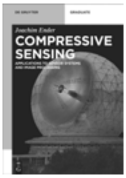
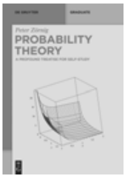
Damir N. Gainanov

Graphs for Pattern Recognition

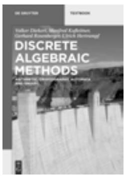
Also of Interest



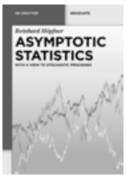
Compressive Sensing. Applications to Sensor Systems and Image Processing
Joachim Ender, 2017
ISBN 978-3-11-033531-6, e-ISBN 978-3-11-033539-2



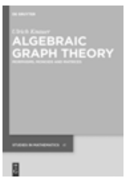
Probability Theory and Statistical Applications. A Profound Treatise for Self-Study
Peter Zörnig, 2016
ISBN 978-3-11-036319-7, e-ISBN 978-3-11-040271-1



Discrete Algebraic Methods. Arithmetic, Cryptography, Automata and Groups
Volker Diekert, Manfred Kufleitner, Gerhard Rosenberger, Ulrich Hertrampf, 2016
ISBN 978-3-11-041332-8, e-ISBN 978-3-11-041333-5



Asymptotic Statistics. With a View to Stochastic Processes
Reinhard Höpfner, 2014
ISBN 978-3-11-025024-4, e-ISBN 978-3-11-025028-2



Algebraic Graph Theory. Morphisms, Monoids and Matrices
Ulrich Knauer, 2011
ISBN 978-3-11-025408-2, e-ISBN 978-3-11-025509-6

Damir N. Gainanov

Graphs for Pattern Recognition

Infeasible Systems of Linear Inequalities

DE GRUYTER

Mathematics Subject Classification 2010
05-02, 06E30, 52-02, 68T10, 90-02

ISBN 978-3-11-048013-9
e-ISBN (PDF) 978-3-11-048106-8
e-ISBN (EPUB) 978-3-11-048030-6
Set-ISBN 978-3-11-048107-5

Library of Congress Cataloging-in-Publication Data
A CIP catalog record for this book has been applied for at the Library of Congress.

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.dnb.de.

Cover image: Islands and Bridges of Feasibility in the Infeasibility Universe, 2016, by Andrey L. Kopyrin, Ekaterinburg, Russia
Typesetting: le-tex publishing services GmbH, Leipzig
Printing and binding: CPI books GmbH, Leck
♾ Printed on acid-free paper
Printed in Germany

www.degruyter.com

Preface

This book deals with mathematical constructions that are foundational in such an important area of *data mining* as *pattern recognition*. A closer look is taken at *infeasible systems of linear inequalities*, whose generalized solutions act as building blocks of geometric decision rules for recognition.

Infeasible systems of linear inequalities proved to be a key object in pattern recognition problems described in geometrical terms thanks to *the committee method*.

Infeasible systems of inequalities represent an important special subclass of *infeasible systems of constraints with monotonicity property* – systems whose multi-indices of feasible subsystems form *abstract simplicial complexes* (*independence systems*), fundamental objects of combinatorial topology. In discrete mathematics, the faces of such complexes are interpreted as zeros of *monotone Boolean functions*. Chapter 1 of the book deals with simplicial complexes and monotone Boolean functions related to common infeasible systems of constraints. The graph-theoretic methods represent a very productive way to study combinatorial and structural properties of infeasible systems of constraints. From the applied point of view, the most important property is the *connectedness* of a specific graph assigned to a family of *maximal feasible subsystems*. For instance, the set of solutions taken one by one for each of the maximal feasible subsystems of an infeasible system, which constitute an *odd cycle* in such a graph, represents a committee for an infeasible system of linear inequalities over $\mathbb{R}^n$ formally describing a pattern recognition problem. Thus, graph-theoretic methods that help us to solve one of the main tasks of committee theory – searching for a committee with the minimal number of elements can be taken as a basis for efficient algorithms of constructing decision rules for pattern recognition. The connectedness of graphs discussed is actually determined by the connectedness of the space $\mathbb{R}^n$; moreover, the connectedness of similar graphs in the context of common topological spaces is also determined by the connectedness of these spaces. The subject matter of Chapter 2 is (hyper)graphs corresponding to facets of common simplicial complexes and to maximal feasible subsystems of infeasible systems of linear inequalities.

