The Book Review Column[1]

by Frederic Green

Department of Mathematics and Computer Science
Clark University
Worcester, MA 01610
email: fgreen@clarku.edu

In this column, we review these three books:

1. **Algorithmic Aspects of Machine Learning**, by Ankur Moitra. A succinct book about theoretical aspects of ML. Review by Sarvagya Upadhyay.
2. **Network Flow Algorithms**, by David Williamson. An examination of many aspects of these important algorithms. Review by S.V. Nagaraj.
3. **The Theory of Quantum Information**, by John Watrous. A new, unified treatment of this vital area. Review by Steve Fenner.

As always, I'm on the lookout for reviewers. *SIGACT News WANTS YOU!!* Please choose from among the books listed on the next page…or not. These are mostly *suggestions*! Please feel free to suggest an appropriate title of your own. Indeed, many of those listed include books I don't have on hand, and can ask the publisher to forward to you. The latter method remains preferable in our present global predicament.

In Memoriam: I was deeply saddened to learn of the recent passing of Alan Selman, a leading light in computational complexity theory. I am very lucky to count Alan as one of my mentors and co-authors. On a personal level, he was especially supportive to me in my early days as a computer scientist. I will miss his warm, friendly greetings and conversations at conferences and visits. You can read more from his many colleagues in Lane Hemaspaandra's column in this issue.

BOOKS THAT NEED REVIEWERS FOR THE SIGACT NEWS COLUMN

Algorithms

1. *The Algorithm Design Manual*, by Steven S. Skiena
2. *Algorithms and Data Structures*, by Helmut Knebl
3. *Beyond the Worst-Case Analysis of Algorithms*, by Tim Roughgarden

Computability, Complexity, Logic

1. *Applied Logic for Computer Scientists: Computational Deduction and Formal Proofs*, by Mauricio Ayala-Rincón and Flávio L.C. de Moura.
2. *Descriptive Complexity, Canonisation, and Definable Graph Structure Theory*, by Martin Grohe.
3. *Mathematics in Computing*, by Gerard O'Regan.
4. *Semigroups in Complete Lattices*, by P. Eklund, J. Gutiérrez García, U. Höhle, and J. Kortelainen.

Miscellaneous Computer Science

1. *Elements of Causal Inference: Foundations and Learning Algorithms*, by Jonas Peters, Dominik Janzing, and Bernhard Schölkopf.
2. *Partially Observed Markov Decision Processes,* by Vikram Krishnamurthy
3. *Statistical Modeling and Machine Learning for Molecular Biology*, by Alan Moses
4. *Language, Cognition, and Computational Models,* Theirry Poibeau and Aline Villavicencio, eds.
5. *Computational Bayesian Statistics, An Introduction,* by M. Antónia Amaral Turkman, Carlos Daniel Paulino, and Peter Müller.
6. *Variational Bayesian Learning Theory,* by Shinichi Nakajima, Kazuho Watanabe, and Masashi Sugiyama.
7. *Knowledge Engineering: Building Cognitive Assistants for Evidence-based Reasoning*, by Gheorghe Tecuci, Dorin Marcu, Mihai Boicu, and David A. Schum.
8. *Quantum Computing: An Applied Approach*, by Jack D. Hidary

Cryptography and Security

1. *Computer Security and the Internet: Tools and Jewels*, by Paul C. van Oorschot

Combinatorics and Graph Theory

1. *Finite Geometry and Combinatorial Applications*, by Simeon Ball
2. *Combinatorics, Words and Symbolic Dynamics,* Edited by Valérie Berthé and Michel Rigo

Programming etc.

1. *Formal Methods: An Appetizer*, by Flemming Nielson and Hanne Riis Nielson
2. *Programming for the Puzzled: Learn to Program While Solving Puzzles*, by Srini Devadas.
3. *Sequential and Parallel Algorithms and Data Structures*, by P. Sanders, K. Mehlhorn, M. Dietzfelbinger, R. Dementiev

Miscellaneous Mathematics

1. *Introduction to Probability*, by David F. Anderson, Timo Seppäläinen, and Benedek Valkó.
2. *Algebra and Geometry with Python*, by Sergei Kurgalin and Sergei Borzunov.

Review of[2]
Algorithmic Aspects of Machine Learning
By
Ankur Moitra
Published by Cambridge University Press
Paperback, 151 pages, $ 34.99

Review by
Sarvagya Upadhyay (supadhyay@fujitsu.com)
Fujitsu Laboratories of America
1240 East Arques Avenue, Sunnyvale CA 94085, USA

1 Overview

Over the past two decades, machine learning has seen tremendous development in practice. Technological advancement and increased computational resources have enabled several learning algorithms to become quite useful in practice. Although many families of learning algorithms are heuristic in nature, their usefulness cannot be understated. Empirical observations coupled with abundance of new datasets have led to development of novel algorithmic techniques that aim to accomplish a variety of learning tasks efficiently on real-world problems.

But what makes these algorithms work on such real-world problems? Clearly, producing correct solutions is one aspect of it. The other aspect is efficiency. While many of these algorithms solve hard problems and cannot be theoretically efficient (under plausible complexity-theoretic assumptions), they seemingly do work on real-world problems. It begets the question: are there conditions under which these algorithms become tractable? Having an answer to this fundamental question sheds light on the power and limitations of these algorithmic techniques.

This book focuses on different learning models and problems, and sets out to capture the assumptions that make certain algorithms tractable. The emphasis is on models and algorithmic techniques that make learning an efficient endeavor.

2 Summary of Contents

This book covers six different topics. Chapter 1 serves as a short introduction and the rest of the eight chapters are devoted to the following topics: non-negative matrix factorization, tensor decomposition, sparse recovery, sparse coding, learning mixture models, and matrix completion. A brief summary of each of the main chapters is given below.

