IAS/Park City Mathematics Series
Volume 25, Pages -9—12
https://doi.org/10.1090/pcms/025/00828

# Introduction

Michael W. Mahoney, John C. Duchi, and Anna C. Gilbert

"The Mathematics of Data" was the topic for the 26th annual Park City Mathematics Institute (PCMI) summer session, held in July 2016. To those more familiar with very abstract areas of mathematics or more applied areas of data—the latter going these days by names such as "big data" or "data science"—it may come as a surprise that such an area even exists. A moment's thought, however, should dispel such a misconception. After all, data must be modeled, e.g., by a matrix or a graph or a flat table, and if one performs similar operations on very different types of data, then there is an expectation that there must be some sort of common mathematical structure, e.g., from linear algebra or graph theory or logic. So too, ignorance or errors or noise in the data can be modeled, and it should be plausible that how well operations perform on data depend not just on how well data are modeled but also on how well ignorance or noise or errors are modeled. So too, the operations themselves can be modeled, e.g., to make statements such as whether the operations answer a precise question, exactly or approximately, or whether they will return a solution in a reasonable amount of time.

As such, "The Mathematics of Data" fits squarely in applied mathematics—when that term is broadly, not narrowly, defined. Technically, it represents some combination of what is traditionally the domain of linear algebra and probability and optimization and other related areas. Moreover, while some of the work in this area takes place in mathematics departments, much of the work in the area takes place in computer science, statistics, and other related departments. This was the challenge and opportunity we faced, both in designing the graduate summer school portion of the PCMI summer session, as well as in designing this volume. With respect to the latter, while the area is not sufficiently mature to say the final word, we have tried to capture the major trends in the mathematics of data sufficiently broadly and at a sufficiently introductory level that this volume could be used as a teaching resource for students with backgrounds in any of the wide range of areas related to the mathematics of data.

The first chapter, "Lectures on Randomized Numerical Linear Algebra," provides an overview of linear algebra, probability, and ways in which they interact fruitfully in many large-scale data applications. Matrices are a common way to model data, e.g., an $m \times n$ matrix provides a natural way to describe $m$ objects, each of which is described by $n$ features, and thus linear algebra, as well as more sophisticated variants such as functional analysis and linear operator theory, are central to the mathematics of data. An interesting twist is that, while work in numerical linear algebra and scientific computing typically focuses on deterministic algorithms that return answers to machine precision, randomness can be used in novel algorithmic and statistical ways in matrix algorithms for data. While randomness is often assumed to be a property of the data (e.g., think of noise being modeled by random variables drawn from a Gaussian distribution), it can also be a powerful algorithmic resource to speed up algorithms (e.g., think of Monte Carlo and Markov Chain Monte Carlo methods), and many of the most interesting and exciting developments in the mathematics of data explore this algorithmic-statistical interface. This chapter, in particular, describes the use of these methods for the development of improved algorithms for fundamental and ubiquitous matrix problems such as matrix multiplication, least-squares approximation, and low-rank matrix approximation.

The second chapter, "Optimization Algorithms for Data Analysis," goes one step beyond basic linear algebra problems, which themselves are special cases of optimization problems, to consider more general optimization problems. Optimization problems are ubiquitous throughout data science, and a wide class of problems can be formulated as optimizing smooth functions, possibly with simple constraints or structured nonsmooth regularizers. This chapter describes some canonical problems in data analysis and their formulation as optimization problems. It also describes iterative algorithms (i.e., those that generate a sequence of points) that, for convex objective functions, converge to the set of solutions of such problems. Algorithms covered include first-order methods that depend on gradients, so-called accelerated gradient methods, and Newton's second-order method that can guarantee convergence to points that approximately satisfy second-order conditions for a local minimizer of a smooth nonconvex function.

The third chapter, "Introductory Lectures on Stochastic Optimization," covers the basic analytical tools and algorithms necessary for stochastic optimization. Stochastic optimization problems are problems whose definition involves randomness, e.g., minimizing the expectation of some function; and stochastic optimization algorithms are algorithms that generate and use random variables to find the solution of a (perhaps deterministic) problem. As with the use of randomness in Randomized Numerical Linear Algebra, there is an interesting synergy between the two ways in which stochasticity appears. This chapter builds the necessary convex analytic and other background, and it describes gradient and subgradient first-order methods for the solution of these types of problems. These methods tend to be simple methods that are slower to converge than more advanced methods—such as Newton's or other second-order methods—for deterministic problems, but they have the advantage that they can be robust to noise in the optimization problem itself. Also covered are mirror descent and adaptive methods, as well as methods for proving upper and lower bounds on such stochastic algorithms.

