

Surveys of Modern Mathematics
Volume III

Application of Elementary Differential Geometry to Influence Analysis

Yat-Sun Poon
Wai-Yin Poon

 International Press
www.intlpress.com

 高等教育出版社
HIGHER EDUCATION PRESS

Surveys of Modern Mathematics, Volume III
Application of Elementary Differential Geometry to Influence Analysis

Yat-Sun Poon
Department of Mathematics, University of California at Riverside
Wai-Yin Poon
Department of Statistics, The Chinese University of Hong Kong

2010 Mathematics Subject Classification. 53-01, 53A07, 53A30, 53A45, 53B20, 62J05, 97K70.

Copyright © 2010, 2012 by International Press, Somerville, Massachusetts, U.S.A., and by Higher Education Press, Beijing, China.

This work is published and sold in China exclusively by Higher Education Press of China.

All rights reserved. Individual readers of this publication, and non-profit libraries acting for them, are permitted to make fair use of the material, such as to copy a chapter for use in teaching or research. Permission is granted to quote brief passages from this publication in reviews, provided the customary acknowledgement of the source is given. Republication, systematic copying, or mass reproduction of any material in this publication is permitted only under license from International Press. Excluded from these provisions is material in articles to which the author holds the copyright. (If the author holds copyright, notice of this will be given with the article.) In such cases, requests for permission to use or reprint should be addressed directly to the author.

ISBN: 978-1-57146-252-7

Printed in the United States of America.

16 15 14 13 12 1 2 3 4 5 6 7 8 9

SURVEYS OF MODERN MATHEMATICS

Series Editors

Shing-Tung Yau
Harvard University
Cambridge, Massachusetts
U.S.A.

Lizhen Ji
University of Michigan, Ann Arbor
U.S.A.

Yat-Sun Poon
University of California at Riverside
U.S.A.

Jie Xiao
Tsinghua University
Beijing, China

Jean-Pierre Demailly
Institut Fourier
Laboratoire de Mathématiques
Saint-Martin d'Hères, France

Eduard J.N. Looijenga
Universiteit Utrecht
The Netherlands

Neil Trudinger
Mathematical Sciences Institute
Australian National University
Canberra, Australia

*Dedicated to our father Yan Ding Poon
for his 88th birthday*

Contents

Preface	xi
 Part I Geometry	
1 Preliminaries	3
1.1 Linear algebra	3
1.1.1 Vectors and matrices	3
1.1.2 Symmetric bilinear forms	5
1.1.3 Vector subspaces	6
1.1.4 Linear maps from \mathbb{R}^n to \mathbb{R}^n	7
1.1.5 A convention	9
1.2 Vector calculus	9
1.2.1 Vector-valued functions and differentials	9
1.2.2 Taylor expansion and extrema	11
1.2.3 Extrema and Lagrange multiplier theorem	12
 2 Euclidean Geometry	15
2.1 Orthogonal transformations	15
2.2 Rigid motions	16
2.3 Translation of vector subspaces	18
2.4 Conformal transformations	20
2.5 Orthonormal basis	20
2.6 Orthogonal projections	23
2.7 Areas and volumes	25
 3 Geometry of Graphs	29
3.1 Graphs in Euclidean spaces	29
3.2 Normal sections	31
3.3 Cross sections in high dimension	33

3.4	First fundamental forms	33
4	Curvatures	35
4.1	Normal curvatures	35
4.1.1	Definition	35
4.1.2	Principal curvatures and principal directions	37
4.2	Sectional curvatures	40
5	Transformations and Invariance	43
5.1	Change of coordinates	43
5.2	Non-linear conformal transformations	44
5.3	Invariant curvatures	46
 Part II Statistics		
6	Discrete Random Variables and Related Concepts	51
6.1	Preliminaries	51
6.2	Discrete random variables	52
6.2.1	Discrete random variables and probability function	52
6.2.2	Relative frequency histogram	55
6.2.3	Cumulative distribution function	55
6.3	Population parameters and sample statistics	56
6.3.1	Population mean and expected value	56
6.3.2	Sample statistic	57
6.3.3	Sample mean	57
6.3.4	Sample and population variances	58
6.4	Mathematical expectations	60
6.5	Maximum likelihood estimation	61
6.6	Maximum likelihood estimation of the probability of a Bernoulli experiment	62
7	Continuous Random Variables and Related Concepts	65
7.1	Continuous random variables	65
7.2	Mathematical expectation for continuous random variables	66
7.3	Mean and variance and their sample estimates	66
7.4	Basic properties of expectations	67
7.5	Normal distribution	68
7.6	Maximum likelihood estimation for continuous variables	72

7.7	Maximum likelihood estimation for the parameters of normal distribution	73
7.8	Sampling distribution	74
8	Bivariate and Multivariate Distribution	77
8.1	Bivariate distribution for discrete random variables	77
8.1.1	Joint probability function	77
8.1.2	Marginal probability function	78
8.1.3	Conditional probability function	79
8.2	Bivariate distribution for continuous random variables	80
8.3	Mathematical expectations	80
8.3.1	Mathematical expectations for the functions of two random variables	80
8.4	Covariance and correlation	82
8.4.1	Sample covariance and correlation	82
8.4.2	Population covariance and correlation	84
8.4.3	Conditional expectations	85
8.5	Bivariate normal distribution	87
8.6	Independence	88
8.7	Multivariate distribution	89
9	Simple Linear Regression	93
9.1	The model	93
9.2	The least squares estimation	95
9.3	The maximum likelihood estimation of regression parameters	98
9.4	Residuals	99
9.5	Coefficient of determination	101
9.6	Weighted least squares estimates	103
10	Topics on Linear Regression Analysis	105
10.1	Multiple regression model	105
10.2	Estimation and interpretation	106
10.3	Influential observations and outliers	110
10.4	Leverage	111
10.5	Cook's distance	113
10.6	Deletion influence, joint influence and masking effect	114
10.7	Derivation of Cook's distances	116
10.7.1	Weighted least squares and Cook's distance	116
10.7.2	Cook's distance – deleting one data point	118