Chapter 2: Nonnegative Matrix Factorization This chapter explores non-negative matrix factorization (NMF). Given a non-negative matrix $M \in \mathbb{R}^{n \times n}$, NMF asks for two non-negative matrices $A \in \mathbb{R}^{n \times r}$ and $W \in \mathbb{R}^{r \times n}$ such that $M = AW$. The definition closely resembles singular value decomposition (SVD); however, NMF is a hard problem in the complexity-theoretic sense. The chapter focuses on fundamental definitions concerning NMF, a heuristic algorithm used in practice (alternating minimization), and fundamental results in solving systems of polynomial equations that lead to an algorithm for accomplishing NMF. Towards the end, the chapter focuses on *topic models* and how NMF can be used to learn their parameters.

Chapter 3: Tensor Decompositions (Algorithms) Tensors are higher-order generalizations of matrices and vectors. Several questions that are raised for matrices either do not make sense for tensors or become intractable. This chapter starts with the well known *rotation problem* that arises in matrix factorization, and proceeds to show how several results in matrices do not automatically translate to tensors. One of the main sections of the chapter is devoted to Jenrich's algorithm that gives a tractable method to find tensor factorization under some assumptions (Theorem 3.3.2 in the book). The final section illustrates the utility of Jenrich's algorithm to the case when the tensor is perturbed by some noise.

Chapter 4: Tensor Decompositions (Applications) This chapter is devoted to applications of tensor decomposition. The first application considered is motivated from evolutionary biology, which describes evolutionary relationships among species: phylogenetic trees. A special case of phylogenetic tree is hidden Markov models (HMMs). The next application considered is community detection, specifically the stochastic block model. The notion of an individual belonging to multiple communities gives rise to mixed membership models, which is considered next. This is divided into two parts: pure topic models and latent Dirichlet allocation (LDA). The chapter concludes with applications of tensor decomposition in independent component analysis (ICA). In all of these applications, the crucial ingredient is Jenrich's algorithm as described in Chapter 3.

Chapter 5: Sparse Recovery This chapter focuses on the following question: Given a system of underdetermined linear equations $Ax = b$, when can we determine x uniquely under the assumption that x is sparse? The question crucially arises in signal processing, where x is the unknown signal that one wishes to recover. The answer lies in conditions imposed on the matrix A. This leads to the next section, which discusses the incoherence principle, a crucial notion that enables us to recover a sparse x exactly. The following two sections showcase different algorithms for solving the problem: (i) an example of pursuit algorithms known as the *orthogonal matching pursuit algorithm*; and (ii) a numerically unstable algorithm called Prony's method. The final section is on compressed sensing that introduces a new assumption on A called the restricted isometry property (RIP).

Chapter 6: Sparse Coding Given a collection of signals that are sparse in an unknown basis, can we still learn the signal? This question is known as sparse coding and forms the focal point of this chapter. The first section introduces two popular methodologies to tackle this problem: (i) the method of optimal directions; and (ii) k-SVD. They are heuristic approaches and can be considered as variants of the alternating minimization technique introduced in Chapter 2. For provable algorithmic guarantees, the chapter considers two separate cases: the undercomplete case and the overcomplete case. The undercomplete case is when the matrix A has full column rank. Under certain stochastic assumptions, one can recover A using convex programming relaxation. For the overcomplete case, the chapter first takes a detour and discusses gradient descent. The overcomplete case is a non-convex problem; however, under certain stochastic assumptions, A can be recovered using gradient descent.

Chapter 7: Gaussian Mixture Models Many natural statistical questions can be modeled as a linear combination of independent Gaussian distributions. Such a combination is known as a mixture of Gaussians. The question of how to learn individual Gaussian parameters from samples drawn from a mixture of Gaussians forms the crux of this chapter. The first technique considered is the method of moments that essentially solves a system of polynomial equations. However, the solution is not unique. Another approach is expectation maximization, which is briefly discussed. For provable guarantees, the first algorithm considered in this chapter is based on clustering. The following section focuses on the weakness of this approach, i.e., generating a sample where one cannot figure out which component generated it. Next, the chapter discusses clustering-free algorithms. The final two sections focus on an algorithm to learn a mixture of univariate Gaussian mixtures.

Chapter 8: Matrix Completion This chapter delves into reconstructing a matrix after observing a few entries of it. Without any restriction on the matrix, this task is impossible, for there are far too many choices. However, when the matrix is low-rank and satisfies the incoherence principle, then simple convex programming relaxation can help reconstruct the matrix uniquely. The first section sets the ground rules for the matrix to be reconstructed. Then the nuclear norm is introduced in following section. The section follows by relaxing the optimization problem for computing the nuclear norm and gives the conditions under which exact recovery is possible. The final section focuses on completing the proof of reconstruction.

3 Evaluation and Opinion

For a 150 page book, the topics covered in this monograph are varied, with a common theme of designing efficient algorithms for learning models under some assumption. Each chapter is more or less independent of the other chapters, except Chapter 4 (which depends on Chapter 3). The variety of topics covered makes the book quite dense from a technical viewpoint. It makes this book a difficult read for a beginner where a fair amount of knowledge in underlying mathematical principles is necessary and often requires pointers to references for further clarification. The book's targeted audience can be graduate students interested in theoretical aspects of machine learning algorithms.

I found the book an interesting read. While there were certain sections that weren't clear to me, the challenges to prove simple but unproven claims and delving deeper into the topics makes it a fascinating read. The exercises at the end of each chapter are few but they are pertinent and worth solving. There were certain topics about which I had tangential knowledge; and knowing a little more about them (from the author's point of interest) was great. For me, one of the best parts of the book is the introduction to each chapter. They thoroughly motivate the topic of the chapters.

