Spatial Tessellations: Concepts and Applications of Voronoi Diagrams

Second Edition

ATSUYUKI OKABE
University of Tokyo, Japan

BARRY BOOTS
Wilfrid Laurier University, Ontario, Canada

KOKICHI SUGIHARA
University of Tokyo, Japan

SUNG NOK CHIU
Hong Kong Baptist University, China

With a Foreword by
D.G. KENDALL

JOHN WILEY & SONS, LTD
Chichester • New York • Weinheim • Brisbane • Singapore • Toronto

This Page intentionally left blank

Spatial Tessellations

Second Edition

WILEY SERIES IN PROBABILITY AND STATISTICS

Established by WALTER A. SHEWHART and SAMUEL S. WILKS

Spatial Tessellations: Concepts and Applications of Voronoi Diagrams

Second Edition

ATSUYUKI OKABE
University of Tokyo, Japan

BARRY BOOTS
Wilfrid Laurier University, Ontario, Canada

KOKICHI SUGIHARA
University of Tokyo, Japan

SUNG NOK CHIU
Hong Kong Baptist University, China

With a Foreword by
D.G. KENDALL

JOHN WILEY & SONS, LTD
Chichester • New York • Weinheim • Brisbane • Singapore • Toronto

Baffins Lane, Chichester,
West Sussex, PO19 1UD, England

National 01243 779777
International (+44) 1243 779777

e-mail (for orders and customer enquiries): cs-books@wiley.co.uk

Visit our Home Page on http://www.wiley.co.uk or http://www.wiley.com

Other Wiley Editorial Offices

John Wiley & Sons Inc., 605 Third Avenue,
New York, NY 10158-0012, USA

Wiley-VCH Verlag GmbH, Pappelallee 3,
D-69469 Weinheim Germany

Jacaranda Wiley Ltd, 33 Park Road, Milton,
Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01,
Jin Xing Distirpark, Singapore 129809

John Wiley & Sons (Canada) Ltd, 22 Worcester Road,
Rexdale, Ontario, M9W 1L1, Canada

Library of Congress Cataloging-in-Publication Data

Okabe, Atsuyuki, 1945–
Spatial tessellations: concepts and applications of Voronoi diagrams / Atsuyuki Okabe ... [et al.] ; with a foreword by D.G. Kendall — 2nd ed.
p. cm. — (Wiley series in probability and statistics)
Includes bibliographical references and index.
ISBN 0-471-98635-6 (alk. paper)
1. Voronoi polygons. 2. Spatial analysis (Statistics). 3. Geometry—Data processing.
I. Title. II. Series.
QA278.2.036 1999
519.5'36—dc21 99–13149
CIP

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN 0–471–98635–6

Typeset in 10/12pt Times by Florence Production Ltd, Stoodleigh, Devon

Contents

Foreword to the First Edition xi
Preface to the Second Edition xiii
Acknowledgements (First Edition) xv
Acknowledgements (Second Edition) xvi

Chapter 1 Introduction 1
1.1 Outline 3
1.2 History of the concept of the Voronoi diagram 6
1.3 Mathematical preliminaries 12
1.3.1 Vector geometry 12
1.3.2 Graphs 24
1.3.3 Spatial stochastic point processes 31
1.3.4 Efficiency of computation 41

Chapter 2 Definitions and Basic Properties of Voronoi Diagrams 43
2.1 Definitions of the ordinary Voronoi diagram 43
2.2 Definitions of the Delaunay tessellation (triangulation) 52
2.3 Basic properties of the Voronoi diagram 57
2.4 Basic properties of the Delaunay triangulation 70
2.5 Graphs related to the Delaunay triangulation 97
2.6 Recognition of Voronoi diagrams 103
2.6.1 The geometric approach 104
2.6.2 The combinatorial approach 106

Chapter 3 Generalizations of the Voronoi diagram 113
3.1 Weighted Voronoi diagrams 119
3.1.1 The multiplicatively weighted Voronoi diagram 120
3.1.2 The additively weighted Voronoi diagram 123
3.1.3 The compoundly weighted Voronoi diagram 127
3.1.4 The power diagram 128
3.1.5 The sectional Voronoi diagram 131
3.1.6 Applications 133

3.2 Higher-order Voronoi diagrams 134
3.2.1 The order-k Voronoi diagram 135
3.2.2 The ordered order-k Voronoi diagram 144
3.2.3 Applications 150
3.3 The farthest-point Voronoi diagram and the kth nearest-point Voronoi diagram 151
3.3.1 The farthest-point Voronoi diagram 151
3.3.2 The kth nearest-point Voronoi diagram 155
3.3.3 Applications 157
3.4 Voronoi diagrams with obstacles 158
3.4.1 The shortest-path Voronoi diagram 158
3.4.2 The visibility shortest-path Voronoi diagram 163
3.4.3 The constrained Delaunay triangulation 165
3.4.4 SP- and VSP-Voronoi diagrams in a simple polygon 168
3.4.5 Applications 168
3.5 Voronoi diagrams for lines 169
3.5.1 Voronoi diagrams for a set of points and straight line segments 171
3.5.2 Voronoi diagrams for a set of points, straight line segments and circular arcs 176
3.5.3 Voronoi diagrams for a set of circles 178
3.5.4 Medial axis 181
3.5.5 Applications 184
3.6 Voronoi diagrams for areas 186
3.6.1 The area Voronoi diagram 186
3.6.2 Applications 188
3.7 Voronoi diagrams with V-distances 189
3.7.1 Voronoi diagrams with the Minkowski metric L_p 189
3.7.2 Voronoi diagrams with the convex distance 194
3.7.3 Voronoi diagrams with the Karlsruhe metric 201
3.7.4 Voronoi diagrams with the Hausdorff distance 202
3.7.5 Voronoi diagram with the boat-on-a-river distance 204
3.7.6 Voronoi diagrams on a sphere 206
3.7.7 Voronoi diagrams on a cylinder 209
3.7.8 Voronoi diagrams on a cone 210
3.7.9 Voronoi diagrams on a polyhedral surface 211
3.7.10 Miscellany 212
3.7.11 Applications 215
3.8 Network Voronoi diagrams 218
3.8.1 The network Voronoi node diagram 219
3.8.2 The network Voronoi link diagram 220
3.8.3 The network Voronoi area diagram 221
3.8.4 Applications 224
3.9 Voronoi diagrams for moving points 224
3.9.1 Dynamic Voronoi diagrams 224
3.9.2 Applications 227

Chapter 4 Algorithms for Computing Voronoi Diagrams 229
4.1 Computational preliminaries 229
4.2 Data structure for representing a Voronoi diagram 235
4.3 The incremental method 242
4.4 The divide-and-conquer method 251
4.5 The plane sweep method 257
4.6 Practical techniques for implementing the algorithms 264
4.6.1 Inconsistency caused by numerical errors 264
4.6.2 Construction of an error-free world 265
4.6.3 Topology-oriented approach 269
4.7 Algorithms for higher-dimensional Voronoi diagrams 275
4.8 Algorithms for generalized Voronoi diagrams 280
4.9 Approximation algorithms 287

Chapter 5 Poisson Voronoi Diagrams 291
5.1 Properties of infinite Voronoi diagrams 295
5.2 Properties of Poisson Voronoi diagrams 299
5.3 Uses of Poisson Voronoi diagrams 300
5.4 Simulating Poisson Voronoi and Delaunay cells 306
5.5 Properties of Poisson Voronoi cells 311
5.5.1 Moments of the characteristics of Poisson Voronoi cells 311
5.5.2 Conditional moments of the characteristics of Poisson Voronoi cells 315
5.5.3 Conditional moments of the characteristics of the neighbouring cells of a Poisson Voronoi cell 324
5.5.4 Distributional properties 331
5.6 Stochastic processes induced by Poisson Voronoi diagrams 350
5.6.1 Point processes of centroids of faces 350
5.6.2 Voronoi growth models 357
5.6.3 The Stienen model 360
5.6.4 Percolation on Poisson Voronoi diagrams and Poisson Delaunay tessellations 361
5.7 Sectional Voronoi diagrams 363
5.8 Additively weighted Poisson Voronoi diagrams: the Johnson–Mehl model 374
5.9 Higher order Poisson Voronoi diagrams 385
5.10 Poisson Voronoi diagrams on the surface of a sphere 389
5.11 Properties of Poisson Delaunay cells 389
5.12 Other random Voronoi diagrams 404

Chapter 6 Spatial Interpolation **411**
6.1 Polygonal methods 416
6.1.1 Nearest neighbour interpolation 417
6.1.2 Natural neighbour interpolation 418
6.2 Triangular methods 427
6.3 Modifying Delaunay triangulations 434
6.4 Approximating surfaces 437
6.5 Delaunay meshes for finite element methods 439
6.5.1 Two-dimensional Delaunay meshes 440
6.5.2 Three-dimensional Delaunay meshes 442
6.6 Ordering multivariate data 446

Chapter 7 Models of Spatial Processes **453**
7.1 Assignment models 454
7.2 Growth models 476
7.3 Spatial-temporal processes 482
7.3.1 Spatial competition models: the Hotelling process 482
7.3.2 Adjustment models 489
7.4 Two-species models 491

Chapter 8 Point Pattern Analysis **495**
8.1 Polygon-based methods 498
8.1.1 Direct approach 498
8.1.2 Indirect approaches 502
8.2 Triangle-based methods 506
8.3 Nearest neighbour distance methods 512
8.3.1 Nearest neighbour distance method for point-like objects 514
8.3.2 Nearest neighbour distance method for line-like objects 517
8.3.3 Nearest neighbour distance method for area-like objects 520
8.3.4 Multi nearest neighbour distance method 521
8.4 The shape of a point pattern 521
8.4.1 Internal shape 521
8.4.2 External shape 523
8.5 Spatial intensity 525
8.6 Segmenting point patterns 527
8.7 Modelling point processes 529

Chapter 9 Locational Optimization Through Voronoi Diagrams 531
9.1 Preliminaries 532
9.1.1 The non-linear, non-convex programming problem 532
9.1.2 The descent method 534
9.1.3 The penalty function method 538
9.2 Locational optimization of points 541
9.2.1 Locational optimization of point-like facilities used by independent users 542
9.2.2 Locational optimization of points in a three-dimensional space 548
9.2.3 Locational optimization of point-like facilities used by groups 549
9.2.4 Locational optimization of a hierarchical facility 551
9.2.5 Locational optimization of observation points for estimating the total quantity of a spatial variable continuously distributed over a plane 555
9.2.6 Locational optimization of service points of a mobile facility 558
9.2.7 Locational optimization of terminal points through which users go to the central point 559
9.2.8 Locational optimization of points on a continuous network 563
9.3 Locational optimization of lines 564
9.3.1 Locational optimization of a service route 564
9.3.2 Locational optimization of a network 567
9.3.3 Euclidean Steiner minimum tree 570
9.4 Locational optimization over time 575
9.4.1 Multi-stage locational optimization 575
9.4.2 Periodic locational optimization 578
9.5 Voronoi fitting and its application to locational optimization problems 581
9.5.1 Method of fitting a Voronoi diagram to a polygonal tessellation 581
9.5.2 Locational optimization for minimizing restricted areas 584

References 585
Index 657

Foreword to the First Edition

I was delighted to be asked to write a preface to this beautiful and outstandingly original book. It is the unique treatise on its subject, it fills a serious gap in the literature and it covers the theory and the huge range of applications in a masterly way.

The authors are right to distinguish Voronoi *diagrams* and Delone *tessellations*. The Delone construction decomposes a Euclidean space of m dimensions, containing a given set of points, into non-overlapping space-filling *simplexes* (not, of course, all of the same shape and size), so that it tessellates the space using tiles that are identical with one another up to linear transformations. The Voronoi construction also splits up the space into polyhedral cells, but now they are much less uniform in character – the number of faces will vary from one cell to another, so that it would be wrong to call the result a tessellation.

These mutually dual procedures give fascinating but different insights into the structure of a set of points in m dimensions, and they have found numerous applications. At the time of writing there is a new application on the largest of all possible scales which throws light on the structure of the Universe as we see it. This will be seen as a particularly interesting development when one recalls that most of the earlier applications (for example, to the study of the structure of metallic composites, and other such aggregates) were on the microscopic scale. The reader of this book is strongly urged to look at a review paper just published by Icke and van de Weygaert (*Quarterly Journal, Royal Astronomical Society*, **32**, 85–112). There it is shown that the Voronoi construction not only gives insight into the distribution of galaxies, but also permits a new approach to the dynamics that mould the shape of the universe we live in.

My own contributions have been in the Delone tradition, and are concerned (for example) with the way in which high-dimensional Delone simplexes pack together around a common vertex. Thus in 15 dimensions the number of such locally associated simplexes turns out to be of the order of 44 million million. This implies a related statement about the Voronoi polyhedra, and there tells us something about the number of faces of an individual cell. It seems likely that the huge number of Delone simplexes in such a local 'fan' can be roughly partitioned into a moderate number of

This Page intentionally left blank

'chunky' simplexes (substantial faces in the Voronoi case), and a vast number of 'needle-like' ones (tiny faces), but we have no precise information on this matter at the moment.

It is a great pleasure to welcome this book to the Wiley series.

David Kendall

Preface to the Second Edition

The First Edition of this book was published in 1992. In 1995, it was reprinted. At that time, we suggested to the publishers that, given the continuing interest in Voronoi diagrams in so many quarters, rather than consider further reprints they allow us to prepare a new, revised, Second Edition. We were pleased to receive a positive reply and so this volume was born.

While this edition maintains the overall structure of the first, there are substantial changes in the content. In particular, on-going growth in research relating to Voronoi diagrams is reflected in the addition of much new material to this volume. Although such additions occur throughout the book, they are most visible in new generalizations of the ordinary Voronoi diagram, new and revised results relating the Poisson Voronoi diagram, and new applications of all forms of Voronoi diagrams. The growth in Voronoi diagram research is also manifest in several other ways. One is the presence of a fourth author, Sung Nok Chiu, without whose contribution the original three authors would probably still be labouring over the revisions. Another is the increase in the number of references from 677 in the First Edition to 1680, 523 of which have appeared since the First Edition was published.

In order to accommodate the new developments we have omitted some material from the First Edition. This is most obvious in the mathematical preliminaries in Chapter 1 where we have omitted the sections relating to matrices, derivatives, integration and probability.

This book is accompanied by a World Wide Web site (`http://okabe.t.u-tokyo.ac.jp/okabelab/Voronoi/index.html`) which provides additional material such as pointers to available Voronoi diagrams and related geometric software and other Web sites featuring Voronoi diagrams. Our WWW page can also be used to notify us of any errors. Although the text has been proofread many times by ourselves and others, it is inevitable that some logical and typographical errors will not have been detected. We will correct any errors we become aware of and provide an Errata list on our WWW page.

December 1998 The authors

Acknowledgments (First Edition)

So many people helped in so many ways during the preparation of this book that it is only possible to acknowledge a few of them individually. First, we are deeply grateful to D.G. Kendall, who read through the draft and encouraged its publication; and to Y. Asami, C.M. Hoffmann, M. Iri, K. Murota and A. Suzuki, who suggested or commented on parts of the draft. Our special thanks also go to D.A. Aboav, F. Aurenhammer, H. Edelsbrunner, S. Egginton, J.D. Embury, M.F. Goodchild, M. Hori, H.-C. Imhof, G. Le Caër, U. Lorz, J. Mecke, R.E. Miles, J. Møller, L. Muche, Y. Ohsawa, N. Rivier, Y.M. Seoung, D. Stoyan, T. Suzuki, M. Tanemura, G. Toussaint, D.S. Wilkinson, H. Yomono and L. Zaninetti, who provided material. We must also express our debt to A. Dawkins, S. Henry H. Honkers, J. Horton, T. Kaneko, O. Kurita, R. Metcalfe, P. Schaus, M. Stone and T. Yoshikawa, among others, who assisted in production. For the help that they have given us, we are indebted to the staff of the publisher, in particular, C. Farmer. S. Gale, J. Narain and H. Ramsey. We should also acknowledge the award of a Book Preparation Grant from Wilfrid Laurier University which helped meet costs incurred during the preparation of the manuscript. Finally we are grateful for academic e-mail networks which made us feel as if we had been working in the same office.

Acknowledgments (Second Edition)

As with the First Edition, so many people helped us in different ways in the preparation of this edition that it is impossible to acknowledge all of them individually. However, we are especially indebted to two individuals who exposed us to significant applications which were either overlooked or received only passing reference in the first edition. Initially by means of a footnote in Oden *et al.* (1993) and later by direct communication, H. Goebl revealed the use of Voronoi diagrams in linguistics, while E. Agrell, by way of his book (Agrell, 1997), showed us how much we had missed on the use of Voronoi diagrams in coding. For their comments and suggestions on the First Edition or drafts of this edition we would like to thank H. Imai, K. Imai, R. Klein, R.C. Lindenbergh, U. Lorz, M. McAllister, L. Muche, T. Roos, M. Schlather, N. Shiode, and C.A. Wang. We would also like to thank those who generously shared their unpublished research or other material with us, C. Gold, U. Lorz, K. McLeod, C. Moukarzel, L. Muche, K. Ohnishi, M. Schlather, D. Stoyan, and D. Watson. Thanks are also due to those who assisted in the production of this edition, especially P. Churcher, S. Horiike, J. Horton, C. Kanasaki, T. Kuroiwa, M. Lefebvre, H. Rayner, P. Schaus, and C. Yoshimoto. Finally, it is again a pleasure to acknowledge the help and guidance we have received from the staff of the publisher, in particular, S. Clutton, S. Corney, and H. Ramsey, who dealt with the idiosyncrasies of four authors scattered around the globe with both patience and good humour.

Spatial Tessellations: Concepts and Applications of Voronoi Diagrams
Edition by ATSUYUKI OKABE & BARRY BOOTS and SUNG NOK CHIU

CHAPTER 1

Introduction

On initial consideration the following problems, which concern a variety of phenomena at disparate scales, would appear to have little in common:

- an astronomer studying the structure of the Universe;
- an archaeologist attempting to identify the parts of a region under the influence of different neolithic clans;
- a meteorologist estimating precipitation at a gauge which has failed to operate;
- an urban planner locating public schools in a city;
- a physicist studying the behaviour of liquid argon;
- a physiologist examining capillary supply to muscle tissue.

However, these problems (all of which, together with many others, are dealt with below) can be resolved by approaches developed from a single concept which forms the subject of this book.

The concept is a simple but intuitively appealing one. Given a finite set of distinct, isolated points in a continuous space, we associate all locations in that space with the closest member of the point set. The result is a partitioning of the space into a set of regions. Figure 1.0.1 shows a simple instance for two-dimensional Euclidean space. Given its widespread occurrence it is not surprising that this concept has been 'discovered' many times in many different places and as a result both the diagram and its constituent regions are known by a plethora of aliases. One immediate problem we faced in writing this book was the choice of an appropriate name which would be recognized by those interested in the concept. We selected the terms Voronoi diagram and Voronoi region, which seem to be the most extensively used, although equivalent terms are identified in Section 1.2.

A second diagram can be constructed from the Voronoi diagram in m-dimensional space by joining those points whose regions share an $(m-1)$-dimensional face (see Figure 1.0.1). We refer to this dual diagram as the Delaunay tessellation, although it too has a number of aliases (see Section 1.2). The Delaunay tessellation may also be constructed directly from the point set by taking each $(m+1)$-ad of points and examining its circumsphere. If the interior of this docs not contain a point of the set, we construct the simplex determined by the $(m+1)$ points, but if it is not empty we do nothing.

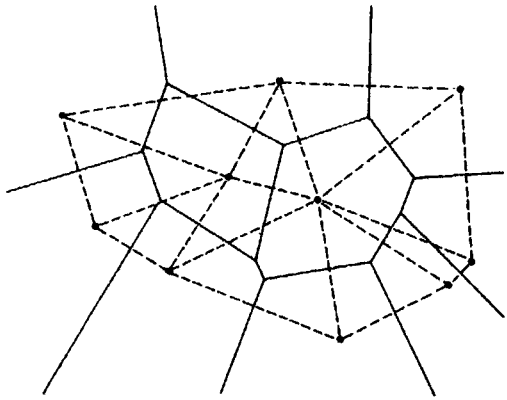


Figure 1.0.1 A planar Voronoi diagram (solid lines) and its dual Delaunay tessellation (dashed lines).

After all possible $(m+1)$-ads have been considered in this way, the result is the Delaunay tessellation.

Voronoi diagrams and Delaunay tessellations are two of a few truly interdisciplinary concepts with relevant material to be found in, but not limited to, anthropology, archaeology, astronomy, biology, cartography, chemistry, computational geometry, crystallography, ecology, forestry, geography, geology, linguistics, marketing, metallography, meteorology, operations research, physics, physiology, remote sensing, statistics, and urban and regional planning. An unfortunate consequence of this is that the material is extremely diffused, of varying mathematical sophistication and quite often idiosyncratic. The amount of material relating to Voronoi diagrams has also been growing steadily since the early 1980s as witnessed by the dominance of recent references in the References. This proliferation results in part from the opening up of new areas of application and in part from the basic concept being extended in a variety of ways (see Chapter 3). Consequently, our major motive for writing this book was to synthesize and to unify as much of the material as possible and to present it in a structured form. This seemed particularly important at the time we wrote the First Edition because no other book had appeared which was devoted exclusively to the topic. The one which came closest was Aurenhammer (1988a) (also published as Aurenhammer, 1990a, 1991) but this was a short (87 pages), highly selective survey which emphasized computational algorithms and only offered brief descriptions of applications. A number of contemporary books also provided limited treatments of Voronoi diagrams as part of more extensive coverage, for example, Ahuja and Schachter (1983), Preparata and Shamos (1985), Iri and Koshizuka (1986), and Stoyan *et al.* (1987) but in most cases Voronoi diagrams accounted for less than ten percent of the books' contents. In addition, the second-last work was in Japanese, which restricted its exposure to an English language audience.