Part III Local Influence Analysis

11 Basic Concepts	123
11.1 Introduction	123
11.2 Perturbation	125
11.3 Likelihood displacement and influence graph	126
12 Measuring Local Influence	129
12.1 Individual influence	130
12.2 Derivation of normal curvature	131
12.3 Case-weight perturbation — an example	133
12.4 Roles of sectional curvature	135
12.5 Joint influence	138
13 Relations Among Various Measures	141
13.1 A bound on influence measures	141
13.2 Individual and overall joint influence	143
13.3 Individual and joint influence measures	146
13.4 Competing eigenvalues	147
13.5 Conclusions	150
14 Conformal Modifications	153
14.1 Modification and invariance	153
14.2 Invariant measures	154
14.3 Benchmarks	155
14.4 Individual's contribution to joint influence — re-visited	157
Appendix A Rank of Hat Matrix	161
Appendix B Ricci Curvature	163
Appendix C Cook's Distance — Deleting Two Data Points	165
Bibliography	167
Index	171

Preface

The two authors of this book have different backgrounds within the mathematical sciences. After fifteen years of collaboration, they find it beneficial to introduce their research areas jointly to students in the early stages of their research careers. Mathematics students will learn that mathematical concepts have an immediate impact on real life situations. Statistics students will appreciate that abstract mathematical concepts are accessible, relevant and valuable. In fact, an early draft of this book was used by one of the authors as lecture notes to run one year-long undergraduate research seminar for a mixture of mathematics and statistics undergraduate students in their senior years.

We hope that this book will encourage young scientists to develop an appreciation of inter-disciplinary research at a time when their research career is just beginning to emerge.

Differential geometry is a broad mathematical subject. Its global aspects are often presented only to mathematics majors or graduate students. However, its local aspects are the foundation of global issues and can be made accessible to all students with rigorous, multi-variable calculus training. To study local geometry, one studies graphs. It is a generalization of the freshman discussion on the concavity of a real-valued function of a single variable. By studying the geometry of graphs, one can proceed to learn global differential geometry as a mathematician.

One may also begin to apply the geometry of a graph to analyze functions arising from concrete problems, such as through statistics. Many statistical analyses involve a hypothesized model. Once a model is specified, the data collected are used to estimate the parameters that characterize the model. Further inference is affected by the hypothesized model and the data collected. Therefore, one must assess the influence based on the perturbation of various aspects of the model inputs. Perturbations can be represented by a set of perturbation parameters, and the function of these perturbation parameters becomes a mathematical object of interest. Therefore, one can apply geometric concepts to study this function and hence, deduce information on the perturbation.

Working through this process prompts theoretical, practical and technical issues. In particular, we must develop measures for the influence of individual

perturbation parameters. We must also develop measures for the joint influence of any two, and then any groups of perturbation parameters. From single to multiple-parameter measures, the process is not a straightforward generalization of lower dimensional geometric problems to higher dimensional problems, because direct geometric generalization may not produce appropriate measures capable of serving the purposes of a statistical analysis. Instead, it is necessary to develop a set of meaningful measures that can constitute a well-structured system that enables further pursuit of relationships among the measures as well as the development of practical tools that facilitate interpretation and data analysis. Typical examples of such practical tools include the establishment of benchmarks for judging and interpreting measures and the search for alternative measures that reduce the computational burden. All such issues must be addressed using geometric techniques in the light of statistical considerations.

In Part I of this book, to fix conventions we recall basics of linear algebra, multi-variable calculus and Euclidean geometry in Chapters 1 and 2. In Chapter 3, we introduce the concept of normal sections, first fundamental forms and second fundamental forms. In Chapter 4, we introduce normal curvature and sectional curvatures. In Chapter 5, we study conformal transformations. This finishes our mathematical preparation.

In Part II, we first review some elementary statistics topics. In Chapters 6 and 7, we introduce basic concepts in relation to univariate distribution for discrete and continuous random variables, including the maximum likelihood estimation method. After generalizing these concepts to the bivariate and multivariate distributions in Chapter 8, we introduce simple linear regression in Chapter 9. Linear regression is the most popular statistical model, and is used as the key example in Part III of this book. To prepare for the illustration, some well-known topics in linear regression are discussed in Chapter 10.

In Part III, we apply the geometric concepts developed in Part I to the statistical issues and models articulated in Part II. The goal is to develop various measures generated by a local perturbation to assess the influence of the perturbation of model inputs. In Chapter 11, we develop the concept of likelihood displacement function. In Chapter 12, we apply the tools developed in Chapters 3 and 4 to analyze the likelihood displacement function. In the process of this development, different or mingled perspectives must be clarified. We use the regression model with some of the most well-known perspectives described in Chapter 10 for illustration. The relations among various measures that are generated in the process must also be analyzed. This is done in Chapter 13. Finally, in Chapter 14, we analyze the modification of perturbations using Chapter 5 as the theoretical foundation.

As we develop our presentation, we often encounter tedious but necessary computations. For the sake of completeness, we have placed computations in

the Appendices.

Graduate students in mathematics may choose to begin reading this book starting with Part II. Likewise, graduate students in statistics may skip Part II. However, undergraduate students from both disciplines will benefit from reading this book in its entirety.

Yat-Sun Poon
Wai-Yin Poon
May, 2012