Finally, these topics are worth expanding and deserve a thorough and comprehensive treatment. I hope that there is a future for subsequent editions of the book.

Review of[3]
Network Flow Algorithms
David P. Williamson
Cambridge University Press, 2019
326 pages, Paperback, ISBN 9781316636831 $39.99

Review by
S.V.Nagaraj (svnagaraj@acm.org)
VIT, Chennai Campus, India

1 Introduction

This book is on algorithms for network flows. Network flow problems are optimization problems where given a flow network, the aim is to construct a flow that respects the capacity constraints of the edges of the network, so that incoming flow equals the outgoing flow for all vertices of the network except designated vertices known as the source and the sink. Network flow algorithms solve many real-world problems. This book is intended to serve graduate students and as a reference. The book is also available in eBook (ISBN 9781316952894/US$ 32.00), and hardback (ISBN 9781107185890/US$99.99) formats. The book has a companion web site www.networkflowalgs.com where a pre-publication version of the book can be downloaded gratis.

2 Summary

The book consists of nine chapters. The first chapter begins with a prelude on shortest path algorithms. Dijkstra's algorithm for non-negative costs and the Bellman-Ford algorithm for negative costs are briefly introduced. Negative cost cycle detection is also discussed.

The second chapter is on maximum flow algorithms. The maximum flow problem, and its dual problem, the minimum s-t cut problem, are the main focus. These two problems have been useful for modeling many problems involving various types of networks. Surprisingly, these two problems have also proven useful in modeling problems that do not apparently involve networks or the flow of material. To exemplify this, three applications are considered. They are carpool sharing, the baseball elimination problem, and finding a maximum density subgraph. Other topics in this chapter include a discussion about most improving augmenting paths (the augmenting paths whose minimum residual capacity arcs are as large as possible), a capacity scaling algorithm, shortest augmenting paths, and the push-relabel algorithm. The push-relabel algorithm, also known as the preflow-push algorithm, is an algorithm for computing maximum flows in a flow network that applies local operations as opposed to (say) Ford-Fulkerson.

The third chapter is on global minimum cut algorithms. Given an undirected graph G(V,E), a global minimum cut is a partition of V into two subsets (A,B) such that the number of edges between A and B is minimized. In this chapter, the global minimum cut problem for directed graphs is also considered. The Hao-Orlin algorithm for finding a global minimum cut in directed graphs, the MA ordering algorithm for

[3] ©2021, S.V.Nagaraj

finding a global minimum cut in undirected graphs, the random contraction algorithm for finding a global minimum cut in undirected graphs, and Gomory-Hu trees are discussed. The Gomory-Hu tree is a data structure related to minimum s-t cuts of an undirected graph.

The fourth chapter discusses some more maximum flow algorithms. Blocking flows in unit-capacity graphs and the Goldberg-Rao algorithm are the key concepts studied in this chapter. The purpose of discussing these additional algorithms is to usher in one of the fastest polynomial-time algorithms known: the Goldberg-Rao algorithm[4]. A type of flow called a blocking flow is discussed and it is demonstrated how blocking flows can be used to develop polynomial-time algorithms for the maximum flow problem. The author remarks that the Goldberg-Rao algorithm was the theoretically fastest polynomial-time algorithm for the maximum flow problem for several years, however, faster algorithms using interior-point methods from linear programming are now available.

The fifth chapter is on minimum-cost circulation algorithms. The focus is on flow problems that involve a cost per unit flow, and in which the goal is to minimize the overall cost of the flow while meeting certain conditions. The idea is to develop a set of conditions that let us know when we have found a circulation of minimum cost. Wallacher's algorithm[5] is then described. In the chapter notes, the author states that the ideas from Wallacher's technical report have been influential, despite never appearing as a journal publication. A minimum-mean cycle canceling algorithm is then looked at. Capacity-scaling algorithms for the minimum-cost circulation problem are also studied. A strongly polynomial-time successive approximation framework for minimum-cost circulation algorithm is then described. The next algorithm of the chapter is for computing a minimum-cost circulation based on a variant of the simplex method for linear programming. This is possible since the minimum-cost circulation problem can be expressed as a linear program. This variant of the simplex method is commonly known as the network simplex algorithm. The chapter then shows how one can use the minimum-cost circulation to solve another flow problem, one that involves a dimension of time, viz., the maximum s-t flow problem over time.

The sixth chapter is on generalized flow algorithms. In this chapter, the discussion is about generalized flow problems; and in particular, the generalized maximum flow problem. In generalized flow problems, for each arc we also have a gain. This gain can be used to model losses on the arcs due to various reasons. There is a discussion on how we can tell whether a given proper flow f is maximum. The analog of an augmenting path for generalized flow is called a generalized augmenting path, or GAP for short. A Wallacher-style GAP-canceling algorithm is then described. A polynomial-time algorithm for generalized flow based on an adaptation of Wallacher's algorithm for the minimum-cost circulation problem is presented. An algorithm for negative-cost GAP detection is also presented. An algorithm to reduce to the case of lossy graphs, Truemper's algorithm, gain scaling, and error scaling are other topics discussed in the chapter.

The seventh chapter is on algorithms for multi-commodity flows. In the maximum flow problem, we try to transport as much of a single good/commodity as possible from the source s to the sink t. In the multi-commodity flow problem, we have multiple goods (or multiple commodities) that need to be sent between distinct sources and sinks, one source and sink per commodity. The author remarks that for the network flow

[4] A. V. Goldberg and S. Rao., Beyond the flow decomposition barrier. Journal of the ACM,45:783-797, 1998

[5] C. Wallacher. A generalization of the minimum-mean cycle selection rule in cycle canceling algorithms. Technical report, Abteilung für Optimierung, Institut für Angewandte Mathematik, Technische Universität Carolo-Wilhelmina, Braunschweig, Germany, 1991

problems discussed in the earlier chapters, nice, combinatorial statements about how to tell when the flow is optimal were provided. However, unfortunately, there are no similar theorems for the multi-commodity flow problem. The two commodity case is then studied. An algorithm that has had many applications in various fields, viz. the multiplicative weights algorithm, is then discussed. The Garg-Könemann algorithm for the maximum multi-commodity flow problem is looked at. The Awerbuch-Leighton algorithm is another multicommodity flow algorithm studied in the chapter.