While texts which devote part of their contents to Voronoi diagrams continue to be published (e.g. de Berg *et al.*, 1997) and others focusing on a particular kind of Voronoi diagram (e.g. Møller, 1994) or a particular area of application (e.g. Agrell, 1997) are starting to appear, we believe that our text remains the only one which attempts to provide a comprehensive and balanced treatment of all aspects of Voronoi diagrams including their definition, history, computation, statistical properties and applications. It is our hope and intention that the book should appeal equally to those whose concern with Voronoi diagrams is theoretical, practical or both. However, in view of the continuing development of Voronoi diagrams, we feel that, like the First Edition, this one should be regarded as a report on the current state of affairs from a most dynamic area rather than an exhaustive study.

1.1 OUTLINE

We have already noted that the concept of the Voronoi diagram is used extensively in a variety of disciplines and has independent roots in many of them. In Section 1.2 we begin our treatment of the Voronoi diagram by tracing its origins and that of the Delaunay tessellation from the nineteenth century up to the present. In the course of this exploration we identify the other names by which both structures are known. Before proceeding further, in order to make the book as self-contained as possible, in Section 1.3 we briefly review some basic features of several topics involved in the subsequent discussion of Voronoi diagrams. These are vector geometry, graphs, stochastic point processes and computational efficiency. The material in this section is not intended to be exhaustive but rather to provide a sufficient introduction for those who may be unfamiliar with such topics.

Chapter 2 begins with a formal definition of both the Voronoi diagram (Section 2.1) and the Delaunay tessellation (Section 2.2). Sections 2.3 and 2.4 present properties of both structures, most of which are exploited in applications described elsewhere in the book. A number of other spatial structures which are related to the Delaunay tessellation and which have been used in a variety of applications including diffusion modelling, numerical taxonomy and the exploratory display and analysis of data are described in Section 2.5. The chapter concludes by considering techniques for determining if a given structure is a Voronoi diagram and, if not, the extent to which it approximates one.

A major reason for the continuing success of the Voronoi diagram is that it can be generalized in a variety of ways. Such generalizations are the subject of Chapter 3 and include weighting the points in various ways (Section 3.1), considering regions associated with subsets of points rather than individual points (Sections 3.2 and 3.3), including obstacles in the space (Section 3.4), considering regions associated with sets of geometric features other than points (Sections 3.5 and 3.6), examining Voronoi diagrams

in non-Euclidean spaces (Section 3.7), on networks (Section 3.8), and for moving points (Section 3.9). Accompanying each extension is a discussion of its applications.

In Chapter 4, after considering some computational preliminaries (Section 4.1), we examine data structures for representing the Voronoi diagram (Section 4.2). Then we present major algorithms for constructing it (Sections 4.3–4.5). For application purposes it is necessary to be able to implement these algorithms and in Section 4.6 we consider various practical techniques for doing this. Up to this point in the chapter the discussion concentrates on the planar Voronoi diagram but in Section 4.7 we consider Voronoi diagrams in higher dimensions and in Sections 4.8 and 4.9 we examine algorithms for generating the various generalized diagrams introduced in Chapter 3.

The Poisson Voronoi diagram (PVD) refers to the situation in which points are located in space 'at random' according to the homogeneous Poisson point process (Section 1.3.3). This diagram has been used extensively both as a descriptive and a normative model in the investigation of a wide range of empirical situations in both the natural and social sciences (Section 5.3) as diverse as galaxies, nesting territories of Royal Terns and mouthbreeder fish, and carbon particles in steels. Since the PVD is an infinite Voronoi diagram, Chapter 5 begins with a consideration of the properties of such structures (Section 5.1). We then focus on the major properties of the PVD (Section 5.2) and its constituent regions (Section 5.5), as well as those of the dual Poisson Delaunay tessellation (Section 5.11), sections of the three-dimensional PVD (Section 5.7) and generalizations of the PVD produced by weighting the points (Section 5.8), considering subsets of points rather than individual points (Section 5.9) and the PVD on the surface of a sphere (Section 5.10). We also consider stochastic processes induced by the PVD (Section 5.6) and briefly examine other types of 'random' Voronoi diagrams (Section 5.12). In view of the importance of the PVD and associated Voronoi diagrams to applications, the emphasis in this chapter is on the presentation of results relating to characteristics of Voronoi regions in such structures. Although some of these are exact results derived analytically, the majority are estimates obtained from Monte Carlo simulation procedures, the different types of which are described in Section 5.4.

In the remaining chapters the emphasis shifts towards applications of Voronoi diagrams. We identify four major areas, spatial interpolation, models of spatial processes, point pattern analysis, and locational optimization, each of which is the subject of a separate chapter. In Chapter 6, under the general heading of spatial interpolation, we consider how the Voronoi diagram has been used to facilitate the presentation and analysis of spatial data (values of one or more variables observed at a set of data sites in space). Such uses include employing Voronoi diagrams and Delaunay tessellations for defining various types of neighbour relationships (Section 6.1), creating meshes for finite element methods (Section 6.5), ordering multivariate data (Section 6.6), and constructing continuous surfaces from a set of data values observed at

punctiform data sites (Sections 6.1–6.3), including the approximation of such surfaces (Section 6.4).

Chapter 7 considers how, by equating the procedures involved in defining the Voronoi diagram with assumptions involved in spatial processes, we may produce simple spatial models. At least four types of spatial processes, assignment (Section 7.1), growth (Section 7.2), some spatial temporal processes (Section 7.3) and spatial competition (Section 7.4) can be represented in this way. The application of such models is extensive and examples from areas as diverse as crystallography, solid state physics, biology, astronomy, physiology, archaeology and urban economics are presented.

Use of the Voronoi diagram and the Delaunay tessellation in point pattern analysis is described in Chapter 8. Section 8.1 considers different methods using the Voronoi diagram while Section 8.2 examines those methods which make use of the Delaunay tessellation. In both of these sections our concern is with how the points in a given set are positioned with respect to each other and the space in which they are located. In Section 8.3 our attention shifts to a consideration of how the members of the point set are located with respect to other objects, not members of the set, which are also located in the same space. These objects may be other points, lines or areas, or a combination of all three. We also examine how Voronoi concepts can be used to describe the shape of a point pattern (Section 8.4), estimate its spatial intensity (Section 8.5) and divide it into sub-units (Section 8.6). Section 8.7 echoes Section 5.6 by considering some stochastic point processes that are induced by Voronoi diagrams.

Finally, Chapter 9 deals with the various ways in which the Voronoi diagram can be applied in solving locational optimization problems. Although problems relating to locational optimization on a network have been the subject of extensive investigation, less work has been undertaken on problems in the plane and it is such problems which are emphasized in this chapter. Planar optimization problems are important in a number of natural and social sciences. Examples range from locating observation sites for estimating a continuous spatial random variable to the location of various types of public facilities such as health clinics, fire stations and public transportation stops. All of these problems involve objects which can be represented as points and they, and other situations, form the subject matter of Section 9.2. Section 9.3 considers problems relating to the location of linear features, while Section 9.4 extends the discussion to situations where the objects under study are not located synchronously in the space. The final section (Section 9.5) illustrates how locational optimization procedures relate to fitting a Voronoi diagram to a given tessellation, a topic which is treated more generally in Section 2.6.

1.2 HISTORY OF THE CONCEPT OF THE VORONOI DIAGRAM

It is quite possible that the concept of the Voronoi diagram is of considerable antiquity. As we show elsewhere in the book (but especially in Chapter 7), many different kinds of natural structures closely resemble Voronoi diagrams and it seems unlikely that such structures would have gone unnoticed by early scientists and observant laymen alike. In his treatment of cosmic fragmentation in both *Le Monde de Mr Descartes, ou Le Trait de la Lumière* published in 1644 (but written between 1629 and 1633) and in Part III of *Principia Philosophiae*, also published in 1644, Descartes uses Voronoi-like diagrams to show the disposition of matter in the solar system and its environs (see Figure 1.2.1). Since these diagrams are not accompanied by any special commentary relating to their construction, it is possible that such figures were not uncommon at that time. In fact, diagrams such as that in Figure 1.2.1 are actually more akin to a generalized version of a Voronoi diagram, known as a weighted Voronoi diagram (see Section 3.1).

The first undisputed comprehensive presentations of the concept that we are aware of appear in the work of Peter Gustav Lejeune Dirichlet (1805–1859) and Georgy Fedoseevich Voronoy (Georges Voronoï) (1868–1908) who, in their studies on positive definite quadric forms (Dirichlet, 1850; Voronoï, 1907, 1908, 1909) considered a special form of the Voronoi diagram. Dirichlet treated two- and three-dimensional cases whereas Voronoï examined the general m-dimensional case. Their concern was with the distribution of points with integer coordinates that give minima of the values of a given quadric form. In that context, they considered the set of points regularly placed in the m-dimensional space generated by linear combinations of m linearly independent vectors with integer coefficients. This set contains infinitely many points, and the Voronoi diagram generated by this set of points gives the partition of the space into mutually congruent polyhedra. Using this concept Voronoï, for example, extended the work of Minkowski (1897) to give a simple proof of the upper bound $2(2^m-1)$ of the number of $(m-1)$-dimensional facets of an m-dimensional polyhedron, the translational repetition of which can fill the whole space without overlap. As far as we can determine, it was the Russian number theorist, Boris Nikolaevitch Delone (1890–1980) (see below), who was the first to associate the names of Dirichlet and Voronoï with such polyhedra. He coined both the term *domaine de Dirichlet* (Dirichlet domain) (Delaunay, 1929a, b) and *Voronoïschen Bereich* (Voronoi region) (Delaunay, 1932) which have subsequently become the two most frequently encountered terms.

Given that the initial developments of the Voronoi diagram concept involved sets of points which were regularly placed in space, it is not surprising that some of its first applications were in crystallography. Work in this area began in the late nineteenth and early twentieth centuries (see Nowacki, 1933, 1976, for reviews) and was dominated by German and Russian researchers. At this time Voronoi regions were known by a number of names.

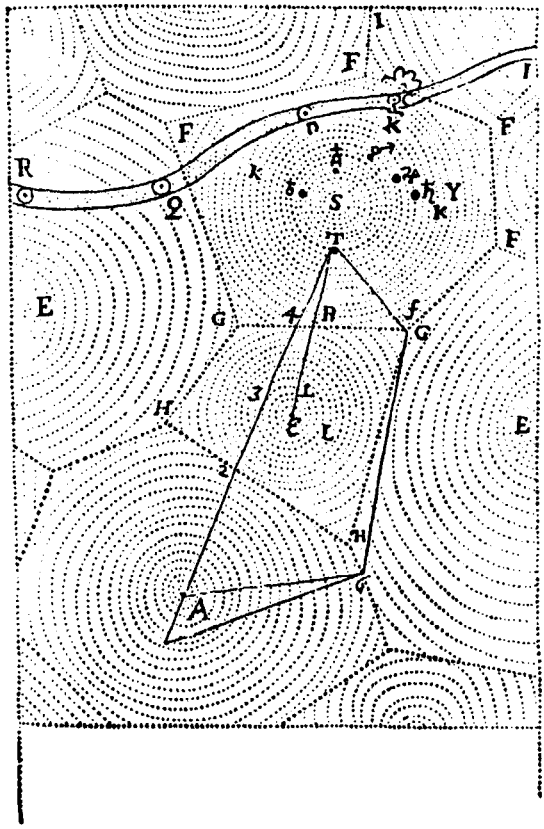


Figure 1.2.1 The disposition of matter in the Solar System and its environs by Descartes (S is the sun; ϵ is a star; RQD represents the path of a comet; polygonal areas represent heavens). (*Source*: Copy of the original in Descartes, 1644, reproduced in Mahoney, 1979.)

For example, Fedorov (1885), who was the first to determine the five combinatorial types of space-filling parallelohedra in $\mathbb{R}^3$ (see Section 7.1), called such regions *stereohedra*, while Schoenflies (1891), Steinitz (1906) and Speiser (1927) used the term *Fundamentalbereich* (fundamental area), and Wulff (1908) used *Wirkungssphäre* (sphere of influence). Later Schoenflies (1923) and Niggli (1927) independently introduced the term *Wirkungsbereich* (domain of action, field of activity, area of influence). The line of research originating in the early work concerned with identifying congruent space-filling domains has continued to the present (see Section 7.1) and Grünbaum and Shephard (1980) have coined the term *plesiohedra* for Voronoi regions encountered in this context.

At the same time as the Voronoi diagram concept was being applied in crystallography, it appears to have been developed independently in at least two other areas, both involving spatial interpolation. The first of these

occurred in meteorology when Thiessen (1911) used Voronoi regions as an aid to computing more accurate estimates of regional rainfall averages (see Section 6.1.1). While acknowledging Thiessen's paper, Horton (1917) claims to have developed the procedure independently in exactly the same context. Notwithstanding this, Whitney (1929) refers to the procedure as 'Thiessen's method' and since then the term *Thiessen polygon* has remained a popular one in two-dimensional applications in meteorology, geography and related social science disciplines.

The second area involved the estimation of ore reserves in a deposit using information obtained from bore holes (see Section 6.1.1). The earliest work on this topic by Boldyrev appeared in Russia in 1909 (Smirnov *et al.*, 1960) and was independently developed in the United States by Davis and Harding (Harding, 1920–21, 1923). Here the Voronoi regions are termed simply *area of influence polygons*, a name which remains current in this area of application (Hayes and Koch, 1984).

Despite this expansion of the use of Voronoi diagrams, further independent rediscoveries continued. In their study of the chemical properties in metallic sodium the physicists Wigner and Seitz (1933) described Voronoi regions for points arranged on a lattice in three-dimensional space. Although they did not label these polyhedra, the term *Wigner–Seitz regions* has come into quite common usage amongst physicists even when they are aware of the precedence of Voronoi diagrams, with the former term being used when the regions are regular polytopes and the latter when they are not. Somewhat ironically, over twenty years later, Voronoi diagrams were rediscovered in a similar context by Frank and Kasper (1958) modelling complex alloy structures as packings of spheres. Their concern focused on the neighbour relationships defined by the Voronoi polyhedra of atoms contiguous to that of the Voronoi polyhedron of a given atom, which they labelled the *domain of an atom*. Later still, Venables and Ball (1971) defined the equivalent of Voronoi regions to approximate capture zones of xenon crystals growing on graphite.

Meijering (1953) also appears to have been unaware of Voronoi concepts since he developed a crystal aggregate model which is a specific instance of the Voronoi Growth Model (see Section 7.2). In his *cell model*, crystals start to grow simultaneously and isotropically from nuclei randomly distributed in space so that the resulting structure is equivalent to a Poisson Voronoi diagram (see Chapter 5).

Voronoi regions were also developed independently in the study of codes. In the communication model introduced by Shannon (1948a,b, 1949), one of C equally likely messages is presented for communication to a transmitter containing a code book with C code words, each of which can be represented by a vector in $\mathbb{R}^3$. A received signal is represented by a vector consisting of the sum of the sent vector (code word) and a noise vector whose components are independent Gaussian variates. The ideal receiver for such a communication system is one which chooses the code word nearest to the received vector, since it can be shown that such a strategy minimizes

the average error probability of the transmitted code words. This operation is equivalent to defining the Voronoi diagram of the set of code words and noting in which Voronoi region a transmitted message is located. Since this type of decoding scheme was labelled minimum distance decoding or maximum likelihood decoding (Shannon, 1959), the resulting Voronoi regions have been labelled *maximum likelihood regions* (Slepian, 1965, 1968). Later still, in the same context Conway and Sloane (1992, p.56) refer to the Voronoi diagram as the *Voronoi honeycomb*.

In a span of only two years in the 1960s, Voronoi diagrams were rediscovered not once but twice more in the field of ecology. First, when estimating the intensity of trees in a forest, Brown (1965) defined a Voronoi region for an individual tree, calling it the *area potentially available (APA)* to a tree. One year later Mead's paper appeared (Mead, 1966) using the same concept for plants and labelling the resulting Voronoi regions as *plant polygons*. This name has remained reasonably common in this area of application despite the early recognition by Jack (1967) that such polygons are equivalent to Voronoi regions.

The final rediscovery that we are aware of is by Hoofd *et al.* (1985) who defined Voronoi regions with respect to the centres of capillaries in sections of tissue, labelling the resulting polygons capillary domains. However, as late as 1987 Icke (Icke and van de Weygaert, 1987) admits to developing them yet again, this time in an astronomical context, only to be saved from claiming such work as novel by his co-author's discovery of the vast field of literature on Voronoi diagrams.

Given the continuing integration of Voronoi concepts into so many fields of application (and the sales of the first edition of this book!), we do not anticipate that there will be any future rediscoveries of the Voronoi diagram. However, since most of its generalizations have a much shorter history and more specialized applications (and some individuals did not contribute to our sales!), it seems almost inevitable that many of these generalizations will inherit the trait of rediscovery. Indeed, weighted Voronoi diagrams (see Section 3.1) are already witness to this, with the additively weighted Voronoi diagram (see Section 3.1.2) being rediscovered by Medvedev (1994) and Goede *et al.* (1997) and the multiplicatively weighted diagram (see Section 3.1.1) by Aparicio and Cocks (1995), Moukarzel (1993), and Gerstein *et al.* (1995), and both diagrams by Parr (1995a,b, 1997a,b).

Although much of the development and many of the initial applications of Voronoi diagrams occurred in the natural sciences, the earliest application of Voronoi concepts known to us appears in a map included in the *Report on the Cholera Outbreak in the Parish of St. James, Westminster, During the Autumn of 1854* (Cholera Inquiry Committee, 1855). This map shows a continuous broken line, described as the 'Boundary of equal distance between Broad Street Pump and other Pumps', which encloses an area around the Broad Street Pump. Although there is no attribution associated with this map, it is most likely the work of John Snow whose map showing the spatial distribution of cholera deaths around the Broad Street Pump in

1854 (Snow, 1855) has become 'the most famous 19th century disease map' (Meade *et al.*, 1988) being reproduced in a variety of forms in many texts in both epidemiology and cartography (e.g. Stamp, 1964; Longmate, 1966; Thrower, 1972; Monmonier, 1996). The remarkable thing about the boundary defined on the map in the *Report* is that distance is not measured in terms of the Euclidean metric but in terms of distance along the street network of Westminster (including allowances for the presence of culs-de-sac), making this a network Voronoi-area diagram (see Section 3.8.3), a concept which would not reappear for nearly one hundred and fifty years.

The next social science application appears in the work of the German dialectologist Carl Haag (1860–1946) who used the Voronoi diagram as a means of visualizing dialect variation and identifying linguistic divides (isoglosses). In his study of dialects in south-west Germany (Haag, 1898), he defined the equivalent of the Voronoi diagram of a set of localities at which dialect data had been collected. He then decorated the Voronoi edge shared by two localities if those localities differed in terms of a dialect feature. Highly decorated edges thus signified the presence of isoglosses. The basis of this approach was adopted by the three main German dialectical institutes and variants are still in current use (see Section 7.1). Although Carl Haag made no mention of this, it is likely that his ideas were influenced by the work of the early chemists, one of whom was his brother, F. Haag, who wrote quite extensively on Voronoi diagrams of point sets in two dimensions (Haag, 1911, 1923, 1924, 1925).

Despite these early studies, most social science applications have been developed only in the last fifty years. The first of these was by Bogue (1949) who used the Voronoi polygons defined about US metropolitan centres (represented as points on a map) as surrogates for their market areas. Aided by other early studies by Snyder (1962) and Dacey (1965), this has remained the main area of application with work extending down to the level of individual retail stores, although geographers, anthropologists and archaeologists have also used the concept to model other types of human territorial systems. Related antecedents to such applications, some dating back to the mid-nineteenth century (see Shieh, 1985, for a review) can be found in spatial economics in the 'law of market areas'. This law considers the form of the boundary between the market areas of two competing centres under various conditions of market prices and transportation costs. Similar considerations are used in the definition of various kinds of weighted Voronoi diagrams (see Section 3.1). Following the suggestion of Evans (1967), geographers were also the first to use Voronoi (and Delaunay) concepts in the analysis of two-dimensional point patterns (Boots, 1974).

So by the 1960s knowledge of the Voronoi concept was current in both the natural and social sciences. However, empirical applications remained somewhat limited because of the lack of a simple and efficient means of constructing them, with practitioners relying on methods involving compass and ruler and fraught with ambiguities (Kopec, 1963). This situation motivated solutions from the rapidly developing area of computer science and by

the early 1970s a number of algorithms had been developed to construct Voronoi diagrams in both two and three dimensions. In turn, this work stimulated other developments in computer science which contributed to the now flourishing field of computational geometry. This endeavour is typified by the seminal paper of Shamos and Hoey (1975). Not only did they present an algorithm for constructing the Voronoi diagram, they also illustrated how it could be used to solve a series of what were then seen as essentially independent problems relating to a finite set of distinct points in Euclidean space, such as finding the minimum spanning tree, identifying the nearest neighbour of each point, and finding the largest circle, containing no points of the set, whose centre is inside the convex hull of the point set (see Section 2.3). In addition, they also suggested ways of generalizing the Voronoi diagram by considering Voronoi regions associated with subsets of k points of the entire point set rather than individual points (see Section 3.2). As a consequence of this and other contemporary initiatives, in the past twenty-five years the basic concept has been extended in a great variety of ways which are detailed in Chapter 3.

The dual concept of the Delaunay tessellation also has a history marked by rediscoveries, although these are not as frequent as in the case of the Voronoi diagram. The concept originated with Voronoï (1908) who defined it by way of the neighbour relationships in the Voronoi diagram, referring to the resulting structure as *l'ensemble (L) de simplexes*. However, it was Delone who first defined the tessellation using the empty sphere method. Delone's initial paper on this topic was included in an International Mathematical Congress held in Toronto, Canada, in 1924. Both this paper and its follow-up (which was dedicated to the memory of Voronoï) were written in French under the name of Delaunay (Delaunay, 1928, 1934), as were other of his earlier papers written in French (Delaunay, 1929a,b) and German (Delaunay, 1932). These circumstances may help to explain why Delaunay is often encountered misspelt Delauney (e.g. Baranyai and Ruff, 1986; Lingas, 1986a,b; Kermode and Weaire, 1990; Cressie, 1991; Finney, 1991; Voigtmann *et al.*, 1994; Blatov *et al.*, 1995) and why Davies and Bell (1996) mistakenly label him as a French mathematician. Delone did not propose a term for the tessellation and followed Voronoï's lead by calling it *la partition de l'espace en tetraedres L*. The term *L-partition* is still in current use in some quarters (e.g. Gruber and Lekkerkerker, 1987).

