Springer Undergraduate Texts in Mathematics and Technology

Springer Undergraduate Texts in Mathematics and Technology (SUMAT) publishes textbooks aimed primarily at the undergraduate. Each text is designed principally for students who are considering careers either in the mathematical sciences or in technology-based areas such as engineering, finance, information technology and computer science, bioscience and medicine, optimization or industry. Texts aim to be accessible introductions to a wide range of core mathematical disciplines and their practical, real-world applications; and are fashioned both for course use and for independent study.

More information about this series at http://www.springer.com/series/7438

Leo Liberti · Carlile Lavor

Euclidean Distance Geometry

An Introduction

Springer

Leo Liberti
CNRS LIX
École Polytechnique
Palaiseau
France

Carlile Lavor
Department of Applied Mathematics
(IMECC-UNICAMP)
University of Campinas
Campinas
Brazil

Part of the work on this book was carried out at the IBM "T.J. Watson" Research Center, Yorktown Heights NY, USA.
Part of the work on this book was carried out at the Dept. of Comp. Sci., Duke University, Durham NC, USA.

ISSN 1867-5506 ISSN 1867-5514 (electronic)
Springer Undergraduate Texts in Mathematics and Technology
ISBN 978-3-319-60791-7 ISBN 978-3-319-60792-4 (eBook)
DOI 10.1007/978-3-319-60792-4

Library of Congress Control Number: 2017943252

Mathematics Subject Classification (2010): 51K99, 51K05

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

This book is dedicated to Prof. Nelson Maculan, who showed us the ropes of Distance Geometry and brought us together.

Contents

Introduction

What do clocks, wireless devices, atoms, and submarines have in common? They move: the clocks move time forward (so to speak), wireless devices usually move on a plane (like an office floor), or at least on a two-dimensional surface (like a mountainous region), atoms move in three-dimensional space, and so do submarines. We are interested in these seemingly disparate entities when they move together: each computer on a network has a clock, wireless devices move as part of a wireless network, atoms form molecules, and the sort of submarines we look at are unmanned, and they move in fleets. A second very distinctive feature of these sets of moving entities is that it is useful to know the position of each entity with respect to the others: we want to know the absolute time of each clock in order to synchronize them, we need to trace each wireless device in the network for routing purposes, we want to find the geometrical shape of each molecule as it largely determines its function with respect to its environment (e.g., a cell), and we would like to control a fleet of unmanned submarines in order to accomplish a given mission. The third decisive common feature is that for each of these entities we can obtain estimates of some of their pairwise distances. The fundamental question underlying this book is the following: *given a subset of pairwise distances and the dimension of the surrounding space* (the one-dimensional time line for clocks, the two-dimensional plane for wireless devices, and the three-dimensional space for atoms and submarines), *can we find positions for all entities yielding the given distances in the space of the given dimension?*

Throughout this book, we look at the static version of this problem: namely, we suppose we can access a "snapshot" of some of the distances, including values and incidence to entities, and we want to compute corresponding spatial positions for all the entities. This is an accurate representation for the clock synchronization problem, where distances correspond to time discrepancies between pairs of clocks, as well as for the protein conformation problem, as long as we pretend that atoms do not vibrate too much that the protein is at rest and that we can trust the nuclear magnetic resonance experiments to provide some of the interatomic distances. Proteins move in space, but the relative positions of its atoms are often unchanged, i.e., they often display a *rigid motion*. Wireless sensors networks and fleets of unmanned submarines move in space, but rarely do they keep all of their pairwise distances fixed. One way to deal with the issue is to make sure that the algorithms that compute the positions from the given distances work very efficiently (so they can run in a time-step).

In abstract terms, we are given a weighted graph and a number K of dimensions, and we want to find positions for the vertices in $\mathbb{R}^K$ such that the edges, drawn as segments, have lengths equal to the corresponding weights. This is the fundamental problem of a field called *Euclidean Distance Geometry*. The name "Distance Geometry" refers to a concept of geometry based on distances rather than points. As for the word "Euclidean," it has something to do with our choice of drawing edges as segments, without constraining their incidence angles. If we had decided to draw edges only using vertical and horizontal segments incident at right angles, the word "Euclidean" might have been replaced by "taxicab" or "Manhattan"—but this setting has fewer applications.