The eighth chapter is on algorithms for electrical flows. This chapter reviews concepts of electrical flows, and then shows how they can be applied to computing maximum flows in undirected graphs and for making graphs sparse. The chapter presents an algorithm for computing such a flow. On certain occasions, it is useful to get fast, almost exact solutions to network flow problems. One way we can do this is to work with sparse representations of the original input to the problem. A multiplicative weights algorithm for computing an approximate s-t flow via electrical s-t flows is presented.

The ninth chapter is on open questions. This chapter concludes the book, by listing some significant open problems. Five open problems are presented:

1. A simple $O(mn)$ time maximum flow algorithm
2. A Gomory-Hu tree without $n - 1$ flow computations.
3. A strongly polynomial-time algorithm for the generalized minimum-cost circulation problem.
4. A combinatorial, polynomial-time, exact algorithm for multi-commodity flow
5. Combinatorial minimum-cost circulation algorithms as fast as interior-point algorithms

3 Opinion

The author of this book is known for his work on optimization and approximation algorithms for which he won many awards. In the preface of this book, he justifies its creation while acknowledging the existence of the book *Network Flows: Theory, Algorithms, and Applications* by Ahuja, Magnanti, and Orlin (First Edition, Pearson, 1993, ISBN 978-0136175490). He considers their book as being definitive and states that it is not easy for a book to be both definitive and succinct. He aimed for a succinct book. This book is based on courses that were taught by Williamson. He notes that results that were either too long or too complex to be covered in a single lecture were not included in the book. He admits that some parts of network flow theory such as applications or algorithms without polynomially-bounded running times are not the focus of this book. For such topics, he advises readers to refer the book by Ahuja et al. He also states that some new topics covered by this book are not covered by Ahuja et al., for example, the work of Goldberg and Rao, and Wallacher (both cited above). New polynomial-time algorithms for global minimum cut, generalized maximum flow, and multi-commodity flow problems that emerged after the publication of the book by Ahuja et al. are found in this book.

The book contains stimulating exercises and useful notes at the end of chapters. There are more than 200 references to the literature. The author index and the subject index are helpful. The open questions posed at the end of book provide enough challenges for graduate students and researchers. The book includes many lemmas and theorems with proofs. It provides a succinct, amalgamated view of a broad mixture of effective combinatorial algorithms for network flow problems, including many topics not found in other textbooks.

The book presents the latest body of work on computing electrical flows along with new applications of these flows to classical problems in network flow theory. It will certainly be a thorough textbook for a course on network flow algorithms and a handy reference for the state of the art in that field. It will be a useful supplement to the book by Ahuja et al. The book is well suited as a textbook for advanced undergraduate and graduate courses on network flows. Instructors who supplement the textbook with numerical examples, computer programming exercises, and optimization software will enhance the utility of the book and make it more useful for solving real-world problems. The author notes that since his research has largely not been on network flows, he feels he can serve as an "unbiased outside observer." At the same time the text reveals his passionate belief that these are "truly beautiful and useful algorithmic ideas that build on each other in a very aesthetically pleasing way." This also comes across in his lectures on the subject, posted shortly before this writing at `https://people.orie.cornell.edu/dpw/orie6330/`. This book can also be used for self-study by research scholars looking for thought-provoking research problems. I strongly recommend the book for students and researchers.

Review of[6]
The Theory of Quantum Information
John Watrous
Cambridge, 2018
590 pages, Hardcover, $79.99 on Cambridge Univ. Press site
eBook $64.00 on the same site

Reviewed by
Stephen A. Fenner (fenner.sa@gmail.com)
Computer Science and Engineering Department
University of South Carolina

Overview

This is an extremely clear, carefully written book that covers the most important results in the sprawling field of quantum information. It is perfect for a reference, self-study, or a graduate course in quantum information. It makes no attempt to be broad or encyclopedic, but instead goes deep into the core topics. The definitions and theorems are all precisely worded, and (starting in Chapter 2) all results have complete proofs, making the book largely self-contained. The book focuses heavily on the mathematical results and nuts-and-bolts techniques underpinning current research, and as such gives the reader a thorough and flexible toolkit for proving new results. If you are just looking for a broad but cursory survey of the field, then this is probably not the book for you. If, however, you want a working knowledge of the core results and proof techniques of quantum information with an eye toward doing cutting-edge research in the field, then this book will be an indispensable addition to your library.

The mathematical theory of quantum information studies the ultimate abilities and limits of transmitting and processing information using the laws of quantum mechanics. It owes much of its motivation to classical information theory, which was largely developed by Claude Shannon in the mid 20th century, and to quantum mechanics itself (of course). It addresses basic questions like: how much information can be transmitted through quantum channels, noisy or otherwise, and how entanglement helps. The theory informs, and is informed by, its sister disciplines of quantum computation and quantum communication (which overlap with physics and computer science), although in some sense it is more fundamental. Though he occasionally mentions applications to these other areas, Watrous seats his book squarely in the realm of pure mathematics.

Some more initial impressions

There is much to like about this book. Perhaps what strikes me the most is the highly consistent and coherent approach it takes to the topics it covers. Notation is carefully chosen and unified throughout (I give some examples below), and there are surprisingly few errors, given the length of the book. Supporting lemmas are precisely stated in enough generality to be useful in several places. The approach is more bottom-up than top-down, carefully building up the foundations that can be applied generally before getting to the "end" results. Some patience is required of the reader, therefore, but that patience pays off richly later on.