One of the properties of the Delaunay tessellation in two dimensions is that the individual triangles are as equilateral as possible (see Section 2.4, Property D15). Thus, when Whitney (1929) adopted this criterion to select triangles for a triangulation for use in spatial interpolation, he unknowingly constructed a Delaunay tessellation.

Smith (1964) seems to have rediscovered the concept in his work on three-dimensional aggregates, labelling the structure the *simplical graph*. Nevertheless, a year later he recognized the precedence of Delaunay's work and re-labelled the constituent regions as *Delaunay simplices* (Smith, 1965), a term which appears to have been coined by Rogers (1964) in his compre-

hensive work on tessellations. Since then, the term Delaunay tessellation has been increasingly used which, given Delone's pivotal role in developing the concept, seems entirely appropriate.

There has been at least one recent rediscovery of the structure since this naming, that of Christ *et al.* (1982a) in the context of random lattice field theory. Although they recognized that the structure is the dual of the Voronoi diagram, they did not give a name to it.

1.3 MATHEMATICAL PRELIMINARIES

In this text, the reader is assumed to be familiar with basic mathematical notions related to sets, matrices, derivatives, geometry, integration, probability and statistics (such as Bartle, 1964 (real analysis); Feller, 1957 (probabilities) and Freund, 1962 (statistics)). For the reader's convenience, this section presents some basic geometrical notions that will be commonly used in this text. The major notions are those related to vector geometry, graphs, spatial stochastic point processes and computational efficiency. Since space is limited, some developments may not be in full detail. The reader who wishes to understand these notions with full formal derivations should consult the relevant mathematical textbooks (such as Royden, 1963 (real analysis); Busacker and Saaty, 1965 (graph theory); Wilks, 1962 (statistics); Karlin, 1969 (stochastic processes); Knuth, 1968 (algorithms)).

1.3.1 Vector Geometry

Consider an m-dimensional real Cartesian space, denoted by $\mathbb{R}^m$ ($\mathbb{R}^2$ in Figure 1.3.1). In this space an m-tuple or the Cartesian coordinates $(x_1, \ldots, x_m)$ indicate a *point*, denoted by p ($p = (x_1, x_2)$ in Figure 1.3.1). If there are n points in $\mathbb{R}^m$, these points are denoted by $p_1, \ldots, p_n$ with Cartesian coordinates $(x_{i1}, \ldots, x_{im})$, $i = 1, \ldots, n$ (for brevity, $i = 1, \ldots, n$ will be referred to as $i \in I_n$ hereafter, where $I_n = \{1, \ldots, n\}$ is a set of integers from 1 to n). Points p and p_i may also be represented by vectors

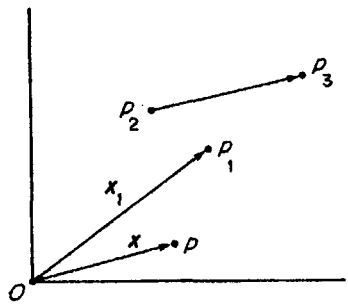


Figure 1.3.1 Vectors.

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \quad \text{and} \quad \boldsymbol{x}_i = \begin{bmatrix} x_{i1} \\ \vdots \\ x_{im} \end{bmatrix}, \tag{1.3.1}$$

respectively. Note that the corresponding row vectors are indicated by $\boldsymbol{x}^{\mathrm{T}} = (x_1, \ldots, x_m)$ and $\boldsymbol{x}_i^{\mathrm{T}} = (x_{i1}, \ldots, x_{im})$, respectively, where T is a transpose operator; by definition, $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{im})^{\mathrm{T}}$. In some contexts, it is more convenient to represent a point p as the arrow from the origin o of $\mathbb{R}^m$ to the point p, denoted by $\overrightarrow{op}$. The length of the arrow in $\mathbb{R}^2$ is, using the Pythagorean theorem, given by $\sqrt{x_1^2 + x_2^2}$. Similarly, in $\mathbb{R}^m$, the length of the arrow, or the *norm* of a vector $\boldsymbol{x}$, denoted by $\|\boldsymbol{x}\|$, is given by

$$\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}} = \sqrt{\sum_{i=1}^{m} x_i^2} \tag{1.3.2}$$

(generally, $\boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{x}_2 = (x_{11}x_{21} + \ldots + x_{1m}x_{2m})^{\mathrm{T}}$ indicates the *inner product* of vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$). If a vector is designated as an arrow, it may be shifted provided that its length and direction are maintained; that is, an arrow may be shifted in a parallel way. In Figure 1.3.1, for example, $\overrightarrow{p_2p_3}$ and $\overrightarrow{op}$ have the same length and the same direction, and so these two arrows indicate the same vector. It should be noted, however, that if a vector is designated as a point, then the point cannot be shifted. For example, in Figure 1.3.1 a point p is indicated by $\overrightarrow{op}$ but not by $\overrightarrow{p_2p_3}$. We sometimes call a vector indicating a point a *location vector* or a *point vector*.

The scalar multiplication of $\boldsymbol{x}$ by a positive scalar λ, i.e. $\lambda\boldsymbol{x} = (\lambda x_1, \ldots, \lambda x_m)^{\mathrm{T}}$, is represented by the arrow obtained from extending (shrinking) $\overrightarrow{op}$ λ times in the same direction (Figure 1.3.2(a)). The scalar multiplication of $\boldsymbol{x}$ by a negative scalar $-\lambda$, i.e. $-\lambda\boldsymbol{x}$, is represented by the arrow obtained from extending (shrinking) $\overrightarrow{op}$ λ times in the opposite direction (Figure 1.3.2(a)). The vector addition, $\boldsymbol{x}_1 + \boldsymbol{x}_2 = (x_{11} + x_{21}, \ldots, x_{1m} + x_{2m})^{\mathrm{T}}$, is represented by the diagonal $\overrightarrow{op_3}$ of the parallelogram constructed from $\overrightarrow{op_1}$ and $\overrightarrow{op_2}$ (Figure 1.3.2(b)). The vector subtraction, $\boldsymbol{x}_1 - \boldsymbol{x}_2$, is obtained from the addition $\boldsymbol{x}_1 + (-\boldsymbol{x}_2) = (x_{11} - x_{21}, \ldots, x_{1m} - x_{2m})^{\mathrm{T}}$, where the vector $-\boldsymbol{x}_2$ is obtained from the vector $\boldsymbol{x}_2$ multiplied by a negative scalar -1; that is, $\boldsymbol{x}_1 - \boldsymbol{x}_2$ is indicated by the diagonal $\overrightarrow{op_5}$ of the parallelogram constructed from $\overrightarrow{op_1}$ and $\overrightarrow{op_4}$ (Figure 1.3.2(c)). Alternatively, the vector subtraction, $\boldsymbol{x}_1 - \boldsymbol{x}_2$, is represented by the diagonal $\overrightarrow{p_2p_1}$ of the parallelogram constructed from $\overrightarrow{op_1}$ and $\overrightarrow{op_2}$ (observe that $\overrightarrow{op_5}$ and $\overrightarrow{p_2p_1}$ represent the same vector $\boldsymbol{x}_1 - \boldsymbol{x}_2$).

Suppose that two vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are given by those in Figure 1.3.3(a). Then, we can find a scalar λ that satisfies $\boldsymbol{x}_2 = \lambda\boldsymbol{x}_1$. On the other hand, we cannot find such a λ for the vectors shown in Figure 1.3.3(b). We say that the vectors in panel (a) are *linearly dependent* and those in panel (b) are *linearly independent*. This notion can be extended to n vectors in $\mathbb{R}^m$. We say that vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are *linearly dependent* in $\mathbb{R}^m$ if there exists a set $\Lambda = \{\lambda_1, \ldots, \lambda_n\}$ of n scalars which are not all zero such that

$$\lambda_1\boldsymbol{x}_1 + \ldots + \lambda_n\boldsymbol{x}_n = 0. \tag{1.3.3}$$

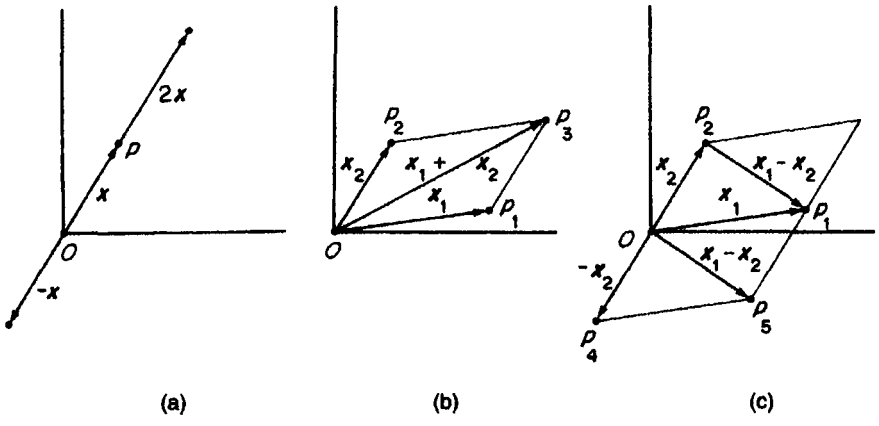


Figure 1.3.2 Vector operations: (a) scalar multiplication; (b) addition; (c) subtraction.

On the other hand, if the only Λ for which equation (1.3.3) holds is $\lambda_1 = \ldots = \lambda_n = 0$, then the vectors are said to be *linearly independent*. The number of linearly independent vectors in $\mathbb{R}^m$ is at most m. For example, in $\mathbb{R}^2$, as is shown in Figure 1.3.3(c), two vectors x_1 and x_2 are linearly independent, but three vectors x_1, x_2 and x_3 are linearly dependent, because, as is indicated by the broken arrows in Figure 1.3.3(c), $\lambda_1 x_1 + \lambda_2 x_2 - x_3 = 0$ holds.

When two vectors x_1 and x_2 form the right angle as in Figure 1.3.4(a), we say that the vectors x_1 and x_2 are *orthogonal*, and denote it by $x_1 \perp x_2$. From Figure 1.3.4(a) it is evident that vectors x_1 and x_2 are orthogonal if and only if $\|x_1 + x_2\|$ equals $\|x_1 - x_2\|$, i.e. $(x_1 + x_2)^T (x_1 + x_2) = (x_1 - x_2)^T (x_1 - x_2)$. After a few steps of calculation, we obtain the following relation:

$$x_1 \perp x_2 \text{ if and only if } x_1^T x_2 = 0. \tag{1.3.4}$$

In Figure 1.3.4(b) we perpendicularly project p_2 or the line containing $\overrightarrow{op}$. Let p_2' be the projected point and x_2' be its location vector. Then we say that p_2' is the *perpendicular projection* or *orthogonal projection* of p_2 on the line containing $\overrightarrow{op_1}$. Since $x_2' = \lambda x_1$, and $(\lambda x_1 - x_2)^T (\lambda x_1) = 0$, from equation (1.3.4), the vector x_2' is written as

$$x_2' = \frac{x_2^T x_1}{\|x_1\|} x_1. \tag{1.3.5}$$

Equation (1.3.5) becomes simpler when p_i with coordinates (x_{i1}, x_{i2}) in $\mathbb{R}^2$ is projected orthogonally onto the x_1-axis. In this case the orthogonal projection of p_i onto the x_1-axis is given by $(x_{i1}, 0)$. Similarly, the orthogonal projection of p_i with coordinates (x_{i1}, x_{i2}, x_{i3}) in $\mathbb{R}^3$ onto the x_1–x_2 plane is given by $(x_{i1}, x_{i2}, 0)$. Furthermore, we can project a set S in $\mathbb{R}^3$ orthogonally onto the x_1–x_2 plane by $S' = \{(x_1, x_2, 0) \mid (x_1, x_2, x_3) \in S\}$ (note that a set of elements constrained by conditions is denoted by {elements | conditions}). In this case we say that S' is the *orthographic projection* of S onto the x_1–x_2

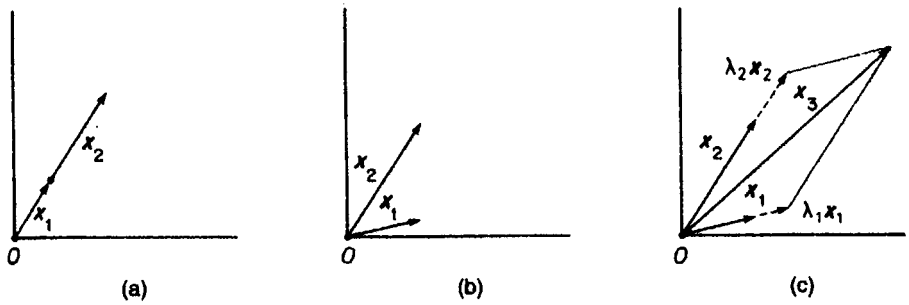


Figure 1.3.3 (a) Two vectors being linearly dependent; (b) linearly independent vectors; (c) three vectors being linearly dependent.

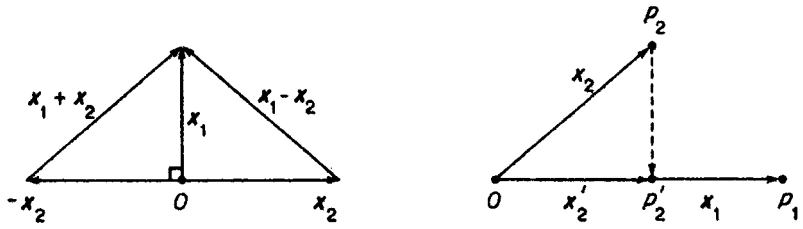


Figure 1.3.4 (a) Orthogonal vectors; (b) an orthogonal projection.

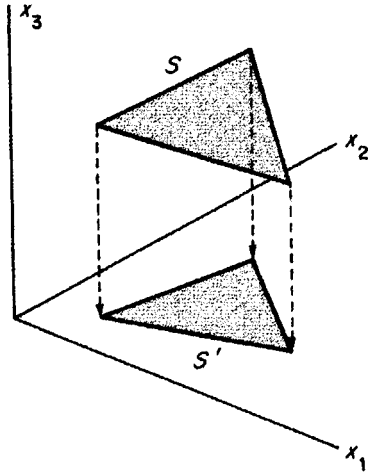


Figure 1.3.5 Orthographic projection.

plane. Figure 1.3.5 depicts the orthographic projection of a triangle placed in $\mathbb{R}^3$ onto the x_1–x_2 plane.

As is seen in Figure 1.3.2(c), $\overrightarrow{p_2p_1}$ and $\overrightarrow{op_5}$ both represent the vector $\boldsymbol{x}_2 - \boldsymbol{x}_1$. Since the length of $\overrightarrow{op_5}$ is given by equation (1.3.2), where $\boldsymbol{x} = \boldsymbol{x}_2 - \boldsymbol{x}_1$, we understand that the *Euclidean distance* between a point $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{im})^{\mathrm{T}}$ and a point $\boldsymbol{x}_j = (x_{j2}, \ldots, x_{jm})^{\mathrm{T}}$ is given by

$$\begin{aligned} \|\boldsymbol{x}_i - \boldsymbol{x}_j\| &= \sqrt{(\boldsymbol{x}_i - \boldsymbol{x}_j)^T (\boldsymbol{x}_i - \boldsymbol{x}_j)} \\ &= \sqrt{\textstyle\sum_{k=1}^{m} (x_{ik} - x_{jk})^2}\,. \end{aligned} \tag{1.3.6}$$

We call the m-dimensional real Cartesian space with the Euclidean distance the m-dimensional *Euclidean space*.

In the Euclidean space, we now show a few topological concepts. For a point $\boldsymbol{c}$ in $\mathbb{R}^m$ and $\varepsilon > 0$, we define the set

$$N_\varepsilon(\boldsymbol{x}) = \{\boldsymbol{x} \mid \|\boldsymbol{x} - \boldsymbol{c}\| < \varepsilon\}, \tag{1.3.7}$$

which is called an *open ball* (in particular, an *open disk* for $m = 2$, as in Figure 1.3.6(a)) with radius ε centred at $\boldsymbol{c}$. Using this term, we say that a set A in $\mathbb{R}^m$ is *open* if every point in A is the centre of some open ball entirely contained in A (Figure 1.3.6(c)); a set A in $\mathbb{R}^m$ is *closed* if $\mathbb{R}^m \setminus A$ is open in $\mathbb{R}^m$ where $\setminus$ means a complement operator (i.e. for two sets A and B, $A \setminus B$ means a set of points which are included in A but not in B); a point $\boldsymbol{x}$ in $\mathbb{R}^m$ is a *boundary point* of A if every open ball centred at $\boldsymbol{x}$ contains a point of A and a point of $\mathbb{R}^m \setminus A$ (Figure 1.3.6(d)). The set of all boundary points of A is called the *boundary* of A, and is denoted by ∂A. If a set A includes the boundary of A, then every point in $\mathbb{R}^m \setminus A$ is the centre of some open ball entirely contained in $\mathbb{R}^m \setminus A$. Thus a set A is closed if it contains its boundary. A point $\boldsymbol{x}$ in $\mathbb{R}^m$ is called an *interior point* of a set A if there exists an open ball centred at $\boldsymbol{x}$ that contains only points in the set A (point $\boldsymbol{x}$ in Figure 1.3.6(c)). If $\boldsymbol{c}$ is a point in $\mathbb{R}^m$ ($\boldsymbol{c} \in \mathbb{R}^m$), then any set that contains an open set containing $\boldsymbol{c}$ is called a *neighbourhood* of $\boldsymbol{c}$ in $\mathbb{R}^m$ (generally, $a \in A$ means that a is an element of the set A). The open ball defined by equation (1.3.7) is a neighbourhood of $\boldsymbol{c}$, and is sometimes referred to as an *ε-neighbourhood*. For a set A, the intersection of all closed sets including A is called a *closure* of A.

In conjunction with an open ball, we define the sets

$$\overline{N}_\varepsilon(\boldsymbol{c}) = \{\boldsymbol{x} \mid \|\boldsymbol{x} - \boldsymbol{c}\| \le \varepsilon\} \quad \text{for } \boldsymbol{c} \in \mathbb{R}^m, \varepsilon > 0 \tag{1.3.8}$$

(notice that $<$ in equation (1.3.7) is replaced by $\le$), and

$$\partial\overline{N}_\varepsilon(\boldsymbol{c}) = \{\boldsymbol{x} \mid \|\boldsymbol{x} - \boldsymbol{c}\| = \varepsilon\} \quad \text{for } \boldsymbol{c} \in \mathbb{R}^m, \varepsilon > 0, \tag{1.3.9}$$

which is a hypersphere in $\mathbb{R}^m$ (specifically, a circle in $\mathbb{R}^2$, a sphere in $\mathbb{R}^3$). The set $\partial\overline{N}_\varepsilon(\boldsymbol{c})$ is the boundary of $\overline{N}_\varepsilon(\boldsymbol{c})$. Since the set $\overline{N}_\varepsilon(\boldsymbol{c})$ includes its boundary, the set $\overline{N}_\varepsilon(\boldsymbol{c})$ is a closed set. Thus we call $\overline{N}_\varepsilon(\boldsymbol{c})$ a *closed ball* (specifically, a *closed disk* for $m = 2$, as in Figure 1.3.6(b)) centred at $\boldsymbol{c}$ with radius ε.

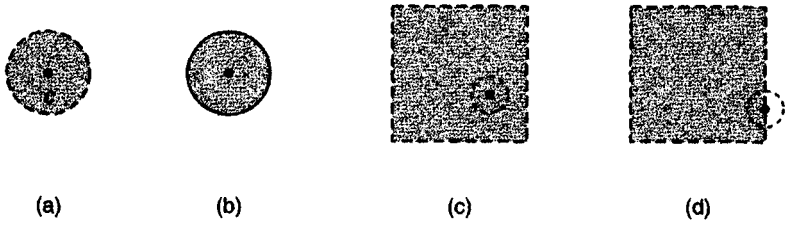


Figure 1.3.6 (a) An open disk; (b) a closed disk; (c) an open set; (d) a closed set.

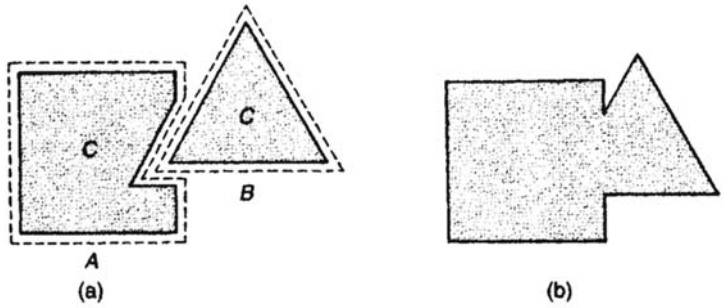


Figure 1.3.7 (a) A disconnected set C; (b) a connected set C.

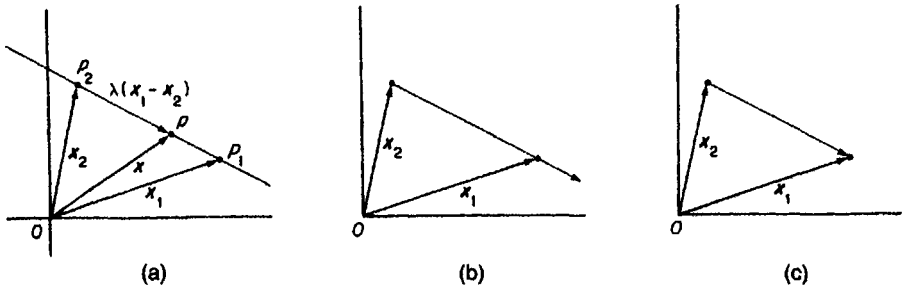


Figure 1.3.8 Lines (bold lines) defined by vectors: (a) a line; (b) a half line; (c) a line segment.