The fourth chapter, "Randomized Methods for Matrix Computations," goes into more detail on randomized methods for computing efficiently a low-rank approximation to a given matrix. One often wants to decompose a large $m \times n$ matrix $A$, where $m$ and $n$ are both large, into two lower-rank more-rectangular matrices $E$ and $F$ such that $A \approx EF$. Examples include low-rank approximations to the eigenvalue decomposition or the singular value decomposition. While low-rank approximation problems of this type form a cornerstone of traditional applied mathematics and scientific computing, they also arise in a broad range of data science applications. Importantly, though, the questions one asks of these matrix decompositions (e.g., whether one is interested in numerical precision or statistical inference objectives) and even how one accesses these matrices (e.g., within the RAM model idealization or in a single-pass streaming setting where the data can't even be stored) are very different. Randomness can be useful in many ways here. This chapter describes randomized algorithms that obtain better worst-case running time, both in the RAM model and a streaming model, how randomness can be used to obtain improved communication properties for algorithms, and also several data-driven decompositions such as the Nyström method, the Interpolative Decomposition, and the CUR decomposition.

The fifth chapter, "Four Lectures on Probabilistic Methods for Data Science," describes modern methods of high dimensional probability and illustrates how these methods can be used in data science. Methods of high-dimensional probability play a central role in applications for statistics, signal processing, theoretical computer science, and related fields. For example, they can be used within a randomized algorithm to obtain improved running time properties, and/or they can be used as random models for data, in which case they are needed to obtain inferential guarantees. Indeed, they are used (explicitly or implicitly) in all of the previous chapters. This chapter presents a sample of particularly useful tools of high-dimensional probability, focusing on the classical and matrix Bernstein's inequality and the uniform matrix deviation inequality, and it illustrates these tools with applications for dimension reduction, network analysis, covariance estimation, matrix completion, and sparse signal recovery.

The sixth and final chapter, "Homological Algebra and Data," provides an example of how methods from more pure mathematics, in this case topology, might be used fruitfully in data science and the mathematics of data, as outlined in the previous chapters. Topology is—informally—the study of shape, and topological data analysis provides a framework to analyze data in a manner that should be insensitive to the particular metric chosen, e.g., to measure the similarity between data points. It involves replacing a set of data points with a family of simplicial complexes, and then using ideas from persistent homology to try to determine the large scale structure of the set. This chapter approaches topological data analysis from the perspective of homological algebra, where homology is an algebraic compression scheme that excises all but the essential topological features from a class of data structures. An important point is that linear algebra can be enriched to cover not merely linear transformations—the 99.9% use case—but also sequences of linear transformations that form complexes, thus opening the possibility of further mathematical developments.

Overall, the 2016 PCMI summer program included minicourses by Petros Drineas, John Duchi, Cynthia Dwork and Kunal Talwar, Robert Ghrist, Piotr Indyk, Mauro Maggioni, Gunnar Martinsson, Roman Vershynin, and Stephen Wright. This volume consists of contributions, summarized above, by Petros Drineas (with Michael Mahoney), Stephen Wright, John Duchi, Gunnar Martinsson, Roman Vershynin, and Robert Ghrist. Each chapter in this volume was written by a different author, and so each chapter has it's own unique style, including notational differences, but we have taken some effort to ensure that they can fruitfully be read together.

Putting together such an effort—both the entire summer session as well as this volume—is not a minor undertaking, but for us it was not difficult, due to the large amount of support we received. We would first like to thank Richard Hain, the former PCMI Program Director, who first invited us to organize the summer school, as well as Rafe Mazzeo, the current PCMI Program Director, who provided seamless guidance throughout the entire process. In terms of running the summer session, a special thank you goes out to the entire PCMI staff, and in particular to Beth Brainard and Dena Vigil as well as Bryna Kra and Michelle Wachs. We received a lot of feedback from participants who enjoyed the event, and Beth and Dena deserve much of the credit for making it run smoothly; and Bryna and Michelle's role with the graduate steering committee helped us throughout the entire process. In terms of this volume, in addition to thanking the authors for their efforts and (usually) getting back to us in a timely manner, we would like to thank Ian Morrison, who is the PCMI Publisher. Putting together a volume such as this can be a tedious task, but for us it was not, and this is in large part due to Ian's help and guidance.