As you go through this book, you will find yourself flipping back to previous sections quite often. Perhaps anticipating this, the author does two things that make this task easy: there are copious citations to

[6] ©2021, Stephen A. Fenner

previous definitions, lemmas, theorems, etc. in the proofs of later results; each definition, lemma, theorem, etc. is stated in a completely self-contained manner, with all ingredients fully quantified over in the statement itself. The latter relieves the reader of the aggravation (common with other texts) of having to scour the surrounding prose for the context necessary to understand the result she is looking up.

Some things are done differently.

The book does some things in a nonstandard way, especially with choice of notation. Most obviously, Watrous avoids Dirac notation completely, opting for more standard mathematical notation. You will not find a single bra or ket anywhere in the text—unusual (to say the least) for a book with "Quantum" in the title. This choice put me off at first, but I gradually came to appreciate its wisdom; in so many places, with vectors and operators of different types acting on different spaces, Dirac notation would just get in the way. In expressions involving the norm of a vector or operator, for example, using $\|\cdot\|$ with a ket would lead to vertical bar fatigue before long.

Another notational difference is writing A^* instead of $A^\dagger$ for the adjoint of an operator A. Without Dirac notation, this leads to some notational unification. The adjoint (or dual) of a vector u is written u^*, same as with an operator, whereas with Dirac notation, the adjoint of a vector $|\varphi\rangle$ is written $\langle\varphi|$, unlike with operators. There are other, more subtle advantages to avoiding notation common in physics, including unambiguity of expressions. In a physics paper, one may encounter an expression like $\langle\alpha|\langle\beta||\gamma\rangle|\delta\rangle$. To parse this correctly, you need to know which vectors combine with which and which juxtapositions represent the inner product and which represent the tensor product. Letting $u = |\alpha\rangle$, $v = |\beta\rangle$, $w = |\gamma\rangle$, and $x = |\delta\rangle$, Watrous would write this as $(u^* \otimes v^*)(w \otimes x)$, which equals the scalar $\langle u, w\rangle\langle v, x\rangle$, rather than, say, $u^* \otimes v^* \otimes w \otimes x$ or $\langle v, w\rangle\langle u, x\rangle$. The book consistently uses $\otimes$ for tensor product, and it maintains the order of the factors when multiplying two tensor products together: $(A \otimes B)(C \otimes D) = AC \otimes BD$ for any conformant objects A, B, C, D.

Watrous defines what others may call a finite-dimensional Hilbert space somewhat differently. He defines a *complex Euclidean space* as the vector space $\mathbb{C}^\Sigma$ for Σ being any nonempty finite set (an *alphabet*). A basis $\{e_a \mid a \in \Sigma\}$ for $\mathbb{C}^\Sigma$ is cooked right into the definition itself, where $e_a(a) = 1$ and $e_a(b) = 0$ for all distinct $a, b \in \Sigma$. Taking $\{e_a\}_{a\in\Sigma}$ to be orthonormal determines the Hermitian inner product $\langle\cdot,\cdot\rangle$ on $\mathbb{C}^\Sigma$. This approach differs in two ways from what is normally done. First, a Hilbert space would normally be defined in an axiomatic, basis-independent manner, which would not be as convenient for many constructions. Second, using an arbitrary alphabet Σ instead of, say, $\{1, \dots, n\}$ as an index set allows for much more flexibility and makes (for example) defining the tensor product of two spaces particularly easy:

$$\mathbb{C}^\Sigma \otimes \mathbb{C}^\Gamma = \mathbb{C}^{\Sigma\times\Gamma}$$

for any alphabets Σ and Γ, with $e_a \otimes e_b = e_{(a,b)}$ for all $a \in \Sigma$ and $b \in \Gamma$.

There are other little consistencies which make for a smooth read. Letters and fonts are sensibly chosen. For example, two quantum registers X and Y have associated complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, respectively, and x may be chosen to represent an arbitrary vector in $\mathcal{X}$ and y chosen to represent a vector in $\mathcal{Y}$. Expressions and statements are "strongly typed" in the sense that types of objects are explicit, either in the expression itself or in the quantifiers; for example, the identity operator acting on a space $\mathcal{X}$ is almost always written $\mathbb{1}_\mathcal{X}$ instead of just the more commonly-used $\mathbb{1}$ or I, the latter being briefer but less readable. I came to really appreciate this strong typing; it took away a lot of the confusion about what was acting on what and in what way. Other literature is not nearly as careful about this, particularly in physics, in my experience.

Topics by chapter

There are eight chapters. Starting with Chapter 2, each chapter ends with a collection of exercises of varying degrees of difficulty, as well as detailed bibliographic notes, which give some historic background as well as citing sources. Starting with Chapter 3, each chapter treats a different core topic of research in quantum information.

Chapter 1 contains the background information and notational conventions needed for the rest of the book. In this chapter only, results are given without proof. Chapter 1 covers a lot of ground and provides a broad survey of the various background concepts and results used repeatedly throughout the book, including lots of linear algebra (of course), as well as the various norms of operators (including the trace norm, Euclidean norm, and the operator norm), some differential calculus, probability, measure and integration, convexity, and positive semidefinite programming. If you are comfortable with most of the topics given here (even if you don't know how the results are proved), you have enough to make it through the rest of the book (you won't need measure and integration until Chapter 7, though). I referred back to this chapter frequently.