Equally interesting results are obtained from an analysis of infeasible systems of linear inequalities by methods of combinatorial geometry. In Chapter 3, the notion of *diagonal* of a polytope, which is traditional for plane geometry, is generalized to multi-dimensional convex polytopes. A dual correspondence between diagonals and facets of polytopes, on the one hand, and multi-indices of maximal feasible and minimal infeasible subsystems of inequalities, on the other hand, is described. This duality is used, in particular, to obtain different estimates of the number of subsystems.

In Chapter 4, the correspondence between infeasible systems of inequalities and monotone Boolean functions motivates us to construct *algorithms for optimal inference* of functions. Several criteria for optimality of algorithms of inference are considered, and algorithms satisfying these criteria are constructed.

In Chapter 5, the algorithmic approach to constructing an optimal committee of an infeasible system of linear inequalities is considered; it is based on such principal features of graphs as the connectedness and the existence of odd cycles. A brief review of *alternative covers* in the second half of this chapter provides a new look at collective solutions to infeasible systems of constraints.

The aim of this book is to present a mathematical toolset finding an application to the construction of pattern recognition complexes that solve the recognition problem in its geometric setting.

Such complexes of pattern recognition start their work with preprocessing of a training sample, that is, a massive collection of vectors from a high-dimensional feature space. Because the vectors of the training sample are preliminarily divided into groups that partially represent logically uniform classes or categories, they reflect a certain knowledge domain in the boundaries of which every new unclassified vector entering into the complex must be referred to one of the classes. At consecutive stages of preprocessing, the groups from the training sample are aggregated, with the use of hierarchical tree-like structures, into two extended groups that partially represent the corresponding generalized classes. The task of the recognition complex consists in the search for a geometric object that has a relatively simple formal description and, at the same time, strictly separates the vectors from distinct extended groups of the training sample. In the context of the book, the above-mentioned task can be interpreted, for example, as the search for a separating hyperplane in an Euclidean feature space. In practice, information contained in almost any training sample leads to a situation where a unique separating hyperplane cannot be found, because the linear inequality system underlying the problem of the discrimination of the two extended groups turns out to be infeasible. By means of some dimensional increase of the input data, the inequalities become homogeneous; their strictness is motivated by the stability demands that must be satisfied by the decision rules generated by the pattern recognition complex. This is how the infeasible system of homogeneous strict linear inequalities comes to the stage in the contradictory two-class pattern recognition problem, which has to be solved by the complex. The system as a whole has no solution, but any of its feasible subsystems can be solved by the software of the recognition complex that implements modern powerful techniques of linear optimization. The smart committee strategy of the recognition complex consists in the finding of solutions to a few maximal feasible subsystems and in their combining into a committee decision rule which operates with arrangements of separating hyperplanes. On the one hand, such a rule always allows the complex to correctly discriminate the vectors from the two extended groups of the training sample and, on the other hand, it makes it possible to apply the procedure of committee voting to a new vector entering into the complex; the majority decision rule, governed by the committee, refers the new vector to a generalized class. The recognition complex implements various effective techniques for constructing the separating committees, by exploiting specific properties of the (hyper)graphs of the maximal feasible subsystems of infeasible systems of linear inequalities. With the help of these techniques, the complex repeatedly solves the two-class pattern recognition problem for each higher level extended group of vectors from the training sample, adding at every step some committee decision rule to a resulting hierarchical tree-like structure. This structure represents the machine for recognition of new vectors, and it correctly recognizes any vector of the training sample.

This edition is the extended translation of the book *Combinatorial Geometry and Graphs in an Analysis of Infeasible Systems and Pattern Recognition* published by *Nauka*, Moscow, in 2014.