For a set C, if there exist two open sets A and B such that $(A \cap C) \cap (B \cap C) = \varnothing$ and $(A \cap C) \cup (B \cap C) = C$, then the set C is said to be *disconnected* (Figure 1.3.7(a); note that $\cap$ is an intersection operator, i.e. for two sets A and B, $A \cap B$ means a set of points that are included in A and B; $\cup$ is a union operator, i.e. $A \cup B$ means a set of points that are included in A or B; $\varnothing$ denotes a set of no elements or an empty set). If a set is not disconnected, then the set is said to be *connected* (Figure 1.3.7(b)).

As we showed in Figure 1.3.2(c), $\overrightarrow{p_2 p_1}$ represents the vector $x_1 - x_2$. We extend or shrink this arrow by multiplying by a scalar λ keeping its initial point at p_2 (Figure 1.3.8(a)). This extended or shrunken arrow represents the

vector $\lambda(x_1 - x_2)$. Let x be the location vector indicating the head point, p, of this vector. Recalling the vector addition depicted in Figure 1.3.2(b), we notice that $\overrightarrow{op}$ is obtained by adding $\overrightarrow{op_2}$ and $\overrightarrow{p_2p}$ (Figure 1.3.8(a)). Since these arrows represent the vectors x, x_2 and $\lambda(x_1 - x_2)$, respectively, we obtain the equation $x = \lambda(x_1 - x_2) + x_2 = \lambda x_1 + (1 - \lambda)x_2$. From this equation the *(straight) line* passing through p_1 and p_2 ($p_1 \neq p_2$) is written as

$$L_1 = \{x \mid x = \lambda x_1 + (1 - \lambda)x_2, \lambda \in \mathbb{R}\}. \tag{1.3.10}$$

Observing that $x = x_2$ if $\lambda = 0$, and $x = x_1$ if $\lambda = 1$, we notice that the *half line* radiating from p_2 passing through p_1 (Figure 1.3.8(b)) is given by the set

$$L_2 = \{x \mid x = \lambda x_1 + (1 - \lambda)x_2, \lambda \geq 0\}, \tag{1.3.11}$$

and the *(straight) line segment* connecting p_1 and p_2 (Figure 1.3.8(c)) is given by the set

$$L_3 = \{x \mid x = \lambda x_1 + (1 - \lambda)x_2, 0 \leq \lambda \leq 1\}. \tag{1.3.12}$$

The points p_1 and p_2 are called the *end points* of L_3. The line segment L_3 contains its end points p_1 and p_2.

Let us next consider a plane in $\mathbb{R}^3$ passing through points p_1, p_2 and p_3 which are not all on the same line, or equivalently, $x_2 - x_1$ and $x_3 - x_1$ are linearly independent. A point v on the line passing through p_1 and p_2 is, from equation (1.3.10), given by $v = \alpha x_1 + (1 - \alpha)x_2$; a point w on the line passing through p_1 and p_3 is given by $w = \beta x_1 + (1 - \beta)x_3$ (Figure 1.3.9). Thus a point x on the plane is given by $x = \gamma v + (1 - \gamma)w = (\alpha\gamma + (1 - \gamma)\beta)x_1 + \gamma(1 - \alpha)x_2 + (1 - \beta)(1 - \gamma)x_3$ (Figure 1.3.9). It follows from this equation that if we define the set by

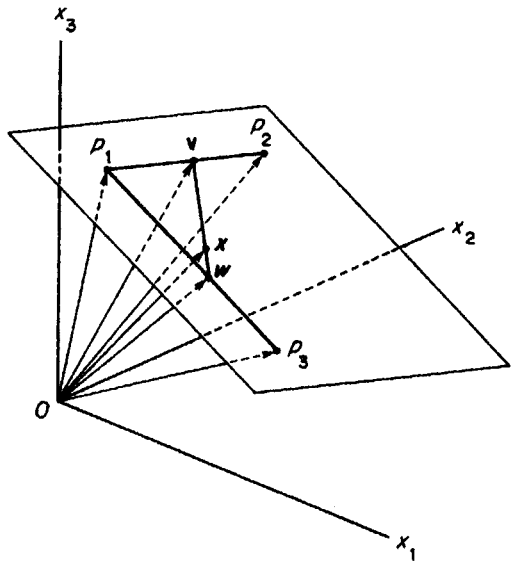


Figure 1.3.9 A plane in $\mathbb{R}^3$.

$$A_1 = \{x \mid x = \lambda_1 x_1 + \lambda_2 x_2 + (1 - \lambda_1 - \lambda_2) x_3, \lambda_1, \lambda_2 \in \mathbb{R}\}, \tag{1.3.13}$$

then the set A_1 represents a plane. More generally, we define the set in $\mathbb{R}^m$ by

$$A_2 = \left\{x \mid x = \sum_{i=1}^{m-1} \lambda_i x_i + \left(1 - \sum_{i=1}^{m-1} \lambda_i\right) x_m, \lambda_i \in \mathbb{R}, i \in I_{m-1}\right\}, \tag{1.3.14}$$

where $x_2 - x_1, \ldots, x_m - x_1$ are linearly independent. We call the set A_2 a *hyperplane* in $\mathbb{R}^m$.

In the Cartesian plane $\mathbb{R}^2$, a line is given by equation (1.3.10). The equation on the right-hand side of equation (1.3.10) is written as $x_1 = \lambda x_{11} + (1 - \lambda)x_{12}$ and $x_2 = \lambda x_{21} + (1 - \lambda)x_{22}$. Cancelling λ from these equations, we obtain $(x_{21} - x_{22})x_1 + (x_{11} - x_{12})x_2 = (x_{21} - x_{22})x_{12} + (x_{11} - x_{12})x_{22}$. Thus, a line is alternatively written as $L_1 = \{(x_1, x_2) \mid a_1 x_1 + a_2 x_2 = b\}$, where a_1 and a_2 are constants, at least one of which is non-zero, and b is a constant. Similarly, in $\mathbb{R}^3$, a plane is alternatively written as $A_1 = \{(x_1, x_2, x_3 \mid a_1 x_1 + a_2 x_2 + a_3 x_3 = b\}$, where a_1, a_2 and a_3 are constants, at least one of which is non-zero. Extending this expression to $\mathbb{R}^m$, we can alternatively write the hyperplane A_2 as

$$A_2 = \{x \mid a^T x = b, a, x \in \mathbb{R}^m, \|a\| \neq 0\}. \tag{1.3.15}$$

A line splits a plane into two disjoint regions (Figure 1.3.8(a)); a plane splits a space into two disjoint regions (Figure 1.3.9); generally, a hyperplane splits a space into two disjoint regions. We call one of the regions with the hyperplane, i.e. the region defined by

$$H = \{x \mid a^T x \leq b, a, x \in \mathbb{R}^m, \|a\| \neq 0\} \tag{1.3.16}$$

a *(closed) half space* (in particular, a *(closed) half plane* for $m = 2$; Figure 1.3.10(a)).

Observing Figure 1.3.10(a), we readily understand that the line segment between any two points in a half plane is included in the half plane. This

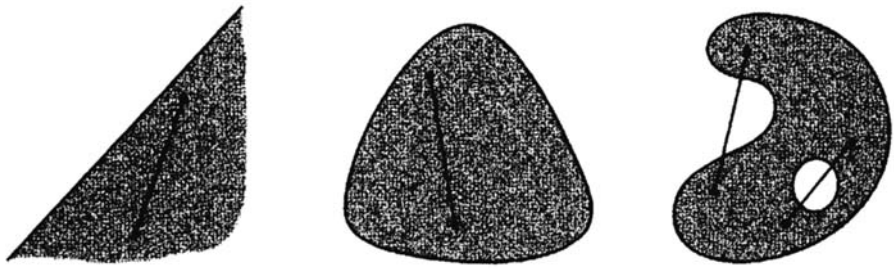


Figure 1.3.10 Convex and non-convex sets: (a) a half plane; (b) a convex set; (c) a non-convex set.

property holds not only for a half plane but also for other figures, such as Figure 1.3.10(b). Unlike Figure 1.3.10(c), Figure 1.3.10(b) does not have holes or its boundary does not bend into the figure. In such a case we say that a geometrical figure is *convex*. Formally, if for any two points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ in A the line segment joining the points is contained in A, i.e.

$$\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2 \in A \quad \text{for all } 0 \le \lambda \le 1, \tag{1.3.17}$$

then we call the set A a *convex set*; otherwise, a *non-convex set*. In particular, if for any two points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ in A the relation

$$\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2 \in A \backslash \partial A \quad \text{for all } 0 \le \lambda \le 1 \tag{1.3.18}$$

holds, we call the set A a *strictly convex set*.

A half space (equation (1.3.16)) is convex, because $\boldsymbol{a}^{\mathrm{T}}(\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2) = \lambda \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}_1 + (1-\lambda)\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}_2 \le \lambda b + (1-\lambda)b = b$ for $\boldsymbol{x}_1, \boldsymbol{x}_2 \in H$ and $0 < \lambda < 1$. The intersection of two convex sets is also convex. To prove it, let $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ be arbitrary points in the intersection of two convex sets A_1 and A_2. Then, $\boldsymbol{x}_1, \boldsymbol{x}_2 \in A_1$ and $\boldsymbol{x}_1, \boldsymbol{x}_2 \in A_2$. Since A_1 and A_2 are convex sets, from relation (1.3.17) the line $\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2$, $0 \le \lambda \le 1$, is in A_1 and in A_2. Therefore, the line $\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2$, $0 \le \lambda \le 1$, is in $A_1 \cap A_2$. Applying this proof, we can see that the intersection of convex sets is also convex. It is left as an exercise to show that a convex polygon A with vertices $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ is given by

$$A = \left\{ \boldsymbol{x} \,\middle|\, \sum_{i=1}^{n} \lambda_i \boldsymbol{x}_i \quad \text{where} \quad \sum_{i=1}^{n} \lambda_i = 1, \lambda_i \ge 0, i \in I_n \right\} \tag{1.3.19}$$

(recall the derivation leading to equation (1.3.13)).

A type of convex sets which is of particular interest is the so-called convex hull. To be explicit, consider the geometric figure in Figure 1.3.11(a) and suppose that this figure is encircled by a rubber band. Then the rubber band becomes like the heavy line in Figure 1.3.11(b). We fill the inside of this rubber band with points. Then we obtain Figure 1.3.11(c), which we call a *convex hull*. To give a formal definition of the convex hull of a set A, consider

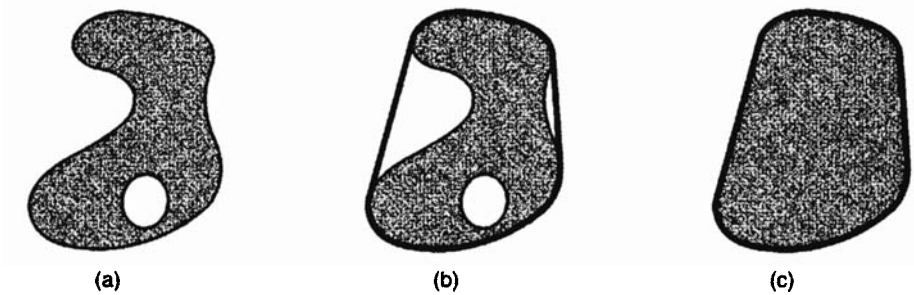


Figure 1.3.11 A convex hull: (a) a non-convex figure A; (b) the boundary of the convex hull of A (the heavy line); (c) the convex hull of A.

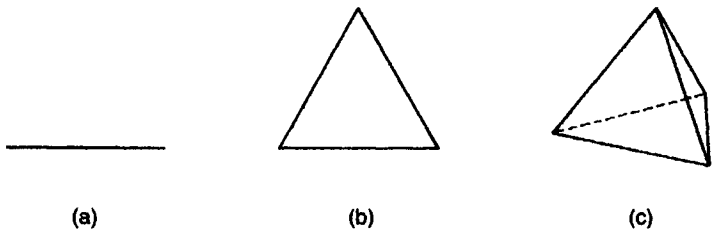


Figure 1.3.12 Simplices.

all possible convex sets that contain the set A. The number of such sets is infinite. We take the intersection of all such convex sets. Then the resulting set is called the *convex hull* of the set A. Since the intersection of convex sets is a convex set, the convex hull of the set A is a convex set. Moreover, it is the 'smallest' convex set that contains the set A. Obviously, if the set A is a convex set, then the convex hull of A is A itself.

As a special type of a convex hull, we define a simplex. A *simplex* in $\mathbb{R}^m$ is the convex hull of any set of $m+1$ points which do not all lie on one hyperplane in $\mathbb{R}^m$. If $m = 0$, the simplex is a point itself, called the *zeroth-order simplex*; if $m = 1$, the simplex is the straight line segment connecting two points, called the *first-order simplex* (Figure 1.3.12(a)); if $m = 2$, the simplex is a triangle, called the *second-order simplex* (Figure 1.3.12(b)); if $m = 3$, the simplex is a tetrahedron, called the *third-order simplex* (Figure 1.3.12(c)), and so forth. The first-order simplex is written as equation (1.3.12). Similarly, the second-order simplex is written as equation (1.3.19), where $n = 3$ and $\boldsymbol{x}_2 - \boldsymbol{x}_1$ and $\boldsymbol{x}_3 - \boldsymbol{x}_1$ are linearly independent. In general, the mth-order simplex is written as equation (1.3.19), where $\boldsymbol{x}_2 - \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{m+1} - \boldsymbol{x}_1$ are linearly independent. If a set is contained in a simplex, we say that the set is *bounded*; otherwise *unbounded*. We sometimes say that the region given by a bounded set is a *finite region* and the region given by an unbounded set is an *infinite region*.

Let A be a subset of $\mathbb{R}^m$. We call an element $\boldsymbol{a}_u$ in $\mathbb{R}^m$ an *upper bound* of A if $\boldsymbol{a}_u > \boldsymbol{x}$ for all $\boldsymbol{x}$ in A. Similarly, we call an element $\boldsymbol{a}_l$ in $\mathbb{R}^m$ a *lower bound* of A if $\boldsymbol{a}_l < \boldsymbol{x}$ for all $\boldsymbol{x}$ in A. When the set A has an upper bound, we say that A is *bounded above*. Similarly, if the set A has a lower bound, we say that A is *bounded below*. If the set A is bounded above in $\mathbb{R}^m$, we say that an upper bound of A is a *supremum* of A if it is less than any other upper bound of A. Similarly, we say that a lower bound of A is an *infimum* of A if it is greater than any other lower bound of A.

The concept of convexity can be introduced into a function. Let $f(\boldsymbol{x})$ be a function from $\mathbb{R}^n$ to $\mathbb{R}$, and $\{(\boldsymbol{x}, y) \mid y \geq f(\boldsymbol{x}), \boldsymbol{x} \in S\}$, where S is the domain of the function f. This set is called the *epigraph* of the function f. Two examples of the epigraph are shown in Figure 1.3.13. If the epigraph of the function f is convex, we call the function f a *convex function* (Figure 1.3.13(a)); otherwise, a *non-convex function* (Figure 1.3.13(b)).

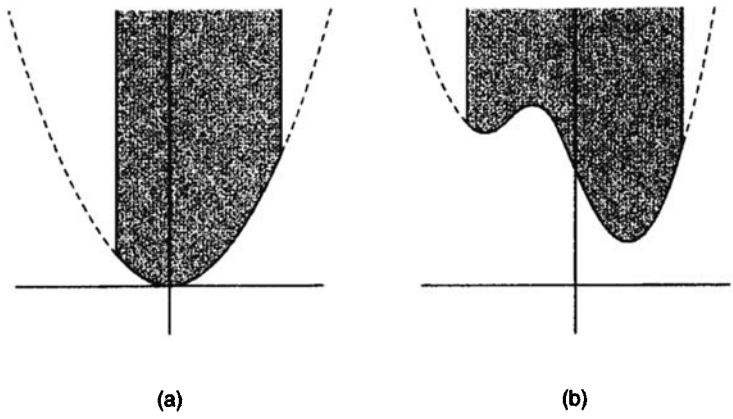


Figure 1.3.13 (a) The epigraph of a convex function, and (b) that of a non-convex function.

Suppose, as is shown in Figure 1.3.14, that five half planes, $H_1, \ldots, H_5$, are given. Using a finite number of intersection and union operations with respect to these half planes, we may construct a connected region. For example, $(H_1 \cap H_2 \cap H_3) \cup (H_1 \cap H_4 \cap H_5)$ in Figure 1.3.14(a). We call this region a *polygon*. Stated a little more generally, if a region constructed from a finite number of intersection and union operations with respect to a finite number of half planes is a non-empty connected set, then we call the region a *polygon*. The boundary of a polygon consists of straight line segments. These line segments are called *edges* and their end points are called *vertices*. If a polygon does not have a hole in it, and each vertex of a polygon has exactly two edges and every edge intersects only at vertices, we call the polygon a *simple polygon*. The polygon in Figure 1.3.14(a) is a simple polygon, but that in Figure 1.3.14(b) is not.

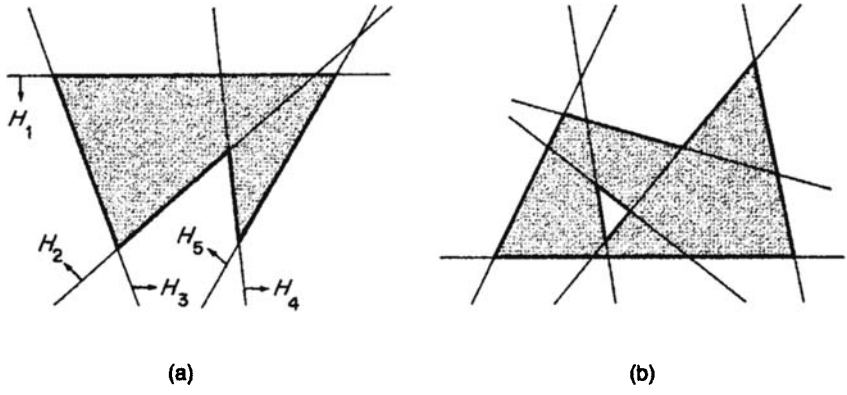


Figure 1.3.14 A polygon constructed from half planes: (a) a simple polygon; (b) a polygon that is not simple.

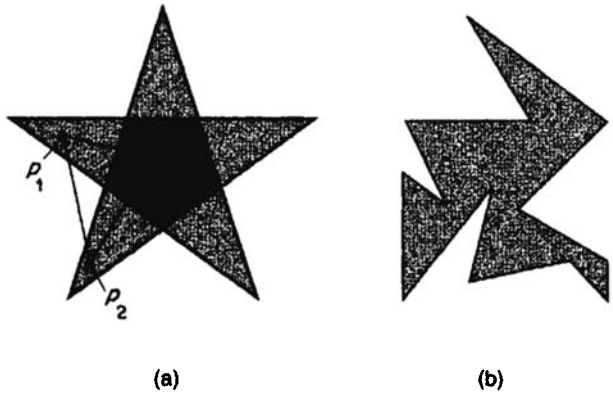


Figure 1.3.15 (a) A star-shaped polygon (the shaded region) and its kernel (the densely shaded region); (b) a polygon that is not star-shaped.

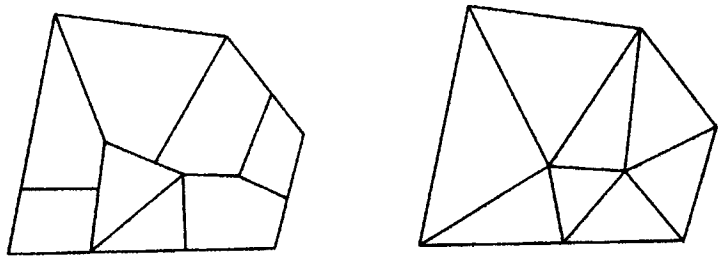
The notion of a polygon defined in $\mathbb{R}^2$ can be extended to $\mathbb{R}^m$. If a set constructed from a finite number of intersection and union operations on a finite number of half spaces is a non-empty connected bounded set, we call the set a *polyhedron*. A polyhedron (polygon) may be convex or non-convex. If a polyhedron is convex and bounded, it is sometimes called a *polytope*. A simplex (for example, the tetrahedron shown in Figure 1.3.12(c)) is a polytope.

The boundary of a polyhedron in $\mathbb{R}^3$ consists of polygons, which are called *faces*. Generally, the boundary of a polyhedron in $\mathbb{R}^m$ consists of polyhedrons in $\mathbb{R}^{m-1}$, which are called $(m-1)$-*faces*; the boundary of an $(m-1)$-face consists of polyhedrons in $\mathbb{R}^{m-2}$, which are called $(m-2)$-faces; and so forth. Note that 0-faces are *vertices*, and 1-faces are *edges*, and that $(m-1)$-faces are sometimes called *facets*.

Recalling the definition of a convex set (equation (1.3.17)), we notice that the sets shown in Figure 1.3.15 are non-convex, because the line segment connecting points p_1 and p_2 is not contained in the set. However, if we fix one end point at p_0, as in Figure 1.3.15(a), the line connecting p_0 and any point in the set is contained in the set. We call a set with this geometric property *star-shaped*. Stated precisely, the set A is *star-shaped* if there exists a point $\boldsymbol{x}_0$ in A for which $\lambda \boldsymbol{x}_0 + (1-\lambda)\boldsymbol{x} \in A$ holds for all $0 \le \lambda \le 1$ and for all $\boldsymbol{x} \in A$. The set of such points, $\{\boldsymbol{x} \mid \lambda \boldsymbol{x}_0 + (1-\lambda)\boldsymbol{x} \in A, 0 \le \lambda \le 1, \boldsymbol{x} \in A\}$, is called the *kernel* of the set A. The kernel of the star-shaped polygon of Figure 1.3.15(a) is shown by the densely shaded region. Obviously, a convex polygon is star-shaped, and its kernel is the convex polygon itself.

Let S be a closed subset of $\mathbb{R}^m$, S_i be a closed subset of S and $\mathcal{S} = \{S_1, \ldots, S_n\}$ (when we deal with an infinite n, we assume that only finitely many S_i hit a bounded subset of $\mathbb{R}^m$). If elements in the set $\mathcal{S}$ satisfy

$$[S_i \backslash \partial S_i] \cap [S_j \backslash \partial S_j] = \varnothing, \quad i \neq j, i, j \in I_n, \tag{1.3.20}$$



(a) (b)

Figure 1.3.16 Tessellations: (a) a tessellation that is not a triangulation; (b) a triangulation.