Chapter 2 discusses quantum states, purifications thereof, and quantum registers, as well as quantum channels, which are *the* central concept in quantum information theory. (Quantum channels are also sometimes called *quantum operations* or, somewhat mistakenly, *superoperators*, but these latter terms are somewhat ambiguous and the book does not use them.) This is a big, dense chapter; if you read it and nothing else, you will already have gotten a lot out of the book. For complex Euclidean spaces $\mathcal{X}$ and $\mathcal{Y}$, $\mathrm{L}(\mathcal{X},\mathcal{Y})$ denotes the space of linear maps from $\mathcal{X}$ into $\mathcal{Y}$. $\mathrm{L}(\mathcal{X})$ is shorthand for $\mathrm{L}(\mathcal{X},\mathcal{X})$, and $\mathrm{T}(\mathcal{X},\mathcal{Y})$ is shorthand for $\mathrm{L}(\mathrm{L}(\mathcal{X}),\mathrm{L}(Y))$, the space that includes quantum channels from a quantum register X to a quantum register Y. A *quantum state* (or *density operator*) in a space $\mathcal{X}$ is defined to be a positive (semidefinite) operator[7] $\rho \in \mathrm{L}(\mathcal{X})$ with unit trace. A *channel* from a quantum register with space $\mathcal{X}$ to one with space $\mathcal{Y}$ is defined as a map $\Phi \in \mathrm{T}(\mathcal{X},\mathcal{Y})$ that is trace-preserving and completely positive:

- $\mathrm{Tr}\,(\Phi(X)) = \mathrm{Tr}\,(X)$ for all $X \in \mathrm{L}(\mathcal{X})$, and
- $(\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Z})})(Z)$ is a positive semidefinite operator (in $\mathrm{L}(\mathcal{Y}\otimes\mathcal{Z})$), for every complex Euclidean space $\mathcal{Z}$ and positive semidefinite $Z \in \mathcal{X} \otimes \mathcal{Z}$.

Thus channels are linear maps that map states to states, even when combined with an extra, non-acting system.

The chapter builds the discussion of channels from the ground up, first defining four standard representations of an arbitrary $\Phi \in \mathrm{T}(\mathcal{X},\mathcal{Y})$—the natural representation, Choi representation, (Kraus) operator sum representation, and Stinespring representation—and shows how they are interrelated. It then gives several equivalent conditions for both complete positivity and for trace preservation in terms of the four representations above. Several examples of quantum channels are given, including unitary ($X \mapsto UXU^*$ for some unitary U), replacement, completely depolarizing, completely dephasing, and extremal channels.

Much discussion in this chapter is devoted to measurements. There are different ways to approach the subject: Do you only want the classical information and don't care about the post-measurement state? Do you only care about the post-measurement state and ignore the classical information? Do you want both? The book takes the first approach for the "official" definition of a measurement on a register with space $\mathcal{X}$—a map $\mu : \Sigma \to \mathrm{L}(\mathcal{X})$ for some alphabet Σ (the set of possible classical outcomes) such that $\mu(a)$ is positive

[7] An operator $A \in \mathrm{L}(\mathcal{X})$ is *positive*, or *positive semidefinite* iff $A = B^*B$ for some conformant operator B.

for all $a \in \Sigma$ and $\sum_{a \in \Sigma} \mu(a) = \mathbb{1}_{\mathcal{X}}$. This is also called a positive operator-valued measure (or POVM) in the literature. If a state $\rho \in \mathrm{L}(\mathcal{X})$ is measured with μ, then the probability of obtaining an outcome $a \in \Sigma$ is $p(a) = \langle \mu(a), \rho \rangle$, and the register holding ρ is destroyed. After characterizing measurements as quantum-to-classical channels, different types of measurements are discussed, including information-complete, projective, and product measurements. After proving Naimark's theorem and its corollary, which shows how an arbitrary measurement can be realized by a projective measurement on a bigger space, the chapter ends with a discussion of nondestructive measurements (where the post-measurement state is retained), extremal measurements, and ensembles of states.

Chapter 3 addresses state and channel discrimination, which are fundamental tasks in quantum information. Here is a typical scenario: Suppose you are given a quantum register in one of two states ρ or σ, chosen uniformly at random by someone else. You know what ρ and σ are, but you don't know which one you were given. How can you maximize your probability of making a correct guess by measuring the register? Chapter 3 proves the Holevo-Helstrom theorem, which shows that the maximum probability of a correct guess is

$$\frac{1}{2} + \frac{1}{4} \| \rho - \sigma \|_1$$

(where $\| \cdot \|_1$ denotes the trace norm), and describes a measurement that achieves this maximum. The chapter goes into much more generality, including nonuniform probability distributions and ensembles of more than two states, discussing the so-called *pretty good measurement* and proving a related theorem of Barnum & Knill.

The trace distance is commonly used to gauge how two states are distinct (as in the scenario above). A somewhat inverse measure is the fidelity function

$$F(P, Q) = \left\| \sqrt{P} \sqrt{Q} \right\|_1 ,$$

defined for all positive operators P and Q over the same space. Applied to states, F takes values in $[0, 1]$, and $F(\rho, \sigma) = 1$ if and only if $\rho = \sigma$. Thus F measures how similar two states are to each other. The chapter goes into depth with the fidelity function, giving different characterizations of it and results regarding it, including joint concavity and the fact that fidelity cannot decrease by the action of any quantum channel. Also included here are the Fuchs-van de Graaf inequalities, which show how the trace distance and the fidelity are inversely related.

The chapter next turns to channel discrimination, defining the completely bounded trace norm of two maps $\Phi, \Psi \in \mathrm{T}(\mathcal{X}, \mathcal{Y})$ (this is also known as the $\diamond$-norm), and uses it to prove a channel analogue of the Holevo-Helstrom theorem. The chapter ends with a number of topics about channel discrimination, including a semidefinite program for maximum output fidelity.