Moscow and *Ekaterinburg* — *Damir N. Gainanov*

October 2016

Contents

Pattern recognition, infeasible systems of linear inequalities, and graphs

Full-function complexes of pattern recognition allow a human or technological user to mine relevant feature data in two main directions that can be considered interconnected, depending on the goals that must be achieved by the complexes.

One direction of data mining in pattern recognition is most often referred to as *unsupervised learning*. The complex deals with a massive collection of vectors whose components represent qualitative or quantitative descriptions of various parameters that are specific for the problem domain of the user. Although some categorical labels could have preliminarily been assigned to those vectors, in order to reflect a knowledge of the domain, the complex treats the data without using any earlier classifying information. Instead, the task consists in an exploratory analysis of the massive amount of high-dimensional vectors from the feature space, which aims at the elucidation of the inner structure of the data cloud. Typically, one is interested in how many relatively dense and isolated subclouds, called *clusters*, can be discovered in the whole data cloud, and how each of them can be given a concise characterization in working terms of the problem domain.

Various strong mathematical mechanisms, as well as heuristics, are involved for preprocessing the input sample of vectors and obtaining a resulting hierarchical picture of the data cloud. Let us mention just two questions that must be answered by the designers of a complex of pattern recognition. How incomplete or missing information on the components of vectors from the feature space should be dealt with? Is there any possibility to artificially decrease the complexity of the data sample by means of an information-preserving map of the sample into a derived feature space of much lower dimension? It is clear that for obtaining concise descriptions of relatively isolated data subclouds, outermost vectors, say the vectors lying on the boundaries of the convex hulls of the subclouds, are most relevant; for this reason certain methods of thinning irrelevant vectors may be provided.

The essential topics in unsupervised learning are the choice of *metrics* that allow the recognition complex to measure the *similarity* or *distance* between vectors and between *clusters* of vectors, and the choice of the presentation format for the cluster hierarchy revealed to the user. It is convenient to visualize the hierarchy with the help of interactively scaled tree-like graphical structures that make it possible to easily reveal information on the cluster membership and on metric intercluster dissimilarities.

Although the exploratory cluster analysis surely plays an important role in data mining, the result of unsupervised learning of the recognition complex should consist in the generation of decision rules, which would allow the complex to refer any new vector of the feature space to a large isolated cluster, thus recognizing the new vector as a representative of a certain category. Such a recognition rule is based on the procedure of comparison of the similarities or distances between the new unclassified vector and the large isolated clusters.

The aim of this book is to present a mathematical toolset finding an application to the construction of pattern recognition complexes that solve the recognition problem, in its *geometric setting*, in the *supervised learning* mode.

Such complexes of pattern recognition begin their work with preprocessing of a *training sample*, that is, a massive collection of vectors from a high-dimensional feature space that are preliminarily divided into groups that partially represent logically uniform *classes* or *categories*. These groups reflect a certain knowledge domain in the boundaries of which every new unclassified vector entering into the complex must be referred to one of the classes.

The variety of approaches to supervised recognition learning includes such universally accepted methodologies as *nearest-neighbor classifiers*, *neural networks*, and *support vector machines*.

At consecutive stages of preprocessing, the groups from the training sample are aggregated, with the use of hierarchical tree-like structures, into two extended groups that partially represent the corresponding generalized classes.

Given an odd integer k, a *k-nearest-neighbor classifier* finds, for a new unclassified vector from the feature space, its k distinct nearest neighbors from the training sample; a majority of these neighbors belongs to one of the extended groups and, as a consequence, that group votes for the referring of the vector to the generalized class represented by the group. A hierarchically organized procedure of making similar k-nearest-neighbor decisions, that is applied to each of the extended subgroups of the training sample, allows the complex to recognize the new vector as a representative of the class partially described by a group from the training sample.