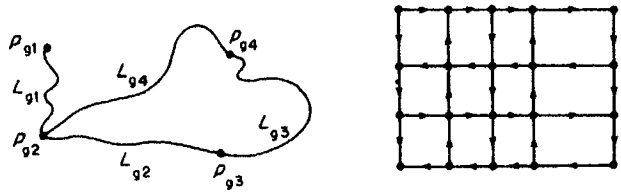
and

$$\bigcup_{i=1}^{n} S_i = S_1 \cup \cdots \cup S_n = S, \tag{1.3.21}$$

then we call the set $\mathcal{S}$ a *tessellation* of S. Specifically, we call a tessellation in $\mathbb{R}^2$ a *planar tessellation* of S. Figure 1.3.16 shows two planar tessellations. The tessellation in panel (a) consists of polygons with three or more vertices, whereas that in panel (b) consists of only triangles. We call a planar tessellation like panel (b), i.e. a tessellation in which S_i in $\mathcal{S}$ is a triangle for all $i \in I_n$, a *triangulation* of S.

1.3.2 Graphs

Suppose, as is shown in Figure 1.3.17(a), that there are four towns $p_{g1}, \ldots, p_{g4}$ (indicated by the filled circles) in a region and those towns are connected by roads $L_{g1}, \ldots, L_{g4}$ (indicated by line segments). We are living in the town p_{g1} and wish to visit towns p_{g3} and p_{g4} in this order without going through the other towns, and return to the town p_{g1}. To examine this type of trip feasibility, the distance between towns or the shape of roads are not essential; the essential information is whether or not there exists a direct road



(a) (b)

Figure 1.3.17 (a) Towns and roads between towns, and (b) one-way streets, which can be regarded as geometric graphs.

Table 1.3.1 Roads and their incident towns

Roads L_{gi}	Towns incident with road L_{gi}
L_{g1}	p_{g1}, p_{g2}
L_{g2}	p_{g2}, p_{g3}
L_{g3}	p_{g3}, p_{g4}
L_{g4}	p_{g4}, p_{g2}

between towns. This information is summarized in Table 1.3.1. Mathematically, we can treat the incidence of roads with towns listed in Table 1.3.1 using graph theory, which is to be introduced here.

To define a graph formally, let $P = \{p_1, \ldots, p_{n_p}\}$ be a non-empty set, and $\{p_i, p_j\}$ be an unordered pair, where $p_i, p_j \in P$. The set of all distinct unordered pairs $\{\{p_i, p_j\} \mid p_i, p_j \in P\}$ is called the *unordered product* of P, and denoted by $P * P$ (note that $\{p_i, p_i\}$ is allowed). Let $L = \{L_1, \ldots, L_{n_l}\}$ be a set which is possibly empty, and f be a mapping of L into $P * P$, i.e. $\{p_j, p_k\} = f(L_i)$. Just for notational simplicity, we write $L_i \sim \{p_j, p_k\}$ for $\{p_j, p_k\} = f(L_i)$. We call the paired sets P and L with the mapping f an *(undirected) abstract graph*, or simply a(n) *(undirected) graph*, and denote it by $G(P, L, f)$, $G(P, L)$ or G. To distinguish a geometric graph to be defined below, we sometimes use G_a for an abstract graph. An example of an abstract graph is given by

$$G_a(P, L, f) = G_a(\{p_1, p_2, p_3, p_4\}, \{L_1, L_2, L_3, L_4\}, \{L_1 \sim \{p_1, p_2\}, L_2 \sim \{p_2, p_3\}, L_3 \sim \{p_3, p_4\}, L_4 \sim \{p_2, p_4\}\}). \tag{1.3.22}$$

If the sets P and L in G are finite, we call the graph a *finite graph*. In this text, we shall deal with only finite graphs.

Given a graph $G(P, L, f)$, we can construct a graph, called a *subgraph*, consisting of selected links and nodes of G with the same incidences as those of G satisfying that the selected nodes include all of the end points of the selected links. Mathematically, we call $G(P_s, L_s, f_s)$ a *subgraph* of G if and only if $G(P_s, L_s, f_s)$ satisfies the following three conditions: (i) $P_s \subset P$ and $L_s \subset L$; (ii) $f(L_i) = f_s(L_i)$ for every $L_i \in L_s$; (iii) if $L_i \in L_s$ and $\{p_j, p_k\} = f(L_i)$, then $p_j \in P_s$ and $p_k \in P_s$.

Inspecting the similarity between the function f of equation (1.3.22) and Figure 1.3.17(a) or Table 1.3.1 (which actually shows a function), we notice that the above abstract graph may be represented by a set of geometric points, and a set of geometric line segments connecting those points. For this representation, let us introduce the notion of a geometric graph.

Let $P_g = \{p_{g1}, \ldots, p_{gn_p}\}$ be a set of n_p distinct points in the m-dimensional Euclidean space and $L_g = \{L_{g1}, \ldots, L_{gn_l}\}$ be a set of n_l non-self-intersecting line segments with end points in P_g. The line segments in L_g may be straight or may be curved, but they should satisfy the following three properties.

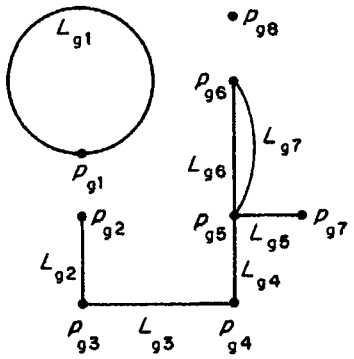


Figure 1.3.18 A geometric graph $G_g = G_g(\{p_{g1}, \ldots, p_{g8}\}, \{L_{g1}, \ldots, L_{g7}\})$.

(i) Every line segment that forms a loop (a non-self-intersecting line segment whose end points coincide) contains exactly one point of P_g (for example, a loop with p_{g1} in Figure 1.3.18).

(ii) Every line segment that does not form a loop contains exactly two points of P_g, and these points agree with its end points (for example, L_{g2} with p_{g2} and p_{g3} in Figure 1.3.18).

(iii) Every line segment in L_g has no common points except for points of P_g (for example, L_{g2} and L_{g3} share p_{g3} in Figure 1.3.18).

We call the paired sets (P_g, L_g) satisfying the above properties a *geometric graph* and denote it by $G(P_g, L_g)$ or G. In particular, when L_g is a set of open straight line segments, $G(P_g, L_g)$ is called a *planar straight-line graph* and abbreviated to PSLG (Preparata and Shamos, 1985). When we distinguish a geometric graph from an abstract graph G_a, we sometimes use G_g. As is noticed from this definition, the end points of every line segment in L_g are included in P_g, but a point in P_g is not necessarily an end point of a line segment in L_g (such as p_{g8} in Figure 1.3.18).

From the definition of a geometric graph, we notice that the geometric points and the geometric line segments shown in Figure 1.3.16(a) actually form a geometric graph, i.e.

$$G_g(P_g, L_g, f_g) = G_g(\{p_{g1}, p_{g2}, p_{g3}, p_{g4}\}, \{L_{g1}, L_{g2}, L_{g3}, L_{g4}\}, \{L_{g1} \sim \{p_{g1}, p_{g2}\}, L_{g2} \sim \{p_{g2}, p_{g3}\}, L_{g3} \sim \{p_{g3}, p_{g4}\}, L_{g4} \sim \{p_{g2}, p_{g4}\}\}). \quad (1.3.23)$$

Moreover, we notice that there is a one-to-one correspondence between the vertices in the abstract graph G_a of equation (1.3.22) and those in the geometric graph G_g of equation (1.3.23) such that the adjacency relations are preserved.

To state this correspondence more precisely, let us consider two graphs $G(P_\alpha, L_\alpha, f_\alpha)$ and $G(P_\beta, L_\beta, f_\beta)$. First, when $G(P_\alpha, L_\alpha, f_\alpha)$ is an abstract graph and $G(P_\beta, L_\beta, f_\beta)$ is an abstract graph or a geometric graph, if there exist a one-to-one mapping φ_p of P_α onto P_β, and a one-to-one mapping φ_l of L_α onto L_β such that $\{p_j, p_k\} = f_\alpha(L_i)$, $L_i \in L_\alpha, p_j, p_k \in P_\alpha$, if and only if $\{\varphi_p(p_j), \varphi_p(p_k)\} = f_\beta(\varphi_l(L_i))$, then we say that the graphs $G(P_\alpha, L_\alpha, f_\alpha)$ and

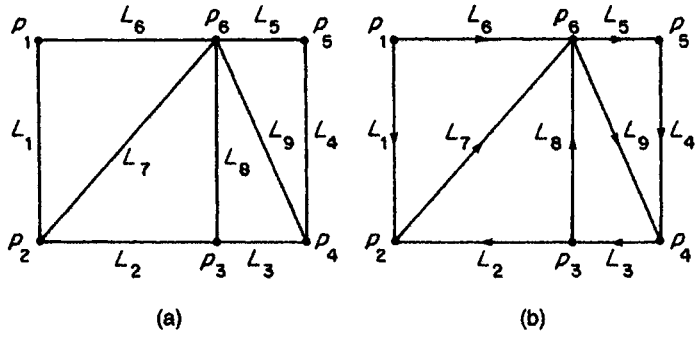


Figure 1.3.19 (a) An undirected graph and (b) a directed graph.

$G(P_\beta, L_\beta, f_\beta)$ are *isomorphic*. If an abstract graph G_a is isomorphic to a geometric graph G_g, then we say that graph G_g is a *geometric realization* of the abstract graph G_a. If a graph has a geometrical realization in $\mathbb{R}^2$, we call the graph a *planar graph*; otherwise, a *non-planar graph*. Although graphs are not always planar graphs, every finite graph has a geometric realization in $\mathbb{R}^3$. Comparing equations (1.3.22) and (1.3.23), we readily notice that the abstract graph G_a and the geometric graph G_g are isomorphic, and hence the geometric graph G_g is a geometric realization of the abstract graph G_a. Moreover, since the geometric graph G_g can be drawn on a plane (Figure 1.3.17(a)), it is a planar graph.

Second, when two graphs $G_g(P_\alpha, L_\alpha, f_\alpha)$ and $G_g(P_\beta, L_\beta, f_\beta)$ are both geometric graphs, the notion of isomorphic becomes a little stronger than the above. If $G_g(P_\beta, L_\beta, f_\beta)$ is obtained by deforming $G_g(P_\alpha, L_\alpha, f_\alpha)$ by moving the vertices continuously in $\mathbb{R}^2$ in such a way that the edges are also moved as their end points are moved (the edges may be extended or shrunken) and that no edges hit a vertex other than its own end points, then we say that graphs $G_g(P_\alpha, L_\alpha, f_\alpha)$ and $G_g(P_\beta, L_\beta, f_\beta)$ are *isomorphic*.

In the example of Figure 1.3.17(a) we implicitly assume that every road is two-way; we can drive, say, from p_{g1} to p_{g2} as well as from p_{g2} to p_{g1}. In the downtown of a city, however, streets are often one-way, such as in Figure 1.3.17(b). To take this directional feature into account in a graph, we introduce a direction on links in L.

Let $P = \{p_1, \ldots, p_{n_p}\}$ be a non-empty set, (p_i, p_j) be an ordered pair, where $p_i, p_j \in P$ (cf. $\{p_i, p_j\}$), $L = \{L_1, \ldots, L_{n_l}\}$ be a set which is possibly empty. The set of all distinct ordered pairs $\{(p_i, p_j) \mid p_i, p_j \in P\}$ is called the *ordered product* of P, and denoted by $P \times P$. Let f be a mapping of L into $P \times P$, i.e. $(p_i, p_j) = f(L_k)$, which is simply written as $L_k \sim (p_i, p_j)$. We call the paired sets P and L with mapping f a *directed graph*, which should be distinguished from the undirected graph defined above. The graphs in Figures 1.3.19(a) and 1.3.19(b) are an undirected graph and a directed graph, respectively.

The elements of P and L in G are called *nodes* and *links*, respectively, and f is called an *incidence mapping*. In particular, in a directed graph, p_j and p_k

in $L_i \sim (p_j, p_k)$ are called an *initial node* and a *terminal node*, respectively. For example, in the directed graph in Figure 1.3.19(b), p_1 is the initial node of L_1, and p_2 is the terminal node of L_1. If $L_i \sim \{p_j, p_k\}$ (for an undirected graph) or $L_j \sim (p_j, p_k)$ (for a directed graph), then L_i is said to be *incident* with each of p_j and p_k. In particular, in a directed graph, L_i is said to be *positively incident* with p_j and *negatively incident* with p_k. The nodes incident with a link are called *end points*. For $L_i \sim \{p_j, p_k\}$ or $L_i \sim (p_j, p_k)$, if $p_j = p_k$, then L_i is called a *loop* (L_{g1} in Figure 1.3.18 is a loop). If $L_i \sim \{p_j, p_k\}$ or $L_i \sim (p_j, p_k)$ and $L_l \sim \{p_j, p_k\}$ or $L_l \sim (p_j, p_k)$, then L_i and L_l are called *parallel links* (L_{g6} and L_{g7} are parallel links in Figure 1.3.18).

If $L_i \sim \{p_j, p_k\}$ or $L_i \sim (p_j, p_k)$, then p_j and p_k are called *adjacent nodes* (for example, p_2 and p_3 in Figure 1.3.19(a),(b) are adjacent nodes). If L_i and L_j share at least one common end point, then L_i and L_j are called *adjacent links* (for example, L_2 and L_3 in Figure 1.3.19(a),(b) are adjacent links). The number of links incident with p_i is called the *degree* of p_i, and denoted by $\delta(p_i)$ (for example, $\delta(p_2) = 3$ in Figure 1.3.19(a),(b)). In particular, in a directed graph, the number of positively incident links with p_i, denoted by $\delta^+(p_i)$, is called the *positive degree* of p_i, and the number of negatively incident links with p_i, denoted by $\delta^-(p_i)$, is called the *negative degree* of p_i (for example, $\delta^+(p_2) = 1$ and $\delta^-(p_2) = 2$ in Figure 1.3.19(b)). Obviously, the equation $\delta(p_i) = \delta^+(p_i) + \delta^-(p_i)$ holds. A node in P whose degree is zero is called an *isolated node* (for example, p_{g8} in Figure 1.3.18).

The adjacency of nodes may be indicated in terms of a matrix, called an *adjacency matrix*. To be explicit, consider a graph $G(\{p_1, \ldots, p_n\}, L)$ and let C be an $n \times n$ matrix (n columns and n rows),

$$C = \begin{bmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{n1} & \cdots & c_{nn} \end{bmatrix}, \tag{1.3.24}$$

which is sometimes written as $[c_{ij}]$ for short. When G is an undirected graph, c_{ij} denotes the number of links incident with both nodes p_i and p_j. When G is a directed graph, let c_{ij} be the number of links directed from a node p_i to a node p_j. Then, the adjacency of nodes in graph G can be indicated by the matrix C, which we call an *adjacency matrix* of nodes. The adjacency matrices C_1 of the undirected graph shown in Figure 1.3.19(a) and C_2 of the directed graph shown in Figure 1.3.19(b) are given by

$$C_1 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}, \quad C_2 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix},$$

respectively.

From the definition of c_{ij} and that of $\delta(p_i)$, it follows that the equation $\delta(p_i) = \sum_{j=1}^{n} c_{ij}$ holds for an undirected graph, and the equations $\delta^+(p_i) = \sum_{j=1}^{n} c_{ij}$ and $\delta^-(p_i) = \sum_{j=1}^{n} c_{ji}$ hold for a directed graph.

For a given graph, if we can traverse from node to node on successively adjacent links without visiting the same links more than once, we call such a set of links a *chain* in an undirected graph, and a *path* in a directed graph (note that in a directed graph, we are supposed to follow directions). For example, L_7, L_9, L_4, L_5, L_6 in Figure 1.3.19(a) is a chain, and L_7, L_9, L_3, L_8, L_5 in Figure 1.3.19(b) is a path. In traversing a chain or a path, if we visit every node only once, the chain or the path is called a *simple chain* or a *simple path*. The chain or the path in the above example is not a simple chain nor a simple path, because we visit p_6 twice. An example of a simple chain in Figure 1.3.19(a) is L_7, L_9, L_3, and that of a simple path in Figure 1.3.19(b) is L_7, L_9, L_3. Note that we allow a chain or a path that returns to the same node, for example $L_1, L_7, L_9, L_4, L_5, L_6$ in Figure 1.3.19(a), and L_7, L_9, L_3, L_2 in Figure 1.3.19(b). Such a chain and a path are called a *circuit* and a *cycle*, respectively. If we can traverse a circuit and a cycle visiting every node only once, we call them a *simple circuit* and a *simple cycle*, respectively. The circuit in the above example is not a simple circuit because we visit p_6 twice, but the cycle in the above example is a simple cycle.

In the graph of Figure 1.3.19(a) we can traverse from any node in P to any node in P through links. In this sense, the graph is connected. Stated a little more precisely, an undirected graph $G(P, L)$ is *connected* if there exists at least one chain between every pair of nodes in P; the graph G is *disconnected* if there exists no chain between at least one pair of nodes in P. For example, the graph in Figure 1.3.19(a) is connected; the graph in Figure 1.3.18 is disconnected. In a directed graph $G(P, L)$, if there exists at least one path from any node to any node in P, we say that the graph is *strongly connected*. The directed graph in Figure 1.3.19(b) is not strongly connected because there is no path from node p_2 to node p_1. A connected subgraph which is not contained in any larger connected subgraph is called a *component*. The graph in Figure 1.3.18(a) consists of three components.

The graph shown in Figure 1.3.20 is a connected graph, which looks like a tree (we shall often show a tree upside down). Obviously, an actual tree does not have looped twigs. In graph theory, if a graph is connected and does not have circuits, the graph is called a *tree (graph)*. When a graph is a tree, the following equation holds:

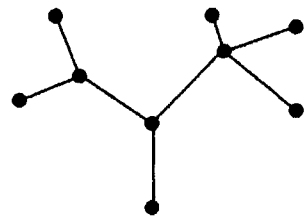


Figure 1.3.20 A tree graph

$$n_p - n_l = 1, \tag{1.3.25}$$

where n_p is the number of nodes and n_l the number of links in a tree.

When a graph is not a tree, equation (1.3.25) does not necessarily hold, but we have a more general equation, called Euler's formula. When a graph is a planar graph, links of the graph may partition the plane into disjoint regions one of which is unbounded. Let n_r be the number of those regions. Note that when a graph is a tree, the links of the tree do not partition the plane and hence we have only one (unbounded) region ($n_r = 1$). In terms of n_p, n_l and n_r, *Euler's formula* is stated as follows.

Euler's formula For a connected planar graph with n_p nodes, n_l links and n_r regions, the equation

$$n_p - n_l + n_r = 2 \tag{1.3.26}$$

holds.

Euler's formula can be generalized in $\mathbb{R}^m$.

The Euler–Poincaré formula (the Euler–Schlaefli formula) Let $\mathcal{S}$ be a tessellation of a bounded set S in $\mathbb{R}^m$, and n_i be the number of i-faces in $\mathcal{S}$. Then the equation

$$\sum_{i=0}^{m} (-1)^i n_i = 1 + (-1)^m \tag{1.3.27}$$

holds.

We call this equation the *Euler–Poincaré formula* or the *Euler–Schlaefli formula.*

A planar graph is always associated with a 'dual graph'. To define it explicitly, let $G_g(P_g, L_g)$ be a planar graph with k regions $R_i, i \in I_k$. We take a point q_i in each region R_i (the unfilled circles in Figure 1.3.21(a)) and let $P_g^* = \{q_i, i \in I_k\}$. For each link which is shared by regions R_i and R_j, we generate a line segment combining q_i and q_j which crosses the common link only once and has no point in common with any other links of L (the broken

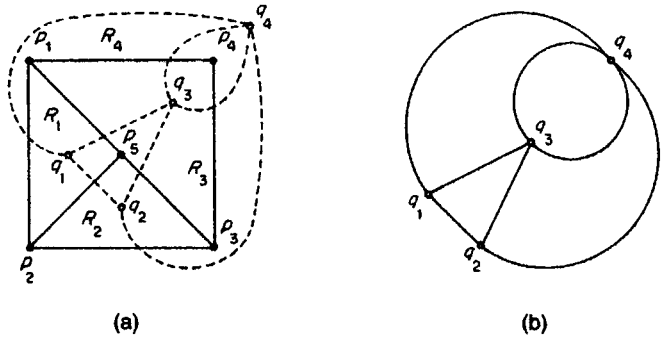


Figure 1.3.21 The primal and its dual graph.

line in Figure 1.3.21(a)). Let L_g^* be the set of resulting links. Then we obtain the new geometric graph G_g^* with P_g^* and L_g^* (Figure 1.3.21(b)). We call $G_g(P_g^*, L_g^*)$ the *dual graph* of $G_g(P_g, L_g)$. For the dual graph G_g^*, G_g is called the *primal graph*. Note that the dual graph and the primal graph are a relative notion. We can say that the dual graph of G_g^* is G_g. In this case, the primal graph is G_g^*.

1.3.3 Spatial stochastic point processes

A *stochastic point process* is a probabilistic generation of points in a space according to a probability distribution function defined over the space. From the term 'process', the reader might call to mind a process over time. In the study of *spatial stochastic point processes*, however, a time process is usually implicit. When we explicitly deal with a spatial stochastic point process over time, we call it a *spatial–temporal stochastic point process*.

Let f be a probability density function defined over S. The meaning of $f(\boldsymbol{x})$ is that the probability of a point being placed in a very small volume (area), ΔS, around $\boldsymbol{x}$ is given by $f(\boldsymbol{x})\,\Delta S$. An example of f is the *uniform distribution* given by

$$f(\boldsymbol{x}) = \begin{cases} \dfrac{1}{|S|}, & \boldsymbol{x} \in S, \\ 0, & \boldsymbol{x} \notin S, \end{cases} \tag{1.3.28}$$

where $S \subset \mathbb{R}^m$ and $|S|$ is the area or volume of S ($|S| < \infty$). We call the process in which points are generated over S according to the uniform distribution a *binomial point process*. Figure 1.3.22 shows the binomial point process with $n = 50$ points in a unit square.

