Euclidean Space	266
11.3 Set Learning via Normalized Laplacian Dimension Reduction and Diffusion Distance	270

11.4 Main Result: Bounds for Discrepancy, Diffusion and Fill Distance Metrics .....	271
References .....	272
<b>12 Geometric Optimization Methods for the Analysis of Gene Expression Data</b>	
<i>Michel Journée, Andrew E. Teschendorff, Pierre-Antoine Absil, Simon Tavaré, Rodolphe Sepulchre .....</i>	274
12.1 Introduction .....	274
12.2 ICA as a Geometric Optimization Problem .....	276
12.3 Contrast Functions .....	277
12.3.1 Mutual Information [8, 10] .....	278
12.3.2 $\mathcal{F}$ -Correlation [14] .....	279
12.3.3 Non-Gaussianity [17] .....	280
12.3.4 Joint Diagonalization of Cumulant Matrices [19] .....	281
12.4 Matrix Manifolds for ICA .....	283
12.5 Optimization Algorithms .....	284
12.5.1 Line-Search Algorithms .....	284
12.5.2 FastICA .....	286
12.5.3 Jacobi Rotations .....	287
12.6 Analysis of Gene Expression Data by ICA .....	288
12.6.1 Some Issues About the Application of ICA .....	288
12.6.2 Evaluation of the Biological Relevance of the Expression Modes .....	290
12.6.3 Results Obtained on the Breast Cancer Microarray Data Set .....	292
12.7 Conclusion .....	294
References .....	294
<b>13 Dimensionality Reduction and Microarray data</b>	
<i>David A. Elizondo, Benjamin N. Passow, Ralph Birkenhead, Andreas Huemer .....</i>	297
13.1 Introduction .....	297
13.2 Background .....	299
13.2.1 Microarray Data .....	299
13.2.2 Methods for Dimension Reduction .....	300
13.2.3 Linear Separability .....	301
13.3 Comparison Procedure .....	304
13.3.1 Data Sets .....	304
13.3.2 Dimensionality Reduction .....	306
13.3.3 Perceptron Models .....	307
13.4 Results .....	307
13.5 Conclusions .....	310
References .....	311

<b>14 PCA and K-Means Decipher Genome</b>	
<i>Alexander N. Gorban, Andrei Y. Zinovyev</i>	313
14.1 Introduction	313
14.2 Required Materials	314
14.3 Genomic Sequence	315
14.3.1 Background	315
14.3.2 Sequences for the Analysis	315
14.4 Converting Text to a Numerical Table	316
14.5 Data Visualization	317
14.5.1 Visualization	317
14.5.2 Understanding Plots	317
14.6 Clustering and Visualizing Results	319
14.7 Task List and Further Information	320
14.8 Conclusion	322
References	323
<b>Index</b>	329

---

## List of Contributors

**Pierre-Antoine Absil**

Department of Mathematical  
Engineering  
Université catholique de Louvain  
Batiment Euler - Parking 13  
Av. Georges Lemaître 4  
1348 Louvain-la-Neuve  
Belgium

**Coryn Bailer-Jones**

Max-Planck-Institut für Astronomie  
Königstuhl 17  
69117 Heidelberg  
Germany  
[calj@mpia-hd.mpg.de](mailto:calj@mpia-hd.mpg.de)

**Wesam Barbakh**

Applied Computational  
Intelligence Research Unit  
The University of Paisley  
Paisley PA1 2BE  
Scotland  
United Kingdom  
[wesam.barbakh@paisley.ac.uk](mailto:wesam.barbakh@paisley.ac.uk)

**Ralph Birkenhead**

Centre for Computational  
Intelligence  
School of Computing  
Faculty of Computing Sciences  
and Engineering

De Montfort University

The Gateway  
Leicester LE1 9BH  
United Kingdom  
[rab@dmu.ac.uk](mailto:rab@dmu.ac.uk)