An issue which appears to be quite well studied in Euclidean Distance Geometry is the uniqueness of the solution. This feature is obviously very desirable in the wireless network and submarine fleet applications, since our objective is to recover the actual position in space of devices and submarines, which cannot help but occupy a single position in space. With proteins, however, which are really our "pet application," this issue is somewhat less important. Molecules naturally come in different isomers, which means that the same chemical composition can give rise to multiple shapes. Many interesting isomers involve a difference in *chirality*, i.e., two molecules can be geometric reflections of each other. With proteins, which consist of a backbone with some side chains appended, the reflections can also be *partial*: Two isomeric backbones have the same shape until atom $v - 1$, and then the part of the backbone from atom v onwards is a reflection with respect to the plane defined by three atoms before v. These simple geometric operations can make the difference between a safe medicine and a deadly poison. Therefore, it makes sense to work out the shape of all of the proteins that are consistent with the observed interatomic distances. Rather than solution uniqueness, we are interested in finding all solutions.

This book is aimed at intermediate undergraduate students, starting graduate students, any researcher who would like to know something about the theory and algorithms used in Euclidean Distance Geometry, and any practitioner who needs to actually compute positions of entities for which he or she knows some of the pairwise distances. Its objective is to teach the basics of this field, without going in too many details and yet providing readers with some useful methodologies as well as with a sense of *why* they work.

Note that *this is a textbook, rather than a research monograph*. The authors have taught Ph.D.-level courses about this material, but also drew on it to teach B.Sc. and M.Sc. courses in other fields, such as Mathematical Programming. We strove to keep the book short; proofs, when present, are given informally within the text and are often complemented by pictures. All algorithms in this book (and many more) have been implemented in the *Mathematica* computer programming language and are available online, at[1]

http://www.lix.polytechnique.fr/~liberti/intro_dg

Each chapter ends with some exercises. We added a schematic appendix containing all necessary preliminary mathematical notions.

Incredibly, this book is the first of its kind, i.e., it is the first (teaching oriented) textbook on Euclidean Distance Geometry. The first book on the subject of Distance Geometry was written by Blumenthal in 1953 [17], with entirely different objectives: it is part research monograph and part compendium of a budding field. Distance Geometry in 1953 focused—due to the almost total lack of computers at the time—on very different issues than it does today. We also note that Blumenthal wrote a short didactical textbook on Euclidean geometry in 1961 [18], to which he added a chapter on Distance Geometry that borders on the same topics of the 1953 book. A mathematically advanced research monograph on issues similar to those treated here was written by Crippen and Havel in 1988 [31], a book which is now out of print and very hard to find. Like this book, [31] is also motivated by finding the shape of proteins. The most recent research-oriented book on this subject was written by Dattorro around 2004 [33], who keeps the book continually updated: [33] is mostly an online book, although the author will print, bind, and send you a copy (for a fee to cover costs) if you ask for one. Dattorro's book is also about convex optimization and therefore takes an approach that is complementary to ours. The most immediately remarkable difference between [33] and this book is in the

[1] If you are reading this chapter as a PDF on your screen, copying and pasting the URL is likely to result in some wrong characters (particularly the tilde and the underscore).

writing style. Dattorro's book is an almost encyclopedic research monograph: commended for the specialists in the field who need all the details at their fingertips, but perhaps not ideal for the budding Euclidean Distance geometer. Some books about graph rigidity [57, 118] and oriented matroids [16] also cover some of the subjects we discuss. We think that our own surveys [80, 84, 86] nicely complement the material in this book. We also cowrote a very short didactical monograph (directed mostly at starting undergraduates and even finishing high-school students) which covers some of the topics of this book, mostly restricted to the 2D plane and 3D space [77].

The reason why we write "incredibly," in noting this as the first teaching book on Euclidean Distance Geometry, is that the subject itself is as old as the ancient Greeks and involved mathematicians such as Heron, Euler, Cauchy, Cayley, and Gödel [81], to name only the most famous. This field also produced methods such as classic Multidimensional Scaling and the Johnson–Lindenstrauss lemma, both widely used today in the context of the "big data" revolution. It contributed to win the 2002 Nobel Prize in Chemistry to Wütrich [121] "for his development of nuclear magnetic resonance spectroscopy for determining the three-dimensional structure of biological macromolecules in solution." From our point of view, there should be *hundreds* of teaching books on Distance Geometry around!

Our own interest in this field started in 2004, when one of us, then working at Politecnico di Milano, paid a visit to the other, then at the Universidade Estadual do Rio de Janeiro. That visit yielded a modest book chapter [78], after which we scratched our heads and asked ourselves, "what now"? Twelve years and many dozens of research papers later, we realize we only just started to brush the surface. Humbled by the gargantuan size of the task ahead, we decided to spend some of our time teaching other people what we know, in the hope of advancing this wonderful field. This book is the result of our teaching efforts.

New York, Paris — Leo Liberti
Campinas, Durham — Carlile Lavor
2013–2017