The reader may be curious to know why the above point process is called the binomial point process. To see the reason, let A be a subset of S, and $N(A)$ be the number of points placed in A when n points are generated according to the uniform distribution. Since the probability of a point being placed in A is given by $|A|/|S|$, the probability, $\Pr(N(A) = x)$, that x points are placed in A is given by

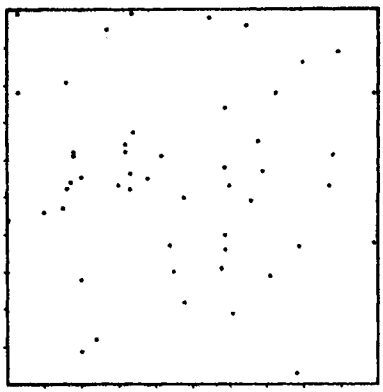


Figure 1.3.22 A binomial point process ($n = 50$).

$$\Pr\left(N(A)=x\right)=\binom{n}{x}\left(\frac{|A|}{|S|}\right)^{x}\left(1-\frac{|A|}{|S|}\right)^{n-x}. \tag{1.3.29}$$

This distribution is the *binomial distribution* with the parameter values n and $\theta = |A|/|S|$, after which the above process is named.

In the above we defined the binomial point process by the uniform distribution. Alternatively, we can directly define the binomial point process by the binomial distribution. The *binomial point process* is defined by the process in which, for any subset A in S, the probability distribution of $N(A)$ is given by equation (1.3.29).

In a similar way, we can define a general stochastic point process. Let S be a non-empty subset of $\mathbb{R}^m$, and $N(A)$ be the number of points in a subset, A, of S. Then a *stochastic point process* is defined by the process in which points are generated according to a probability distribution $\Pr(N(A) = x)$, $x = 0, 1, \ldots$, for any A in S. Specifically, if $m = 2$, the process is called a *planar stochastic point process*. If the probability of two points being coincident is zero under a stochastic point process, we say that the process is *simple*. If any finite region A in $\mathbb{R}^m$ contains a finite number of points under a stochastic point process, the process is said to be *locally finite*. In this text, we deal with stochastic point processes that satisfy these two conditions.

Among many simple and locally finite stochastic point processes, the most fundamental process is the *Poisson point process*, which is defined by the process in which for any subset A in $S = \mathbb{R}^m$, $\Pr(N(A) = x)$ is given by

$$\Pr\left(N(A)=x\right)=\frac{\lambda|A|\,e^{-\lambda|A|}}{x!},\quad x=0,1,2,\ldots. \tag{1.3.30}$$

We can also define the Poisson point process as a limit of the binomial point process. The limit means the expansion of the finite region S to an infinite region keeping $n/|S|$ constant. Since $\lambda = n/|S|$ is the density of points, this expansion implies the expansion of the region S keeping the density of points constant. In the limit, the binomial distribution of equation (1.3.29) approaches the Poisson distribution of equation (1.3.30). We understand from this derivation that the parameter λ in the Poisson point process shows the density of points in a unit volume. The density λ is sometimes referred to as *intensity*.

In the above, we derived the Poisson point process directly from the Poisson distribution. More fundamentally, we can derive the Poisson point process from the following four postulates, through which we can see the characteristics of the Poisson point process explicitly.

Postulate PO1 $0 < \Pr(N(A) = 0) < 1$ if $0 < |A| < \infty$.

Postulate PO2 The probability distribution of $N(A)$ depends on $|A|$ and $\Pr(N(A) \geq 1)$ approaches zero as $|A|$ approaches 0.

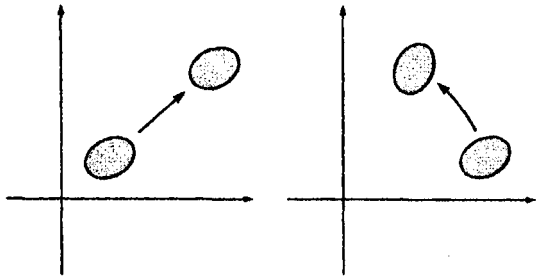


Figure 1.3.23 (a) Translation; (b) rotation.

Postulate PO3 If $A_1, \ldots, A_n$ are disjoint, then $N(A_1), \ldots, N(A_n)$ are mutually independent and $N(A_1 \cup \ldots \cup A_n) = N(A_1) + \ldots + N(A_n)$.

Postulate PO4

$$\lim_{|A| \to 0} \frac{\Pr(N(A) \geq 1)}{\Pr(N(A) = 1)} = 1. \tag{1.3.31}$$

Postulate PO1 asserts that if a region A has a positive volume, then the probability of a point being placed in the region A is not zero. Postulate PO2 asserts that $\Pr(N(A) = x)$ does not depend upon the location of A but only upon its volume (or its area for $m = 2$). There is no tendency for points to occupy particular regions in $\mathbb{R}^m$. Postulate PO3 means no interactions between points. We call this property *independent scattering*. It should be noted that the binomial point process is not independent scattering, because if x points are placed in A, then $n-x$ points are inevitably placed in $S \setminus A$. Postulate PO4 implies that the process is simple. If the random variable $N(A)$ satisfies the above four postulates for $A \subset \mathbb{R}^m$, then $N(A)$ is proved to follow the distribution of equation (1.3.30). Since this proof requires much space, we omit it here; the reader should consult a textbook on stochastic processes, for example Karlin (1969, pp. 337–43).

As is seen in Postulate PO2, the Poisson point process has the invariance property in the sense that $\Pr(N(A_1) = x) = \Pr(N(A_2) = x)$ for $|A_1| = |A_2|$. Generally, we can examine the characteristics of stochastic point processes in terms of invariance properties. First, let us consider the transformation of A in S defined by

$$\varphi_c(A) = \{x + c \mid x \in A\}, \quad A \subset S, \varphi_c(A) \subset S, \tag{1.3.32}$$

where $c \in \mathbb{R}^m$ is a constant vector. If we regard A as a figure in $\mathbb{R}^m$, we may say that $\varphi_c(A)$ moves the figure A without rotation in $\mathbb{R}^m$ (see Figure 1.3.23(a), $m = 2$). We call such a transformation a *translation*. In observing equation (1.3.30), we readily notice that $\Pr(N(A) = x) = \Pr(N(\varphi_c(A)) = x)$ holds; that is, the Poisson point process is invariant under translation.

Generally if a stochastic point process satisfies

$$\Pr\left(N(A_1)=x_1,\ldots,N(A_k)=x_k\right)=\Pr\left(N(\varphi_c(A_1))=x_1,\ldots,N(\varphi_c(A_k)=x_k)\right), \tag{1.3.33}$$

$$A_i\subset S,\ \varphi_c(A_i)\subset S,\quad i=1,\ldots,k,\ k=1,2,\ldots,$$

we say that the stochastic point process is *homogeneous* or *stationary*. The binomial point process and the Poisson point process are homogeneous. This may be intuitively understood from the fact that the processes are derived from the uniform distribution.

We next consider the transformation of A in S defined by

$$\varphi_\theta(A)=\{Bx\mid x\in A,\ |B|=1,\ B^{\mathrm{T}}B=1\},\quad A\subset S,\ \varphi_\theta(A)\subset S, \tag{1.3.34}$$

where B is an $m\times m$ matrix and $|B|$ is its determinant. The transformation φ_θ rotates the figure A around the origin in $\mathbb{R}^m$ as is shown in Figure 1.3.23(b) ($m=2$). We call such a transformation a *rotation*. Since $|A|=|\varphi_\theta(A)|$ holds, $\Pr(N(A)=x)=\Pr(N(\varphi_\theta(A))=x)$ holds for equation (1.3.30). Generally, if a stochastic point process satisfies

$$\Pr\left(N(A_1)=x_1,\ldots,N(A_k)=x_k\right)=\Pr\left(N(\varphi_s(A_1))=x_1,\ldots,N(\varphi_s(A_k)=x_k)\right), \tag{1.3.35}$$

$$A_i\subset S,\ \varphi_s(A_i)\subset S,\quad i=1,\ldots,k,\ k=1,2,\ldots,$$

we say that the stochastic point process is *isotropic*. If a stochastic point process is homogeneous and isotropic, we call the stochastic point process *motion-invariant*. The binomial point process and the Poisson point process are motion-invariant.

The binomial point process and the Poisson point process have a stronger property than the motion-invariant property. If a transformation, φ_v, of A in S satisfies

$$|\varphi_v(A)|=|A|,\quad A\subset S,\ \varphi_v(A)\subset S, \tag{1.3.36}$$

we call φ_v the *volume-preserving transformation* (*area-preserving transformation* for $m=2$). Obviously, translation and rotation are volume-preserving transformations. Generally, if a stochastic point process satisfies

$$\Pr\left(N(A_1)=x_1,\ldots,N(A_k)=x_k\right)=\Pr\left(N(\varphi_v(A_1))=x_1,\ldots,N(\varphi_v(A_k)=x_k)\right), \tag{1.3.37}$$

$$A_i\subset S,\ \varphi_v(A_i)\subset S,\quad i=1,\ldots,k,\ k=1,2,\ldots,$$

we say that the stochastic point process is *volume-preserving*. From equation (1.3.29), we notice that the binomial point process is volume-preserving. The Poisson point process is also volume-preserving because Postulate PO2 is satisfied in that process.

The volume-preserving transformation can be extended to a measure-preserving transformation. To be explicit, consider a set S and a set $\mathcal{A}$ of

subsets of S which satisfy the following three conditions: (i) $S \in \mathcal{A}$; (ii) if $A_i \in \mathcal{A}$, then $S \setminus A_i \in \mathcal{A}$; (iii) if $A_i \in \mathcal{A}$, $i = 1, 2, \ldots$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$. We call the pair $(S, \mathcal{A})$ satisfying these conditions a *measurable space*. If a function, m, from $\mathcal{A}$ to a non-negative real number satisfies $m(\emptyset) = 0$ and $m(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} m(A_i)$, where $A_1, A_2, \ldots$ are mutually disjoint, we say that m is a *measure* on $(S, \mathcal{A})$. A measurable space $(S,\mathcal{A})$ with a measure m is called a *measure space* and is denoted by a triple $(S, \mathcal{A}, m)$. In particular, if $m(S) = 1$, we call $(S, \mathcal{A}, m)$ a *probability space*. If a transformation φ satisfies $\varphi^{-1}(A_i) \in \mathcal{A}$ for $A_i \in \mathcal{A}$, we say that φ is *measurable* (φ^{-1} is the inverse function of φ). If φ is one-to-one, if $\varphi(S) = S$ and if $A_i \in \mathcal{A}$ implies $\varphi^{-1}(A_i) \in \mathcal{A}$, then we say that φ is *invertible*. For a measurable transformation φ in $(S, \mathcal{A}, m)$, if

$$m(\varphi^{-1}(A)) = m(A), \quad A \in \mathcal{A}, \tag{1.3.38}$$

we say that φ is a *measure-preserving transformation* and a point process which satisfies equation (1.3.37) for all measure-preserving transformations is a *measure-preserving stochastic point process*; if φ is invertible, equation (1.3.38) is equivalent to $m(\varphi(A)) = m(A)$. In the following exposition, we assume, for simplicity, that a measure-preserving transformation is invertible. Obviously the volume-preserving transformation is a measure-preserving transformation. Thus the binomial point process and the Poisson point process are measure-preserving point processes.

To sum up, the Poisson point process is simple, independent scattering and invariant under the measure-preserving transformation; consequently it is motion-invariant, i.e. it is homogeneous and isotropic. Because of these properties, we sometimes say that a pattern of points resulting from the Poisson point process is a *complete spatial random pattern* or *complete spatial randomness*, and abbreviated to CSR (Diggle, 1983; note that Daley and Vere-Jones, 1988, refer to independent scattering as complete randomness).

We next introduce another important property of the Poisson point process, called *ergodicity*, which plays an important role in deriving the properties of the Poisson Voronoi diagram in Chapter 5. In the above discussion we regard a point process as a counting process (i.e. counting the number $N(A)$ of points in a region A). Alternatively we may regard a point process as a 'random set' of points. To be explicit, we consider the collection, $\mathbb{M}$, of all simple and locally finite sets of points in $\mathbb{R}^m$, and a collection, $\mathcal{M}$, of subsets of $\mathbb{M}$ that satisfies (i), (ii) and (iii) above. Namely we consider a measurable space, $(\mathbb{M}, \mathcal{M})$. Let U be an element of $\mathcal{M}$ (i.e. a set of sets of points), and Θ be a random variable of which each realization is an element of $\mathbb{M}$ (i.e. a set of points). We call such a set of points Θ a *random set* of points. Let $P(U)$ be the probability of a random set Θ being an element of U, i.e.

$$P(U) = \Pr(\Theta \in U), \quad U \in \mathcal{M}. \tag{1.3.39}$$

Then $(\mathbb{M}, \mathcal{M})$ with P forms a measure space, or more specifically a probability space, $(\mathbb{M}, \mathcal{M}, P)$. In this context we may regard a *stochastic point*

process as a random set Θ determined by the probability distribution P over $\mathcal{M}$.

Let φ_x be the translation transformation, i.e.

$$\varphi_x(U) = \{x_1 + x, x_2 + x, \ldots \mid \{x_1, x_2, \ldots\} \in U\}, \quad U \in \mathcal{M}, \tag{1.3.40}$$

where $x \in \mathbb{R}^m$. The set $\varphi_x(U)$ implies the set of sets of points that are translated by x. For the sets $\varphi_x(U)$ and U, if the following equation holds,

$$P(U) = P(\varphi_x(U)) \quad \text{for any } x \in \mathbb{R}^m \text{ and } U \in \mathcal{M}, \tag{1.3.41}$$

we say that a point process Θ is *stationary*. This definition is found to be equivalent to the definition given by equation (1.3.33) above.

Let $[-a, a]^m$ be a hyper-cube in $\mathbb{R}^m$ whose edge is given by $[-a, a]$. For any sets U and W in $\mathcal{M}$, if the following relation holds,

$$\frac{1}{2^m a^m} \int_{[-a,a]^m} P(U \cap \varphi_x(W))\, dx - P(U)P(W) \to 0 \quad \text{as } a \to 0, \tag{1.3.42}$$

we say that a stationary point process Θ determined by the distribution P is *ergodic*. A homogeneous Poisson point process, Θ_P, is ergodic.

To give an alternative definition of ergodicity, we define a set U, called an *invariant set*, as

$$P(U \setminus \varphi_x(U) \cup \varphi_x(U) \setminus U) = 0 \quad \text{for all } x \in \mathbb{R}^m. \tag{1.3.43}$$

In this term, we may alternatively say that a stationary point process Θ is *ergodic* if $P(U)$ is 0 or 1 for any invariant set $U \in \mathcal{M}$. Stated differently, a stationary point process Θ determined by the distribution P is *ergodic* if P is not a mixture of two distinct stationary distributions, i.e. there do not exist two distinct stationary distributions P' and P'' such that $P = pP' + (1 - p)P''$ for any $0 < p < 1$. Intuitively, this implies that P is not decomposable.

The ergodicity of a spatial stochastic point process enables us to find the spatial averages in individual realizations of the stochastic point process. To be explicit, consider a stationary point process Θ with intensity $\lambda > 0$ in which its distribution is given by P. For a set U in $\mathcal{M}$, we consider a point p_i satisfying that the point is an element of a set θ of points and is included in the cube $[0, 1]^m$ (i.e. $x_i \in \theta \cap [0, 1]^m$) and that the set θ is an element of the set $\varphi_{x_i}(U)$ of sets of points (i.e. $\{x_1, x_2, \ldots\} \in \varphi_{x_i}(U)$, where x_i is the location vector of p_i). We consider the number of such points p_i, i.e. $\#\{i \mid x_i \in \{x_1, x_2, \ldots,\} \cap [0, 1]^m$ and $\{x_1, x_2, \ldots\} \in \varphi_{x_i}(U)\}$, where #(a set) means the number of elements in the set. Then the sum of this number for θ in $\mathbb{M}$ divided by λ is written as

$$P_0(U) = \frac{1}{\lambda} \int_{\mathbb{M}} \#\{i \mid x_i \in \{x_1, x_2, \ldots\} \cap [0,1]^m \text{ and } \{x_1, x_2, \ldots\} \in \varphi_{x_i}(U)\} P(d\theta), \quad U \in \mathcal{M}. \tag{1.3.44}$$

(Note that there are various expressions for $P(d\theta)$; see Stoyan *et al.*, 1995, p.101.) We call this distribution the *Palm distribution*, P_0, of P. This

distribution implies that the point process, Θ_0, determined by the distribution P_0 is a random shift of Θ such that Θ_0 has a point at the origin.

The ergodicity and the Palm distribution defined above have an important relation, which is shown in the following theorem (Daley and Vere-Jones, 1988, Proposition 12.4.I).

Theorem V1 Let P be the distribution of a stationary ergodic point process in $\mathbb{R}^m$ with finite intensity $\lambda > 0$ and P_0 the corresponding Palm distribution. Denote by E_P and E_{P_0} the expectations under the distributions P and P_0, respectively. For non-negative functions f and g defined on $\mathbb{M}$ such that $E_P(f(\Theta))$ and $E_{P_0}(g(\Theta_0))$ are finite, we have the following relations with probability 1.

$$\begin{aligned} \frac{1}{2^m a^m} \int_{[-a,a]^m} f(\varphi_{-x}(\Theta_0))\, dx &\to E_P(f(\Theta)) \quad \text{as } a \to \infty, \\ \frac{1}{n} \sum_{i=1}^{n} g(\varphi_{x_{Ni}}(\Theta)) &\to E_{P_0}(g(\Theta_0)) \quad \text{as } a \to \infty, \end{aligned} \tag{1.3.45}$$

where x_{Ni} is the ith nearest point in Θ to the origin.

Equation (1.3.45) implies that the averages along an individual realization converge to the theoretical mean, which, loosely speaking, is the average of infinitely many realizations. Equation (1.3.45) is particularly important for the study of the Poisson Voronoi diagram (see Chapter 5). Suppose that $g(\Theta)$ is a characteristic of the Voronoi cell (see Chapter 2) containing the origin. Then the left-hand side of equation (1.3.45) means the average of the characteristic of n distinct Voronoi cells in an individual realization. If Θ is ergodic, the limit of the averages exists and is defined to be the mean of the characteristic of a *typical Voronoi cell* called by Cowan (1978, 1980; see Chapter 5). Equation (1.3.45) enables us to calculate the mean by using the Palm distribution (see Chapter 5).

The following theorem is quite useful to simplify the expectation calculation under the Palm distribution of a Poisson point process (Slivnak, 1962).

Theorem V2 A stationary point process Θ is a Poisson point process if and only if $\Theta \cup \{o\}$ has the same distribution as Θ_0 where o is the origin.

The Poisson point process satisfies a stronger property than ergodicity. A stationary point process Θ determined by the distribution P is said to be:

(i) *strongly mixing (α-mixing)* if for any subsets U and V in $\mathcal{M}$,

$$P(U \cap \varphi_c(W)) - P(U)P(W) \to 0 \quad \text{as } \|c\| \to 0; \tag{1.3.46}$$

(ii) *weakly mixing* if for any subsets U and V in $\mathcal{M}$,

$$\frac{1}{2^m a^m} \int_{[-a,a]^m} \left| P(U \cap \varphi_x(W)) - P(U)P(W) \right| dx \to 0 \quad \text{as } a \to 0. \tag{1.3.47}$$

Strong mixing implies weak mixing, which in turn implies ergodicity.

In addition to the above mixing conditions, there are many other mixing conditions which are different in their so-called *mixing coefficients*. Let $\mathcal{U}$ and $\mathcal{W}$ be two collections of subsets of $\mathbb{M}$. Then the *α-mixing coefficients* and the *β-mixing coefficients* are defined as:

$$\alpha(\mathcal{U},\mathcal{W}) = \sup_{U\in\mathcal{U}, V\in\mathcal{W}} \left|\Pr(U\cap W) - \Pr(U)\Pr(W)\right|, \tag{1.3.48}$$

$$\beta(\mathcal{U},\mathcal{W}) = \mathrm{E}\sup_{W\in\mathcal{W}} \left|\Pr(W\,|\,\mathcal{U}) - \Pr(W)\right|, \tag{1.3.49}$$

where E is an expectation operator under the probability Pr and $\Pr(W\,|\,\mathcal{U})$ is the conditional probability of W given $\mathcal{U}$. If $\alpha(\mathcal{U},\mathcal{W})$ or $\beta(\mathcal{U},\mathcal{W})$ goes to zero as the 'distance' (to be defined) between $\mathcal{U}$ and $\mathcal{W}$ tends to infinity, then we say that the α-mixing condition or the *β-mixing* (*absolute regularity*) condition is satisfied, respectively. Other mixing coefficients such as $\psi(\mathcal{U},\mathcal{W})$, $\varphi(\mathcal{U},\mathcal{W})$ and $\varrho(\mathcal{U},\mathcal{W})$ have also been defined in the literature (see, for example, Lin and Lu, 1996).

If $\mathcal{U}$ and $\mathcal{W}$ are the collections of subsets of $\mathbb{M}$ which involve Θ and $\varphi_{-c}(\Theta)$, respectively, and the distance between them is defined to be $\|\boldsymbol{c}\|$, then the point process Θ is α-mixing or β-mixing if the mixing coefficient $\alpha(\mathcal{U},\mathcal{W}) \to 0$ or $\beta(\mathcal{U},\mathcal{W}) \to 0$, respectively, as $\|\boldsymbol{c}\| \to \infty$.