**Ronald Coifman**

Department of Mathematics  
Yale University  
New Haven, CT 06520-8283  
USA  
[coifman@math.yale.edu](mailto:coifman@math.yale.edu)

**Steven B. Damelin**

Department of Mathematical  
Sciences  
Georgia Southern University  
PO Box 8093  
Statesboro, GA 30460  
USA  
[damelin@georgiasouthern.edu](mailto:damelin@georgiasouthern.edu)

**Jochen Einbeck**

Department of Mathematical  
Sciences  
Durham University  
Science Laboratories  
South Road  
Durham DH1 3LE  
United Kingdom  
[jochen.einbeck@durham.ac.uk](mailto:jochen.einbeck@durham.ac.uk)

<b>David A. Elizondo</b> Centre for Computational Intelligence School of Computing Faculty of Computing Sciences and Engineering De Montfort University The Gateway Leicester LE1 9BH United Kingdom <a href="mailto:elizondo@dmu.ac.uk">elizondo@dmu.ac.uk</a>	av. de l'Europe Montbonnot 38334 Saint-Ismier cedex France <a href="mailto:Stephane.Girard@inrialpes.fr">Stephane.Girard@inrialpes.fr</a>
<b>Ludger Evers</b> Department of Mathematics University of Bristol University Walk Bristol BS8 1TW United Kingdom <a href="mailto:l.evers@bris.ac.uk">l.evers@bris.ac.uk</a>	<b>Alexander N. Gorban</b> Department of Mathematics University of Leicester University Road Leicester LE1 7RH United Kingdom, and Institute of Computational Modeling Russian Academy of Sciences Siberian Branch Krasnoyarsk 660036 Russia <a href="mailto:ag153@le.ac.uk">ag153@le.ac.uk</a>
<b>Martin Fraunholz</b> Competence Centre for Functional Genomics Institute for Microbiology Ernst-Moritz-Arndt-University Greifswald F.-L.-Jahn-Str. 15 17487 Greifswald Germany <a href="mailto:Martin.Fraunholz@uni-greifswald.de">Martin.Fraunholz@uni-greifswald.de</a>	<b>Andreas Huemer</b> Centre for Computational Intelligence School of Computing Faculty of Computing Sciences and Engineering De Montfort University The Gateway Leicester LE1 9BH United Kingdom <a href="mailto:ahuemer@dmu.ac.uk">ahuemer@dmu.ac.uk</a>
<b>Colin Fyfe</b> Applied Computational Intelligence Research Unit The University of Paisley Paisley PA1 2BE Scotland United Kingdom <a href="mailto:colin.fyfe@paisley.ac.uk">colin.fyfe@paisley.ac.uk</a>	<b>Serge Iovleff</b> Laboratoire Paul Painlevé 59655 Villeneuve d'Ascq Cedex France <a href="mailto:serge.iovleff@univ-lille1.fr">serge.iovleff@univ-lille1.fr</a>
<b>Stéphane Girard</b> INRIA Rhône-Alpes Projet Mistis Inovallée 655	<b>Michel Journée</b> Department of Electrical Engineering and Computer Science University of Liège B-4000 Liège Sart-Tilman Belgium

**Ioannis G. Kevrekidis**  
Department of Chemical Engineering  
and Program in Applied  
and Computational Mathematics  
Princeton University  
Princeton, NJ 08544  
USA  
[yannis@princeton.edu](mailto:yannis@princeton.edu)

**Uwe Kruger**  
School of Electronics  
Electrical Engineering and  
Computer Science  
Queen's University Belfast  
Belfast BT9 5AH  
United Kingdom  
[uwe.kruger@ee.qub.ac.uk](mailto:uwe.kruger@ee.qub.ac.uk)

**Stephane Lafon**  
Department of Mathematics  
Yale University  
New Haven, CT 06520-8283  
USA  
[stephane.lafon@yale.edu](mailto:stephane.lafon@yale.edu)  
and Google Inc.