Dealing with an extended subgroup of vectors from the training sample, which is, in turn, divided into two subgroups at some stage of a hierarchical learning process, a *neural network* represents a collection of interconnected layers of neurons. *Neurons* are elementary computational operators that reflect vectors of the feature space to weighted values of a *sigmoid function* taken at certain weighted sums of the components of those vectors. As the result of supervised training, the neural network combines the responses of individual neurons into a decision, based on a mechanism of *thresholds*, which refers a new unclassified vector to some generalized subclass.

A *support vector machine* tries to find, at a step of a hierarchically organized procedure, three parallel hyperplanes of the feature space, namely the *maximal-margin hyperplane* which separates the vectors of two subgroups from the training sample and, at the same time, maximizes the distance between two *margin hyperplanes* containing the nearest vectors of the training sample that belong to different subgroups. The quadratic optimization technique allows the recognition complex to find the maximal-margin hyperplanes (when training subgroups are affinely separable) or to motivate the search for nonlinear separating surfaces (when the subgroups cannot be separated by hyperplanes). The hierarchical collection of the separating hyperplanes and surfaces makes it possible to refer new unclassified vectors from the feature space to some classes partially represented by the vectors of the training sample.

Thus, the task of the recognition complex that implements a supervised learning methodology often consists in the search for a geometric object that has a relatively simple formal description and, at the same time, strictly separates the vectors from distinct extended groups of the training sample.

In the context of the book, the above-mentioned task can be seen as the search for a separating hyperplane in an Euclidean feature space. In practice, information contained in almost any training sample leads to a situation where a unique separating hyperplane cannot be found, because the linear inequality system underlying the problem of the discrimination of the two extended groups turns out to be infeasible. Indeed, let $\widetilde{\boldsymbol{B}}$ and $\widetilde{\boldsymbol{C}}$ be the two extended groups of vectors from the training sample, processed at some step of the hierarchical supervised learning procedure. These are just two finite sets of vectors of the feature space $\mathbb{R}^{n-1}$. Let us augment every vector from the sets $\widetilde{\boldsymbol{B}}$ and $\widetilde{\boldsymbol{C}}$ by a new nth component which is equal to 1. We thus obtain two sets $\boldsymbol{B}, \boldsymbol{C} \subset \mathbb{R}^n$, for which the recognition complex tries to find a vector $\mathbf{x} \in \mathbb{R}^n$ such that

$$\begin{cases} \langle \boldsymbol{b}, \mathbf{x} \rangle > 0, & \boldsymbol{b} \in \boldsymbol{B}\,, \\ \langle \boldsymbol{c}, \mathbf{x} \rangle < 0, & \boldsymbol{c} \in \boldsymbol{C}\,, \end{cases} \tag{1}$$

where $\langle \boldsymbol{b}, \mathbf{x} \rangle$ denotes the standard scalar product $\sum_{k\in[n]} b_{ik}\mathrm{x}_k$, and $[n] := \{1, 2, \ldots, n\}$. The *strictness* of these *homogeneous* inequalities is motivated by the stability demands that must be satisfied by the decision rules generated by the pattern recognition complex.

If $\boldsymbol{x}$ is a solution to system (1), then classification of a new vector $\boldsymbol{g} \in \mathbb{R}^n$ (i.e., the referring of $\boldsymbol{g}$ to one of the extended classes partially represented by the sets $\boldsymbol{B}$ and $\boldsymbol{C}$) is performed on the basis of the sign of the scalar product $\langle \boldsymbol{x}, \boldsymbol{g} \rangle$. However, the system under consideration can turn out to be infeasible, and this most frequent case requires the development of special methods of problem-solving.

Even if system (1) as a whole has no solution, any of its feasible subsystems can be solved by the software of the recognition complex that implements techniques of linear optimization.