Finally, let us mention one more useful mixing condition. Suppose that $\mathcal{U}$ and $\mathcal{W}$ are collections of subsets of $\mathbb{M}$ which involve $\Theta \cap [-a, a]^m$ and $\Theta \cap \{\mathbb{R}^m \setminus [-b, b]^m\}$, respectively, and the distance between them is defined to

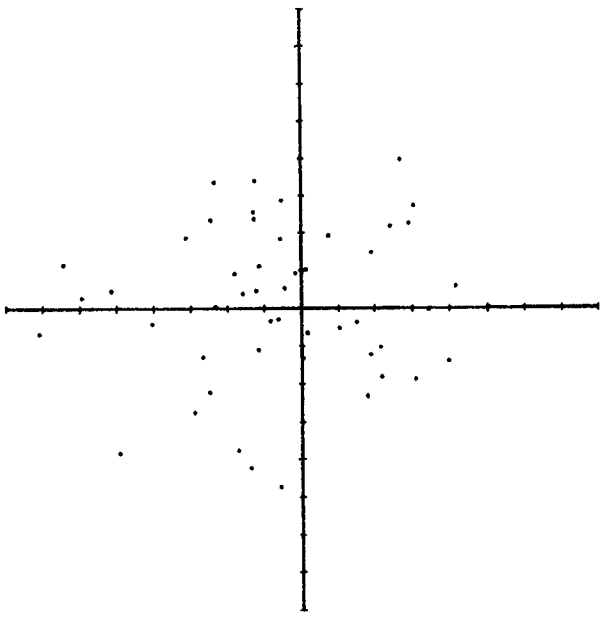


Figure 1.3.24 A general Poisson point process with $\lambda(\boldsymbol{x}) = c_1 \exp(-c_2\sqrt{x_1^2 + x_2^2})$.

be $b - a$. Heinrich (1994) established various theorems for Voronoi diagrams generated by a point process Θ (e.g. a homogeneous Poisson point process) for which $\beta(\mathcal{U}, \mathcal{W}) \to 0$ as $b - a \to \infty$ (see Section 5.6.1).

The Poisson point process may be extended in several directions. First, we may relax homogeneity, and define the point process by

$$\Pr(N(A) = x) = \frac{\lambda(A)|A|e^{-\lambda(A)|A|}x}{x!}, \quad x = 0, 1, 2, \ldots, \tag{1.3.50}$$

where

$$\lambda(A) = \int_A \lambda(x)\, dx. \tag{1.3.51}$$

We call this point process the *general Poisson point process* with the intensity function $\lambda(x)$. An example is shown in Figure 1.3.24, where $\lambda(x) = c_1 \exp(-c_2\sqrt{x_1, + x_2}), c_1, c_2 > 0$.

The general Poisson point process reduces to the Poisson point process of equation (1.3.30) if $\lambda(x) = \lambda$ (a constant). Since λ in equation (1.3.30) indicates the density of points, $\lambda(x)$ implies that the density depends upon where A is placed. Thus the density of points may vary from location to location in a general Poisson point process. This property distinguishes from that of the Poisson point process defined by equation (1.3.30) which we call the *homogeneous Poisson point process*.

In the homogeneous Poisson point process the parameter λ is deterministic. We may relax this assumption, and regard it as probabilistic. We call the Poisson point process with a probabilistic parameter λ a *doubly stochastic Poisson point process*, a *compound Poisson point process*, or a *Cox point process*. Such a point process is in general not ergodic.

We may also relax the property of independent scattering. Examples include the Poisson cluster processes, hard core processes and Gibbs point processes. We briefly mention these point processes here (in Section 5.12, we deal with Voronoi diagrams generated by these point processes).

Poisson cluster processes are point processes in which a cluster of points are independently scattered around a point of a Poisson point process; the clusters are independent and finite with probability 1; points of the Poisson point process are then deleted. If the clusters are not only independent but also identically distributed, such Poisson cluster processes are called the *Neyman–Scott processes*. *Hard core processes* are point processes in which there is a minimum interpoint distance. *Gibbs point processes* are a class of point processes in which interaction between points exists. A rigorous treatment of Gibbs point processes is rather mathematically involved. Here we present the simplest case: a Gibbs point process with a non-random total number of points in a bounded subset of $\mathbb{R}^m$. The distribution P of such a point process (see equation (1.3.39)) is in the form:

$$P(U) = \int \ldots \int_{\{x_1, \ldots x_n\} \in U} f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n, \quad U \in \mathcal{M}, \tag{1.3.52}$$

where the probability density function is given by

$$f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \propto \exp\left(-C(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)\right) \tag{1.3.53}$$

The function C, which does not depend on the order of its arguments, is called the *interaction energy function*. In statistical physics C is usually assumed to be the sum of *pair potentials* given by

$$C(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \sum \sum_{1 \le i \le j \le n} h\left(\|\boldsymbol{x}_i - \boldsymbol{x}_j\|\right). \tag{1.3.54}$$

If the pair potential function is

$$h(\|\boldsymbol{x}_i - \boldsymbol{x}_j\|) = \begin{cases} \infty & \text{if } \|\boldsymbol{x}_i - \boldsymbol{x}_j\| \le r, \\ -b & \text{if } r < \|\boldsymbol{x}_i - \boldsymbol{x}_j\| \le R, \\ 0 & \text{if } \|\boldsymbol{x}_i - \boldsymbol{x}_j\| > R, \end{cases} \tag{1.3.55}$$

then the corresponding Gibbs point process is a hard core point process with minimum interpoint distance r. If $b > 0$ or $b < 0$, there is attraction or repulsion, respectively, between points within a distance R. When the bounded subset is replaced by $\mathbb{R}^m$, the formulation is more sophisticated and so we do not discuss it here. The reader who is interested in general point processes, including the Gibbs point processes, should consult Bartlett (1975), Haggett *et al.* (1977, Chapter 13), Pielou (1977), Getis and Boots (1978), Matthes *et al.* (1978), Cliff and Ord (1981), Diggle (1983), Kallenberg (1983), Upton and Fingleton (1985), Daley and Vere-Jones (1988), Ripley (1988), Cressie (1991), Snyder and Miller (1991), Kingman (1993), Reiss (1993), and Stoyan *et al.* (1995), among others.

A random set of points or a stochastic point process may be generalized to a random set of lines or a stochastic line process. Stated precisely, a *stochastic line process* is a random collection of lines in $\mathbb{R}^m$ and there are only finitely many lines that hit each bounded subset of $\mathbb{R}^m$. A line is uniquely determined by its perpendicular distance to the origin and the angle between the line and the x_1-axis measured in a counterclockwise direction. Thus a line can be represented by a point in $\mathbb{R}^m \times (0, \pi]$. The *Poisson line process* in $\mathbb{R}^m$ has the same distribution as the Poisson point process in $\mathbb{R}^m \times (0, \pi]$. Similarly, a random set of k-dimensional hyperplanes in $\mathbb{R}^m$ $(1 \le k \le m)$ forms a *hyperplane process*, which can be regarded as a point process in $\mathbb{R}^m \times (0, \pi]$. The *Poisson hyperplane process* is a stochastic hyperplane process the distribution of which is the same as the Poisson point process in $\mathbb{R}^m \times (0, \pi]$. A Poisson m-dimensional hyperplane process divides the space $\mathbb{R}^m$ into polytopes. The resultant structure is known as the *Poisson hyperplane tessellation*. In particular, the Poisson m-dimensional hyperplane process (tessellation) is known as the *Poisson line process (tessellation)* if $m = 2$ and the *Poisson plane process (tessellation)* if $m = 3$. Details of the Poisson hyperplane tessellations can be found in Santaló (1976), Miles (1964a,b, 1972a, 1973, 1995), Matheron (1975), Crain and Miles (1976), Mecke (1995, 1999), and Stoyan *et al.* (1995).

In a stochastic line process we consider a random set of lines. Naturally, we can consider a random set of line segments. For example, at each point

of a Poisson point process we put a line segment the length of which is random and the angle between the line segment and the x_1-axis is uniformly distributed on $(0, \pi]$. A line segment can be further generalized to a fibre, which is a segment of a curve. A random set of fibres in $\mathbb{R}^2$ is called a *stochastic fibre process*. Just as a line process in $\mathbb{R}^2$ can be generalized to a hyperplane process in $\mathbb{R}^m$, a stochastic fibre process can be generalized to a *stochastic manifold process* (Mecke, 1981), which is a random set of fragments of manifolds of fixed dimension. A *manifold* is the solution of a system of equations $f_1(x) = 0, \ldots, f_k(x) = 0$. Thus, a random set of fragments of hyperplanes is a stochastic manifold process.

1.3.4 Efficiency of computation

When we carry out a computation using an algorithm with a set of input data, we wish to evaluate how 'efficiently' the algorithm runs. A standard method for evaluating the efficiency is to observe the asymptotic behaviour of the time required by the algorithm with respect to the size of input data (Knuth, 1968; Aho *et al.*, 1974). Let $T(n)$ denote the time required by an algorithm to process input data of size n. In general there are many different input data of the same size and they require different processing times. Hence we usually consider the worst case or the average case. In the worst case, $T(n)$ is the maximum time over all possible inputs of size n, whereas in the average case $T(n)$ is the average time over them. For some positive function $f(n)$ if there exists a constant C such that

$$\frac{T(n)}{f(n)} < C \text{ for all } n, \tag{1.3.56}$$

we write $T(n) = O(f(n))$ and say that $T(n)$ is of *order* $f(n)$. If $T(n) = O(f(n))$, we also say that the *time complexity* of the algorithm is of $O(f(n))$. Intuitively, $T(n)$ being of $O(f(n))$ implies that $T(n)$ does not increase more rapidly that $f(n)$ does as n grows. Hence the algorithm can be considered as being efficient if it admits a slowly increasing function $f(n)$ satisfying equation (1.3.56).

We say, for example, an algorithm with $T(n) = O(1)$ is ' ideally fast', and an algorithm with $T(n) = O(\log n)$ is 'very fast'. An algorithm with $T(n) = O(n)$ is called a *linear time* algorithm, and an algorithm with $T(n) = O(n^q)$ is called a *polynomial time* algorithm.

Turing proposed an abstract model of a computer, which is called a *Turing machine*. There are two types of Turing machines. One is a *deterministic Turing machine*; it corresponds to an actual computer in which the machine behaves deterministically according to a given sequence of instructions (i.e. a program). The other type is a *non-deterministic Turing machine*. It does not correspond to an actual computer; it is more powerful in the sense that this machine can choose one of any possible branches of instruction sequences non-deterministically. See Aho *et al.* (1974), for example, for the strict definition of the Turing machine.

A problem whose answer is either 'yes' or 'no' is called a *decision problem*. For example, for a given set V of vectors, to judge whether V is linearly independent is a decision problem, whereas to find a maximal linearly independent subset of V is not a decision problem. The set of all decision problems that can be solved in polynomial time by a deterministic Turing machine is called a *class P*. On the other hand, the set of all decision problems that can be solved in polynomial time by a non-deterministic Turing machine is called a *class NP*. Since the non-deterministic Turing machine is more powerful than the deterministic one, the class NP includes the class P. It is a big open problem to judge whether the class P is a proper subset of the class NP.

A decision problem p is called *NP-complete* provided that p satisfies the following two conditions:

(i) p belongs to the class NP.
(ii) If p belongs to the class P, all the problems in the class NP belong to the class P.

Condition (ii) intuitively means that if someone finds an algorithm for solving p in polynomial time, then all the problems in the class NP can be solved using this algorithm. Hence, an NP-complete problem can be regarded as one of the most difficult problems in the class NP. The set of all NP-complete problems is called *class NP-complete*.

For a given graph, a cycle visiting all the vertices exactly once is called a *Hamilton cycle*. To judge whether a given graph has a Hamilton cycle is a decision problem belonging to the class NP-complete.

A general problem (i.e. a problem which is not necessarily a decision problem) is said to be *NP-hard* if it contains an NP-complete problem as a subproblem.

For a given geometric graph, finding the Hamilton cycle whose total length is minimum is called the *travelling salesperson problem*. The travelling salesperson problem is NP-hard, because if we solve this problem, we can also judge whether the graph has a Hamilton cycle.

Spatial Tessellations: Concepts and Applications of Voronoi Diagrams
Edition by ATSUYUKI OKABE & BARRY BOOTS and SUNG NOK CHIU

CHAPTER 2

Definitions and Basic Properties of Voronoi Diagrams

In Chapter 1 we discussed the concept of a Voronoi diagram from a historical viewpoint. In this chapter we wish to fully develop this concept from a geometric viewpoint.

The chapter consists of six sections. In Section 2.1 we define a Voronoi diagram and introduce notations to be commonly used in this text. In Section 2.2 we define a Delaunay tessellation, in particular a Delaunay triangulation in terms of the 'dual tessellation' of a Voronoi diagram. Voronoi diagrams and Delaunay tessellations are, as we overviewed in Section 1.2, quite useful in various fields because of their nice geometric properties. In Sections 2.3 and 2.4 we show the basic geometric properties of Voronoi diagrams and Delaunay triangulations, respectively (specific properties only used in each chapter will be shown in that chapter). In Section 2.5 we refer to geometric graphs closely related to a Delaunay triangulation, such as a Gabriel graph, a relative neighbourhood graph and a minimum spanning tree. In Section 2.6 we show how to judge whether or not a given diagram is a Voronoi diagram.

2.1 DEFINITIONS OF THE ORDINARY VORONOI DIAGRAM

For ease of exposition, we first give a fairly intuitive definition of a Voronoi diagram in a plane. We next restate this definition more precisely in mathematical terms. Lastly, extending the definition in $\mathbb{R}^2$ to $\mathbb{R}^m$, we define an m-dimensional Voronoi diagram.

Suppose that a set of points is given in the Euclidean plane (for example, the filled circles in Figure 2.1.1). The number of points is assumed to be two or more but finite and they are all distinct in the sense that no points coincide in the plane. Given this point set, we assign every location in the plane to the closest member in the point set (in Figure 2.1.1, for example, the point p is assigned to the filled circle incident to the heavy broken line). If a location happens to be equally close to two or more members of the point set,

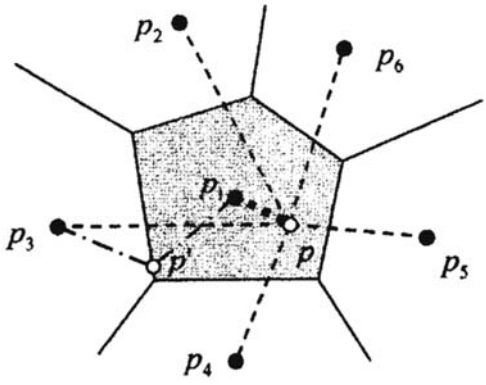


Figure 2.1.1 A planar ordinary Voronoi diagram.

we assign the location to those members (point p' in Figure 2.1.1 is assigned to the two filled circles incident to the dash-dot lines in Figure 2.1.1). As a result, the set of locations assigned to each member in the point set forms its own region (the shaded region in Figure 2.1.1). The resulting regions are collectively exhaustive in the plane because every location is assigned to at least one member in the point set. The set of locations assigned to two or more members in the point set forms the boundaries of the regions (the continuous lines in Figure 2.1.1). Hence adjacent regions overlap only on their boundaries. Thus the set of the regions is collectively exhaustive and mutually exclusive except for boundaries; namely, the set of the regions forms a tessellation. We call this tessellation a *planar ordinary Voronoi diagram*, and the regions constituting the Voronoi diagram *ordinary Voronoi polygons*.

To sum up, we have the following definition of a Voronoi diagram.

Definition V1 (a planar ordinary Voronoi diagram) Given a set of two or more but a finite number of distinct points in the Euclidean plane, we associate all locations in that space with the closest member(s) of the point set with respect to the Euclidean distance. The result is a tessellation of the plane into a set of the regions associated with members of the point set. We call this tessellation the *planar ordinary Voronoi diagram* generated by the point set, and the regions constituting the Voronoi diagram *ordinary Voronoi polygons*.

Note that when the Euclidean plane (space) is understood, or when there is no confusion with generalized Voronoi diagrams (to be shown in Chapter 3), we shall often refer to a planar ordinary Voronoi diagram simply as a *Voronoi diagram* and an ordinary Voronoi polygon as a *Voronoi polygon*. Also note that we exclude from our discussion the point set consisting of only one member, because such a Voronoi diagram is too trivial (it consists of only one Voronoi polygon which is the whole plane).

In the above definitions we define the Voronoi diagram for a set of finitely many points, but we can also define a diagram for a set of infinitely many points in the same manner. We also call the resulting diagram a Voronoi diagram (which is referred to as an *infinite Voronoi diagram* in Chapter 5). One diagram of this kind, the Poisson Voronoi diagram, is the subject of Chapter 5. In other chapters, however, we deal with only the former Voronoi diagram (which may be called a *finite Voronoi diagram*) because it is consistent with the applications to be presented throughout the book and, moreover, we want to avoid lengthy treatments resulting from the distinction between finite and infinite.

Let us now restate Definition V1 in mathematical terms. We consider a finite number, n, of points in the Euclidean plane, and assume that $2 \leq n < \infty$. The n points are labelled by $p_1, \ldots, p_n$ with the Cartesian coordinates $(x_{11}, x_{12}), \ldots, (x_{n1}, x_{n2})$ or location vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$. The n points are distinct in the sense that $\boldsymbol{x}_i \neq \boldsymbol{x}_j$ for $i \neq j$, $i, j \in I_n = \{1, \ldots, n\}$. Let p be an arbitrary point in the Euclidean plane with coordinates (x_1, x_2) or a location vector $\boldsymbol{x}$. Then the Euclidean distance between p and p_i is given by $d(p, p_i) = \|\boldsymbol{x} - \boldsymbol{x}_i\| = \sqrt{(x_1 - x_{i1})^2 + (x_2 - x_{i2})^2}$. If p_i is the nearest point from p or p_i is one of the nearest points from p, we have the relation $\|\boldsymbol{x} - \boldsymbol{x}_i\| \leq \|\boldsymbol{x} - \boldsymbol{x}_j\|$ for $j \neq i, j \in I_n$ (recall the heavy broken and dash-dot lines in Figure 2.1.1). In this case, p is assigned to p_i. Therefore, Definition V1 is written mathematically as follows.

Definition V2 (a planar ordinary Voronoi diagram) Let $P = \{p_1, \ldots, p_n\} \subset \mathbb{R}^2$, where $2 < n < \infty$ and $\boldsymbol{x}_i \neq \boldsymbol{x}_j$ for $i \neq j$, $i, j \in I_n$. We call the region given by

$$V(p_i) = \{\boldsymbol{x} \mid \|\boldsymbol{x} - \boldsymbol{x}_i\| \leq \|\boldsymbol{x} - \boldsymbol{x}_j\| \quad \text{for } j \neq i,\ j \in I_n\} \tag{2.1.1}$$

the *planar ordinary Voronoi polygon* associated with p_i (or the Voronoi polygon of p_i), and the set given by

$$\mathcal{V} = \{V(p_1), \ldots, V(p_n)\} \tag{2.1.2}$$

the *planar ordinary Voronoi diagram* generated by P (or the Voronoi diagram of P). We call p_i of $V(p_i)$ the *generator point* or *generator* of the ith Voronoi polygon, and the set $P = \{p_1, \ldots, p_n\}$ the *generator set* of the Voronoi diagram $\mathcal{V}$ (in the literature, a generator point is sometimes referred to as a *site*).

For brevity we may write V_i for $V(p_i)$. Also, we may use $V(x_{i1}, x_{i2})$ or $V(\boldsymbol{x}_i)$ when we want to emphasize the coordinates or location vector of the generator point p_i. In addition, we may use $\mathcal{V}(P)$ when we want to explicitly indicate the generator set P of $\mathcal{V}$.

In Definition V2 the reader should notice that the relation in equation (2.1.1) is defined in terms of $\leq$, but not $<$. A Voronoi polygon is hence a closed set. Alternatively, we may define a Voronoi polygon as

$$V^{\circ}(p_i) = \{\boldsymbol{x} \mid \|\boldsymbol{x} - \boldsymbol{x}_i\| < \|\boldsymbol{x} - \boldsymbol{x}_j\| \quad \text{for } j \neq i,\ j \in I_n\}, \tag{2.1.3}$$

which is an open set. Both definitions are acceptable, but in this text we

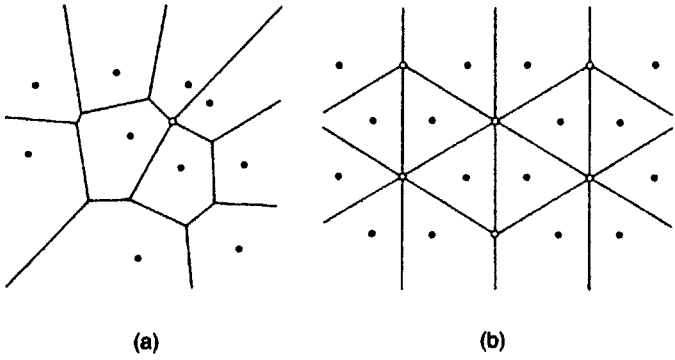


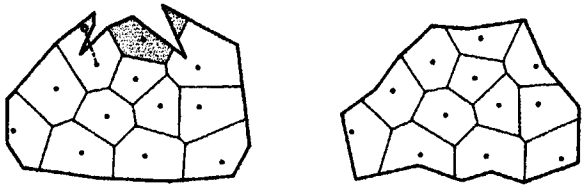
Figure 2.1.2 Degenerate Voronoi diagrams.

define a Voronoi polygon as a closed set. When we want to emphasize this property, we call $V(p_i)$ a *closed Voronoi polygon*, and $V^o(p_i)$ an *open Voronoi polygon.*

Since a Voronoi polygon is a closed set, it contains its boundary, which is denoted by $\partial V(p_i)$. The boundary of a Voronoi polygon may consist of line segments, half lines or infinite lines, which we call *Voronoi edges*. We denote a Voronoi edge by e_i. Note that a Voronoi diagram is sometimes defined by the union of Voronoi edges, i.e. $\cup_{i=1}^{n} \partial V(p_i)$ in place of the set $\{V(p_1), \ldots, V(p_n)\}$. Since the union of Voronoi edges may be regarded as a network, it is sometimes called a *Voronoi network* (Medvedev *et al.*, 1988; Bartkowiak and Mahan, 1995).