Chapter 4 covers unital channels and the majorization relation. A quantum channel $\Phi \in \mathrm{T}(\mathcal{X}, \mathcal{X})$ is *unital* if it fixes the identity, i.e., if $\Phi(\mathbb{1}_{\mathcal{X}}) = \mathbb{1}_{\mathcal{X}}$. I think of a unital channel intuitively as one that arises through a natural physical process—a process that does not involve any artificial manipulation or preparation of the output. Unital channels are common and have some interesting properties given in the chapter. The chapter also covers different types of unital channels, including mixed unitary channels (i.e., convex combinations of unitary channels), which are exactly those channels amenable to environment-assisted channel correction. Also covered are Weyl-covariant channels and Schur channels.

Chapter 4 also discusses majorization, both for vectors (really as a warm-up) and Hermitian operators. There are a few equivalent ways to define majorization of vectors. The book uses the following: Given two

n-dimensional real vectors x and y, one says that x *majorizes* y (written $y \prec x$) if there exists a doubly stochastic[8] $n \times n$ matrix A such that $y = Ax$. Intuitively, y results from x by some kind of random mixing process. Analogously to vectors, a Hermitian operator X *majorizes* a Hermitian operator Y over the same space (written $Y \prec X$) if there exists a mixed unitary channel that maps X to Y. The book proves a theorem of Uhlmann saying, among other things, that X majorizes Y if and only if the spectrum of X majorizes the spectrum of Y (as vectors of real values). The chapter closes with two applications of majorization, including my personal favorite, the Schur-Horn theorem, which says that for n-dimensional real vectors x and y, $y \prec x$ if and only if there exists an $n \times n$ Hermitian matrix with diagonal entries y and eigenvalues x. This theorem is used, among other places, to prove Nielsen's theorem in Chapter 6.

Chapter 5 is all about quantum entropy and its uses in characterizing the limits of quantum source coding. This topic is perhaps closest to its analogue in the classical information theory of Shannon, and the chapter spends a fair amount of time going over the most relevant aspects of the classical theory: Shannon entropy, relative entropy, joint and conditional entropies, mutual information, and their properties. The analogue of classical entropy for a quantum state ρ is its *von Neumann entropy* $\mathrm{H}(\rho)$, which is defined as just the classical entropy of ρ's spectrum. It is one way of measuring how "mixed" ρ is, or put another way, how incompletely ρ describes a pure state (pure states have von Neumann entropy 0). Several things are proven about the quantum entropic quantities, e.g., concavity and subadditivity of the von Neumann entropy. Interestingly, not all results in the classical case go through in the quantum case. Classically, $\mathrm{H}(\mathsf{X}) \le \mathrm{H}(\mathsf{X},\mathsf{Y})$ for any joint distribution on classical sources X and Y. This is not necessarily true in the quantum case, although it is shown that $\mathrm{H}(\mathsf{X}) \le \mathrm{H}(\mathsf{Y}) + \mathrm{H}(\mathsf{X},\mathsf{Y})$ for any state of the joint register (X,Y).

A whole subsection is devoted to proving the joint convexity of the quantum relative entropy. This section is rather technical, but this property of the quantum relative entropy finds many uses later on. For example, the relative entropy between two states cannot increase when the states are sent through a mixed unitary channel. For another example, von Neumann entropy is strongly subadditive, i.e.,

$$\mathrm{H}(\mathsf{X},\mathsf{Y},\mathsf{Z}) + \mathrm{H}(\mathsf{Z}) \le \mathrm{H}(\mathsf{X},\mathsf{Z}) + \mathrm{H}(\mathsf{Y},\mathsf{Z})$$

for all states of the joint quantum register $(\mathsf{X},\mathsf{Y},\mathsf{Z})$. Ignoring the register Z yields the usual subadditivity of H as a special case.

Finally, the chapter takes up quantum source coding in analogy with classical source coding. Here, the fundamental question is: How much information can you send through a quantum channel (with no prior entanglement shared between source and receiver)? The classical version is Shannon's (noiseless) source coding theorem, which is proved for fixed-length codes with bounded error, after which quantum source coding is introduced. A theorem of Schumacher gives the quantum version of Shannon's theorem for quantum states sent encoded through a quantum channel, where the fidelity function is used to describe the error bounds.

The chapter then turns to how much *classical* information can be sent through a quantum channel. Here, classical information is encoded into a quantum state (by Alice, say), which is then decoded by means of a measurement (by Bob). We imagine Alice wanting to send a sequence of independently random, identically distributed letters from some source alphabet Σ to Bob by encoding each $a \in \Sigma$ into a quantum state ρ_a, which she sends to Bob. Bob then measures each ρ_a he receives, with the outcome (hopefully) being a. If $a \in \Sigma$ is sent with probability $p(a)$, then prior to his measurement, all Bob knows about the quantum state he receives is the mixed state $\rho = \sum_{a\in\Sigma} \rho_a$. To help with the analysis, the concepts of *accessible information*

[8] A matrix is *doubly stochastic* if all its entries are nonnegative and each row and column sums to 1.

and *Holevo information* are introduced. Given ρ_a and $p(a)$ for all $a \in \Sigma$, the accessible information is just the mutual information (as defined classically) between Alice and Bob, maximized over Bob's choices of measurement of ρ (and it is shown that the maximum is achievable by a particular measurement). The Holevo information is the mutual information between Alice and the quantum state ρ. Equivalently it is the average amount of information gained (beyond the information already in ρ) when the source symbol a is revealed:

$$\chi(\eta) = \mathrm{H}(\rho) - \sum_{a\in\Sigma} p(a)\mathrm{H}(\rho_a) \ ,$$

where η represents the ensemble of states ρ_a, each weighted by $p(a)$. Then Holevo's theorem is proved, which says that the accessible information is always upper bounded by the Holevo information. Since the latter is bounded by $\log_2 n$, where n is the dimension of the space that ρ resides in, Holevo's theorem implies that a register with an n-dimensional space used to transfer a quantum state from Alice to Bob can carry at most $\log_2 n$ many classical bits. The chapter ends with a discussion of quantum random access codes and a proof of Nayak's theorem.