By means of the passage from system (1) to the infeasible system

$$\begin{cases} \langle \boldsymbol{b}, \mathbf{x} \rangle > 0, & \boldsymbol{b} \in \boldsymbol{B}\,, \\ \langle -\boldsymbol{c}, \mathbf{x} \rangle > 0, & \boldsymbol{c} \in \boldsymbol{C}\,, \end{cases}$$

which we will briefly describe here as the system

$$\{\langle \boldsymbol{a}, \mathbf{x} \rangle > 0 \colon\ \boldsymbol{a} \in \boldsymbol{A}\}\,, \tag{2}$$

the recognition complex deals with the mathematical construction that has the principal feature: if any subsystem, with two inequalities, of system (2) is feasible, then this simple condition guarantees that the recognition complex can involve in its computational arsenal a powerful technique for constructing certain *collective solutions* to infeasible system (2), and further use them as the components of hierarchical decision rules for recognition.

Recall that a *committee* of infeasible system (2) is defined as a finite subset of vectors $\mathcal{K} \subset \mathbb{R}^n$ satisfying the relation

$$|\{\boldsymbol{x} \in \mathcal{K} : \ \langle \boldsymbol{a}, \boldsymbol{x} \rangle > 0\}| > \tfrac{1}{2}|\mathcal{K}| ,$$

for each vector $\boldsymbol{a} \in \boldsymbol{A}$.

Suppose that a committee $\mathcal{K}$ of system (2) is found by the recognition complex. Then an unclassified vector of the feature space $\mathbb{R}^{n-1}$, lifted to the working $(n-1)$-dimensional affine subspace of the space $\mathbb{R}^n$ with the help of the additional nth component 1, can be recognized as an element of the classes, partially represented by the sets $\widetilde{\boldsymbol{B}}$ and $\widetilde{\boldsymbol{C}}$, according to the result of the majority voting procedure performed by the members of the committee $\mathcal{K}$.

The smart committee strategy of the recognition complex consists in the finding of solutions to a few *maximal feasible subsystems* (*MFSs*) of system (2), and in their combining into the committee decision rule, which operates with arrangements of separating hyperplanes.

A feasible subsystem of infeasible system (2) is called *maximal* if any additional inequality from the system turns the resulting collection of inequalities into an infeasible subsystem.

If $[m]$ is the set of indices with which the inequalities from infeasible system (2) are marked, then a *multi-index* $T \subseteq [m]$ corresponds to the subsystem composed of the inequalities with the indices from the set T.

If we let $\mathbf{J}$ denote the family of the multi-indices of all maximal feasible subsystems of system (2), then the *graph of MFSs* of system (2) is defined as the graph with the vertex set $\mathbf{J}$; an unordered pair $\{J, J'\} \subset \mathbf{J}$ is an edge of this graph if and only if the multi-indices J and J' cover the index set of system (2), that is, $J \cup J' = [m]$.

The high efficiency of supervised learning algorithms implemented by the recognition complex, which uses the graph of MFSs, is explained by the following three basic facts:

- The graph of MFSs is *connected*.
- The graph of MFSs is *not bipartite*.
- The complement $[m] - J$ of the multi-index $J \in \mathbf{J}$ of any *MFS* of system (2) is the multi-index of a *feasible subsystem*.

Since the graph of MFSs is not bipartite, it contains at least one *cycle of odd length*.

A fundamental result in the committee theory is formulated as follows: if the multi-indices of some MFSs represent the vertex set of a cycle of odd length in the graph of MFSs, then in order to construct a committee, it suffices to take one vector from the open cone of solutions to each MFS from the vertex set of the cycle.

Thus, the problem of constructing a committee with a small number of members can be reduced to the problem of finding a cycle of short odd length in the graph of MFSs. This derived problem is solved by the software of the recognition complex with the help of various strong and heuristic methods.

On the one hand, the obtained committee decision rule always allows the recognition complex to correctly discriminate the vectors from the two extended groups of the training sample and, on the other hand, it makes it possible to apply the procedure of committee voting to a new vector entering into the complex; the majority decision rule, governed by the committee, refers the new vector to a generalized class.

The complex repeatedly solves the two-class pattern recognition problem for each higher level extended group of vectors from the training sample, adding at every step some committee decision rule to a resulting hierarchical tree-like structure.

This structure represents the machine for recognition of new vectors, and it correctly recognizes any vector of the training sample.