**Boris Mirkin**  
School of Computer Science  
and Information Systems  
Birkbeck College  
University of London  
Malet Street  
London WC1E 7HX  
United Kingdom  
[mirkin@dcs.bbk.ac.uk](mailto:mirkin@dcs.bbk.ac.uk)

**Boaz Nadler**  
Department of Computer Science  
and Applied Mathematics  
Weizmann Institute of Science  
Rehovot 76100  
Israel  
[boaz.nadler@weizmann.ac.il](mailto:boaz.nadler@weizmann.ac.il)

**Benjamin N. Passow**  
Centre for Computational  
Intelligence  
School of Computing  
Faculty of Computing Sciences  
and Engineering  
De Montfort University  
The Gateway  
Leicester LE1 9BH  
United Kingdom  
[passow@dmu.ac.uk](mailto:passow@dmu.ac.uk)

**Marian Peña**  
Applied Computational  
Intelligence Research Unit  
The University of Paisley  
Paisley PA1 2BE  
Scotland  
United Kingdom  
[marian.pena@paisley.ac.uk](mailto:marian.pena@paisley.ac.uk)

**Matthias Scholz**  
Competence Centre  
for Functional Genomics  
Institute for Microbiology  
Ernst-Moritz-Arndt-University  
Greifswald  
F.-L.-Jahn-Str. 15  
17487 Greifswald  
Germany  
[Matthias.Scholz@uni-greifswald.de](mailto:Matthias.Scholz@uni-greifswald.de)

**Joachim Selbig**  
Institute for Biochemistry  
and Biology  
University of Potsdam  
c/o Max Planck Institute  
for Molecular Plant Physiology  
Am Mühlenberg 1  
14424 Potsdam  
Germany  
[Selbig@mpimp-golm.mpg.de](mailto:Selbig@mpimp-golm.mpg.de)

**Rodolphe Sepulchre**

Department of Electrical Engineering  
and Computer Science  
University of Liège  
B-4000 Liège Sart-Tilman  
Belgium  
[r.sepulchre@ulg.ac.be](mailto:r.sepulchre@ulg.ac.be)

**Neil R. Sumner**

Department of Mathematics  
University of Leicester  
University Road  
Leicester LE1 7RH  
United Kingdom  
[nrs7@le.ac.uk](mailto:nrs7@le.ac.uk)

**Simon Tavaré**

Breast Cancer  
Functional Genomics Program  
Cancer Research UK  
Cambridge Research Institute  
Department of Oncology  
University of Cambridge  
Robinson Way  
Cambridge CB2 0RE  
United Kingdom

**Andrew E. Teschendorff**

Breast Cancer  
Functional Genomics Program  
Cancer Research UK  
Cambridge Research Institute  
Department of Oncology  
University of Cambridge  
Robinson Way  
Cambridge CB2 0RE  
United Kingdom

**Lei Xie**

National Key Laboratory  
of Industrial Control Technology  
Zhejiang University  
Hangzhou 310027  
P.R. China  
[leix@iipc.zju.edu.cn](mailto:leix@iipc.zju.edu.cn)

**Hujun Yin**

School of Electrical  
and Electronic Engineering  
The University of Manchester  
Manchester M60 1QD  
United Kingdom  
[hujun.yin@manchester.ac.uk](mailto:hujun.yin@manchester.ac.uk)

**Junping Zhang**

Department of Computer Science  
and Engineering  
Fudan University  
Shanghai 200433  
P.R. China  
[jpzhang@fudan.edu.cn](mailto:jpzhang@fudan.edu.cn)

**Andrei Y. Zinovyev**

Institut Curie  
26 rue d'Ulm  
Paris 75248  
France,  
and Institute of Computational  
Modeling  
Russian Academy of Sciences  
Siberian Branch  
Krasnoyarsk 660036  
Russia  
[andrei.zinovyev@curie.fr](mailto:andrei.zinovyev@curie.fr)