Noticing that = is included in the relation of equation (2.1.1), we may alternatively define a Voronoi edge as a line segment, a half line or an infinite line shared by two Voronoi polygons with its end points. Mathematically, if $V(p_i) \cap V(p_j) \neq \emptyset$, the set $V(p_i) \cap V(p_j)$ gives a Voronoi edge (which may degenerate into a point). We use $e(p_i, p_j)$ for $V(p_j) \cap V(p_j)$, which is read as the Voronoi edge generated by p_i and p_j. Note that $e(p_j, p_j)$ may be empty. If $e(p_i, p_j)$ is neither empty nor a point, we say that the Voronoi polygons $V(p_i)$ and $V(p_j)$ are *adjacent.*

An end point of a Voronoi edge is called a *Voronoi vertex*. Alternatively, a Voronoi vertex may be defined as a point shared by three or more Voronoi polygons. We denote a Voronoi vertex by q_i. When there exists at least one Voronoi vertex at which four or more Voronoi edges meet in the Voronoi diagram $\mathcal{V}$ (the unfilled circle in Figure 2.1.2(a)), we say that $\mathcal{V}$ is *degenerate*; otherwise, we say that $\mathcal{V}$ is *non-degenerate*. The Voronoi diagram in Figure 2.1.2(a) is degenerate and that in Figure 2.1.1 is non-degenerate. A degenerate Voronoi diagram often appears when generator points are regularly spaced, such as in Figure 2.1.2(b). In some derivations, a degenerate Voronoi diagram requires special lengthy treatments which are not always essential. To avoid this difficulty, we shall often make the following assumption.



(a) (b)

Figure 2.1.3 Bounded Voronoi diagrams: (a) a disconnected boundary Voronoi region (the shaded region) and a non-star-shaped boundary Voronoi polygon with respect to its generator (the broken line segment); (b) star-shaped boundary Voronoi polygons with respect to their generator points.

Assumption V1 (the non-degeneracy assumption) Every Voronoi vertex in a Voronoi diagram has exactly three Voronoi edges.

In Definition V1 or V2, we defined a Voronoi diagram in an unbounded plane. In practical applications, however, we often deal with a bounded region S where generator points are placed (the heavy lines in Figure 2.1.3). In this case we consider the set, $\mathcal{V}_{\cap S}$, given by

$$\mathcal{V}_{\cap S} = \{V(p_1) \cap S, \ldots, V(p_n) \cap S\}. \tag{2.1.4}$$

We call this set a *bounded Voronoi diagram* or the *Voronoi diagram bounded by S* (which should be distinguished from the bounded Voronoi diagram defined by Wang and Schubert, 1987; see Section 3.4). If a Voronoi polygon $V(p_i)$ shares the boundary of S, we call the region $V(p_i) \cap S$ a *boundary Voronoi polygon* or *region* (the term 'region' is used when ∂S is curved or when $V(p_i) \cap S$ is not connected).

We should note that a boundary Voronoi region may be disconnected (the shaded region in Figure 2.1.3(a)), and that the line segment joining a point in $V(p_i)$ and p_i may not be contained in $V(p_i) \cap S$ (the broken line segment in Figure 2.1.3(a)). In practical applications, such boundary Voronoi polygons are often problematic, and so we have to define a more appropriate Voronoi diagram. We shall show such an alternative Voronoi diagram in Section 3.4. A bounded Voronoi diagram may be meaningful in practice if every boundary Voronoi region is star-shaped with respect to its generator point (Figure 2.1.3(b); recall Figure 1.3.15). In practice, we usually treat such a well-formed bounded Voronoi diagram.

As we observed in Figure 2.1.1, the ordinary Voronoi diagram consists of polygons. Recalling that a polygon is defined in terms of half planes (Section 1.3), we may alternatively define a Voronoi diagram in terms of half planes. To show this alternative definition, we consider the line perpendicularly bisecting the line segment $\overline{p_i p_j}$ joining two generator points p_i and p_j (Figure 2.1.4). We call this line the *bisector* between p_i and p_j and denote it by

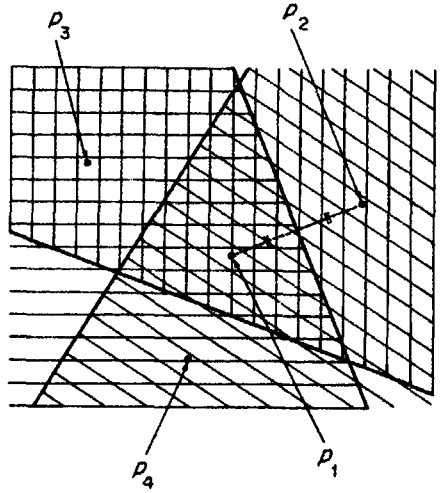


Figure 2.1.4 A Voronoi polygon obtained from half planes.

$b(p_i, p_j)$. Since a point on the bisector $b(p_i, p_j)$ is equi-distant from the generator points p_i and p_j, $b(p_i, p_j)$ is written as

$$b(p_i, p_j) = \{x \mid \|x - x_i\| = \|x - x_j\| \quad j \neq i. \tag{2.1.5}$$

The bisector divides the plane into two half planes and gives

$$H(p_i, p_j) = \{x \mid \|x - x_i\| \leq \|x - x_j\| \quad j \neq i. \tag{2.1.6}$$

We call the region $H(p_i, p_j)$ the *dominance region* of p_i over p_j. In Figure 2.1.4 we indicate the dominance regions of p_1 over p_2, p_3 and p_4 by the horizontally, diagonally and vertically hatched regions, respectively. Obviously, in the dominance region $H(p_i, p_j)$ the distance to the generator p_i is shorter than or equal to the generator p_j. In Figure 2.1.4, therefore, the distance from a point p in the intersection of the vertically, horizontally and diagonally hatched regions to the generator p_1 is shorter than or equal to the distance from p to the generator p_j, $j = 2, 3, 4$. This relation is equivalent to equation (2.1.1), and hence the intersection $H(p_1, p_2) \cap H(p_1 \cap p_3) \cap H(p_1, p_4)$ gives the Voronoi polygon associated with p_1. From this example, we understand that the following definition is an alternative to Definition V2.

Definition V3 (a planar ordinary Voronoi diagram defined with half planes) Let $P = \{p_1, \ldots, p_n\} \subset \mathbb{R}^2$, where $2 \leq n < \infty$ and $x_i \neq x_j$ for $i \neq j, i, j \in I_n$. We call the region

$$V(p_i) = \bigcap_{j \in I_n \setminus \{i\}} H(p_i, p_j) \tag{2.1.7}$$

the *ordinary Voronoi polygon* associated with p_i and the set $\mathcal{V}(P) = \{V(p_1), \ldots, V(p_n)\}$ the *planar ordinary Voronoi diagram* generated by P.

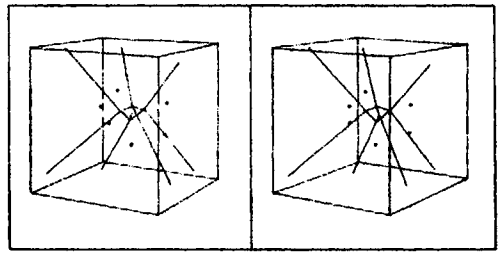


Figure 2.1.5 A stereographic view of a three-dimensional Voronoi diagram (view the left panel with the left eye and the right panel with the right eye at the same time, or use a stereographic viewer). (*Source*: Gen, 1983, p. 178.)

The equivalence of Definition V3 to Definition V2 is apparent, because $\|\boldsymbol{x} - \boldsymbol{x}_i\| \leq \|\boldsymbol{x} - \boldsymbol{x}_j\|$ if and only if $\boldsymbol{x} \in H(p_i, p_j)$ for $j \neq i$.

We can readily extend the above definition to the m-dimensional Euclidean space.

Definition V4 (an ordinary Voronoi diagram in $\mathbb{R}^m$) Let $P = \{p_1, \ldots, p_n\} \subset \mathbb{R}^m$, where $2 \leq n < \infty$ and $\boldsymbol{x}_i \neq \boldsymbol{x}_j$ for $i \neq j, i, j \in I_n$. We call the region

$$V(p_i) = \{\boldsymbol{x} \mid \|\boldsymbol{x} - \boldsymbol{x}_i\| \leq \|\boldsymbol{x} - \boldsymbol{x}_j\| \quad \text{for} j \neq i,\ j \in I_n\} \tag{2.1.8}$$

$$= \bigcap_{j \in I_n \setminus \{i\}} H(p_i, p_j) \tag{2.1.9}$$

the *m-dimensional ordinary Voronoi polyhedron* associated with p_i, and the set $\mathcal{V}(P) = \{V(p_1), \ldots, V(p_n)\}$ the *m-dimensional ordinary Voronoi diagram* generated by P, where $H(p_i, p_j)$ is given by equation (2.1.6) for $p_i, p_j \in \mathbb{R}^m$.

Note that when $\mathbb{R}^m$ is understood, we may simply call $\mathcal{V}(P)$ an *ordinary Voronoi diagram* or just a *Voronoi diagram*, and that Voronoi polyhedra (polygons) are sometimes called *Voronoi cells*.

For the three-dimensional Voronoi diagram, the boundaries of a Voronoi polyhedron consists of facets, which we call *Voronoi facets*. The boundaries of a Voronoi facet consist of line segments, half lines or infinite lines, which we call *Voronoi edges*. The boundaries of a Voronoi edge consist of points, which we call *Voronoi vertices*. Figure 2.1.5 shows a stereographic view of a three-dimensional Voronoi diagram obtained by Gen (1983) (see also Bowyer, 1981). In astronomy, realized Voronoi polyhedra in a three-dimensional space are sometimes called *Voronoi foams* (Icke and van der Weygaert, 1987). In winner-take-all type neural networks, the Voronoi polyhedron $V(p_i)$ is called the *receptive field* of a neural unit i with $\boldsymbol{x}_i$ being the synaptic weight vector of this neural unit (Martinetz and Schulten, 1994).

A visual presentation of an m-dimensional Voronoi diagram becomes difficult when $m \geq 4$. Formally, we call the boundaries of an m-dimensional

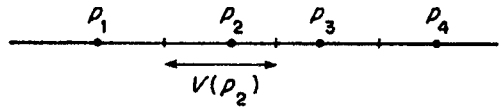


Figure 2.1.6 A Voronoi diagram on a line.

Voronoi polyhedron $(m-1)$-*dimensional Voronoi faces*; the boundaries of an $(m-1)$-dimensional Voronoi face $(m-2)$-*dimensional Voronoi faces* etc.; the boundaries of a three-dimensional Voronoi face *two-dimensional Voronoi faces* or simply *Voronoi faces*; the boundaries of a two-dimensional Voronoi face *one-dimensional Voronoi faces* or *Voronoi edges*; the boundaries of a one-dimensional Voronoi face *zero-dimensional Voronoi faces* or *Voronoi vertices*. An $(m-1)$-dimensional Voronoi face is also called a *Voronoi facet*.

Obviously we allow $m = 1$, a *one-dimensional Voronoi diagram*, or a *Voronoi diagram on a line* (Figure 2.1.6). In this case, a 'one-dimensional Voronoi polyhedron' is a half line or a line segment called a *Voronoi line*, and Voronoi vertices are end points of Voronoi lines. We easily notice that the boundary point between two adjacent Voronoi lines is the midpoint of the generator points of those Voronoi lines; the number of unbounded Voronoi lines is always two; the number of Voronoi lines adjacent to a Voronoi line is one or two. Because of these simple geometric properties, a Voronoi diagram on a line is often adopted in a simplified theoretical context to avoid complicated geometric arguments (for example, the original Hotelling model in Section 7.3). In this text we shall mainly treat a Voronoi diagram whose dimension is two or more.

In the above definitions, a space is represented by a continuous plane. This representation, however, may not be acceptable in some contexts, for instance raster image analysis. In such an analysis a space is represented by a set of square cells forming a grid lattice, and a geometrical figure is represented in terms of these cells. To be explicit, let $c_{ij} = \{(x, y) \mid i \le x \le i+1,\ j \le y \le j+1\}$, $i, j = \ldots, -n, \ldots, -1, 0, 1, \ldots, n, \ldots$. For a set of points representing a geometrical figure, G, we define $\mathrm{Im}(G) = \{c_{ij} \mid c_{ij} \cap G \neq \emptyset\}$, called the *image* of G. In terms of $\mathrm{Im}(G)$, we give the following definition (Dehne, 1989; Dehne *et al.*, 1991).

Definition V5 (a planar digitized Voronoi diagram) Let $\mathcal{V} = \{V(p_1), \ldots, V(p_n)\}$ be the Voronoi diagram defined in Definitions V1–V3. We call the set $\mathrm{Im}\mathcal{V} = \{\mathrm{Im}(V(p_1)), \ldots, \mathrm{Im}(V(p_n))\}$ the *(planar ordinary) digitized Voronoi diagram* of $\mathcal{V}$, $\mathrm{Im}(V(p_i))$ a *digitized Voronoi region*, and $\mathrm{Im}(\partial V(p_i))$ the *border* of a digitized Voronoi region $\mathrm{Im}(V(p_i))$.

An example is shown in Figure 2.1.7. Dehne (1989) shows a computational method for the planar ordinary digitized Voronoi diagram. Dehne (1989) also shows a computational method for a planar digitized Voronoi diagram generated by a set of objects with convex distance functions (see Section

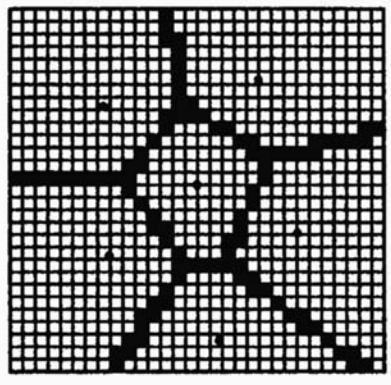


Figure 2.1.7 A planar ordinary digitized Voronoi diagram.

3.7). Embrechts and Roose (1996) develop a parallel algorithm using the Euclidean transformation. Also see Adamatzky (1994, 1996).

Alternatively we may define a Voronoi diagram by assigning cells to their nearest cells (Melkemi and Vandorpe, 1994; Watanabe and Murashima, 1996).

Definition V6 (a planar digital Voronoi diagram) Let $C = \{c_{ij}, i = 1, \ldots, n_h, j = 1, \ldots, n_v\}$ and $P_c = \{p_{c1}, \ldots, p_{cn}\}$ be a subset of C, where $2 \le n \le n_v n_h < \infty$, and $p_{ci} \neq p_{cj}$ for $i \neq j, i, j \in I_n$, and $d(c_{ij}, p_{ck})$ denotes the Euclidean distance between the centre of c_{ij} and that of p_{ck}. We call the region

$$\begin{aligned} V(p_{ci}) = \{c_{kl} \mid & d(c_{kl}, p_{ci}) < d(c_{kl}, p_{cj}) \quad \text{for } j \neq i; \\ & d(c_{kl}, p_{ci}) = d(c_{kl}, p_{cj}) < d(c_{kl}, p_{cm}) \\ & \text{for } i < j \neq m,\ k \in I_{n_h},\ l \in I_{n_v}\} \end{aligned} \tag{2.1.10}$$

the *digital Voronoi polygon* associated with p_{ci} and the set $\mathcal{V}(P_c) = \{V(p_{c1}), \ldots, V(p_{cn})\}$ the *digital Voronoi diagram* generated by P_c. Note that the second condition in equation (2.1.10) means that if a cell c_{kl} is equally distant from the generator cells p_{ci} and p_{cj}, we assign, for convenience, the cell c_{kl} to the generator that has a smaller index, i.e. p_{ci} where $i < j$ (not both as in the ordinary Voronoi diagram). Toriwaki and Yokoi (1988), Schwarzkopf (1989), Melter and Stojmenović (1995), and Watanabe and Murashima (1996) call the digital Voronoi diagram the *discrete Voronoi diagram*.

Recalling Definition V2, we notice that $V(p_i)$ is alternatively written in terms of a function $f_i(\boldsymbol{x}) = \|\boldsymbol{x} - \boldsymbol{x}_i\|$ as

$$V(p_i) = \{\boldsymbol{x} \mid f_i(\boldsymbol{x}) = \min_{j \in I_n} \{f_j(\boldsymbol{x})\}\}. \tag{2.1.11}$$

Geometrically, $z = f_i(\boldsymbol{x})$ shows the cone centred at $\boldsymbol{x}_i$ in the three-dimensional space $\{(x_1, x_2, z)\}$, and hence $z = f_{\min}(\boldsymbol{x}) = \min_{j \in I_n} \{f_j(\boldsymbol{x})\}$ shows the lower envelope of the n cones, $z = f_1(\boldsymbol{x}), \ldots, z = f_n(\boldsymbol{x})$. This envelope forms

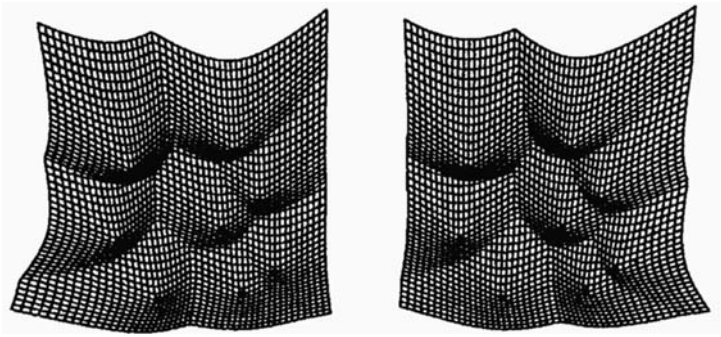


Figure 2.1.8 A Voronoi surface (look at the left-hand figure with the right eye and the right-hand figure with the left eye).

a surface, which we call the *Voronoi surface* of $\mathcal{V}$ (Huttenlocher *et al.*, 1993). An example is shown in Figure 2.1.8 (also see Figure 2 in Webster, 1998). Alternatively we may replace $f_i(\boldsymbol{x}) = \|\boldsymbol{x} - \boldsymbol{x}_i\|$ with $f_i(\boldsymbol{x}) = \|\boldsymbol{x} - \boldsymbol{x}_i\|^2$, and obtain another surface $z = f_{\min}(\boldsymbol{x})$. This surface is differentiable except at points on the boundaries $\cup_{i=1}^{n} \partial V(p_i)$. Siersma (1998) examines the differential topology of this surface through the Morse theory.

2.2 DEFINITIONS OF THE DELAUNAY TESSELLATION (TRIANGULATION)

In the same way that a planar graph has its dual graph, a Voronoi diagram has its 'dual tessellation', called a Delaunay tessellation. In this section we define a Delaunay tessellation, in particular a Delaunay triangulation.

We consider a Voronoi diagram in the Euclidean plane, and assume that generator points of the Voronoi diagram are not on the same line as in Figure 2.2.1(a). Since this assumption will be adopted often in this text, we shall refer to it as:

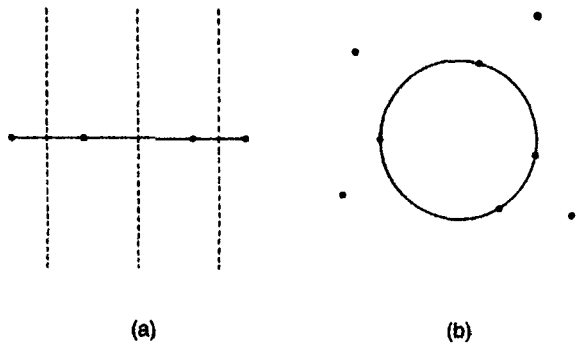


Figure 2.2.1 (a) Collinear generator points; (b) cocircular generator points.

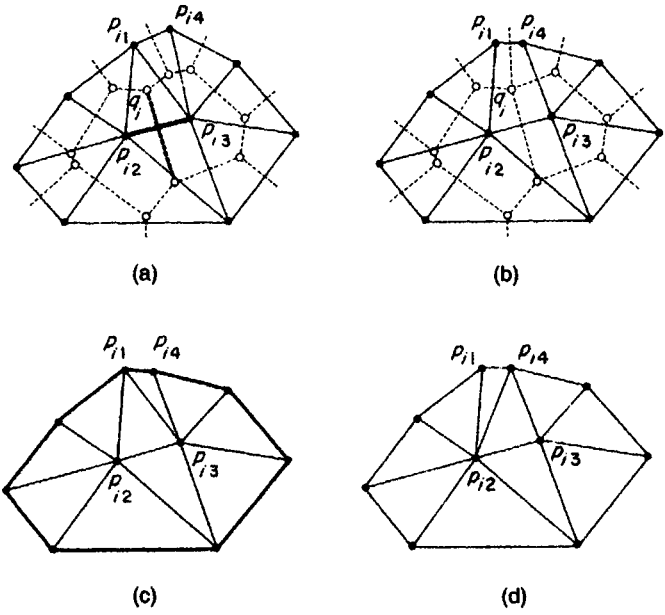


Figure 2.2.2 Voronoi diagrams (the broken lines) and triangulations (the solid lines): (a) a Delaunay triangulation; (b) a Delaunay pretriangulation; (c) a Delaunay triangulation obtained from (b); (d) another Delaunay triangulation obtained from (b).

Assumption D1 (the non-collinearity assumption) For a given set $P = \{p_1, \ldots, p_n\}$ of points, the points in P are not on the same line.

Note that Elbaz and Spehner (1990) show an approximation test for collinearity, which is useful in numerical computation with finite digits.

In addition to this assumption we assume that the number of points is three or more but finite. Note that the non-collinearity assumption implicitly implies $n \geq 3$, because two points are always on the same line. Also note that this assumption was not made in the preceding section, but it is made here because, as will be shown later, without these assumptions we cannot obtain a triangulation. Also note that we can obtain a Voronoi diagram even when $n = 2$ or when generator points are collinear (the broken lines in Figure 2.2.1(a)).

Under the non-collinearity assumption and $3 \leq n < \infty$, we now show how to obtain a Delaunay triangulation from a Voronoi diagram (the broken lines in Figure 2.2.2). First we choose a Voronoi edge in a Voronoi diagram (the heavy broken line in Figure 2.2.2(a)). This Voronoi edge is shared by two Voronoi polygons. We join the generator points (the filled circles) of these Voronoi polygons by a line segment (the heavy solid line in Figure 2.2.2(a)). We carry out this line generation with respect to all Voronoi edges in the Voronoi diagram. As a result, we obtain a second tessellation of the convex hull of the generator points (the solid lines in Figure 2.2.2(a)). In Figure