Holevo's bound, above, assumes no shared prior resources between Alice and Bob, in particular, they share no entangled state. Speaking of which, ...

Chapter 6 is all about bipartite entanglement—how to measure it and what one can do with it. As the author admits, this chapter can only scratch the surface of a large body of research on the subject. It does, however, cover those topics which probably are the most widely viewed as important. The chapter is divided into three major sections, covering separability, entanglement manipulation, and phenomena associated entanglement, respectively. The topics in the first section include: the Horodecki criterion for separability; entanglement rank; separable and LOCC channels and measurements (LOCC stands for Local Operations, Classical Communication); state discrimination via separable and LOCC measurements. I already knew much of this material (in less generality), but I did learn some interesting and counterintuitive things; for example, for any two spaces $\mathcal{X}$ and $\mathcal{Y}$, there is a neighborhood of $\mathbb{1}_{\mathcal{X}} \otimes \mathbb{1}_{\mathcal{Y}}$ in which every positive operator in $\mathrm{L}(\mathcal{X} \otimes \mathcal{Y})$ is separable with respect to $\mathcal{X}$ and $\mathcal{Y}$. This implies that any state of $\mathcal{X} \otimes \mathcal{Y}$ sufficiently close to the completely mixed state is separable.

The second section goes into ways of converting entanglement from one form to another. It first proves Nielsen's theorem, which gives a close connection between majorization and the action of separable channels on pure states, and which implies an equivalence between LOCC channels, one-way LOCC channels, and separable channels when acting on pure states. The section next covers two principal measures of bipartite entanglement: distillable entanglement and entanglement cost (the latter once being called entanglement of formation). The distillable entanglement of a bipartite state $\rho \in \mathrm{L}(\mathcal{X} \otimes \mathcal{Y})$ describes the number of copies of the maximally entangled state τ you can produce from n copies of ρ, asymptotically as $n \to \infty$, with arbitrarily good fidelity via LOCC channels. The flip side of this is the entanglement cost, which describes how many copies of τ you need to produce n copies of ρ, asymptotically as $n \to \infty$, with arbitrarily good fidelity via LOCC channels. For any ρ, the entanglement cost is an upper bound on the distillable entanglement. If ρ is a pure state, then these two values are equal to the reduced-state von Neumann entropy $\mathrm{H}(\mathrm{Tr}_{\mathcal{X}}(\rho))$, which is the same as $\mathrm{H}(\mathrm{Tr}_{\mathcal{Y}}(\rho))$. This last fact has intuitive appeal: if a bipartite system is in a pure state ρ (i.e., $\mathrm{H}(\rho) = 0$), then the amount of entanglement of ρ equals the amount of uncertainty about the state obtained by ignoring (i.e., tracing out) one or the other of the component spaces. The section ends with discussing the curious case of *bound entanglement*—states that are entangled but have zero distillable entanglement. The main result here is that all states with positive partial transpose have no distillable entanglement.

The third and final section covers three phenomena associated with entanglement: teleportation, dense coding, and nonclassical correlations (a.k.a. Bell inequality violations). As teleportation is commonly taught, if Alice and Bob share a maximally entangled 2-qubit state (called an EPR pair, or e-bit) then Alice can transfer (teleport) an arbitrary 1-qubit quantum state to Bob with only two bits of classical communication. Inversely, Alice can transfer two classical bits, (densely) coded into a single qubit that she sends to Bob (and this violates the Holevo bound given in the previous chapter). Both of these operations consume their shared e-bit. This section defines teleportation and dense coding more generally—for any number of dimensions—and proves analogous limits on how much information can be transferred in these ways and what types of measurements for Alice are optimal for teleportation.

The study of nonclassical correlations goes back to the work of John Bell in the 1960s. He found a way to physically test a postulate made earlier by Einstein, Podolsky, and Rosen in the 1930s, which posited that a complete quantum theory must be "local" and "realistic." Bell showed that any local realistic theory must satisfy certain statistical inequalities (now known as Bell inequalities) that are violated according to the standard quantum theory. Deterministic and quantum correlation operators are defined to study these inequalities, and the CHSH inequality—due to Clauser, Horne, Shimony, and Holt—is studied in depth. Finally, the optimal quantum correlation is solved explicitly for a prominent special case using Tsirelson's theorem. This chapter in particular has several good exercises and more extensive bibliographic information.

Chapter 7 deals with symmetric vectors and operators as well as unitarily invariant measures. A vector u in the space $\mathcal{X}^{\otimes n}$ (the n-fold tensor product of an arbitrary space $\mathcal{X}$ with itself) is *symmetric* or *permutation-invariant* if $u = W_\pi u$ for any $\pi \in S_n$, where S_n is the group of permutations on $\{1, \ldots, n\}$ and W_π is the linear operator that permutes the n components of u according to π, i.e.,

$$W_\pi(x_1 \otimes \cdots \otimes x_n) = x_{\pi^{-1}(1)} \otimes \cdots \otimes x_{\pi^{-1}(n)}$$

for all $x_1, \ldots, x_n \in \mathcal{X}$. The set of symmetric vectors in $\mathcal{X}^{\otimes n}$ forms an important subspace of $\mathcal{X}^{\otimes n}$. There is an analogous subspace of operators in $\mathrm{L}(\mathcal{X}^{\otimes n})$ (or $\mathrm{L}(\mathcal{X})^{\otimes n}$—the two spaces are canonically isomorphic). It is shown that this subspace is spanned by $\{Y^{\otimes n} : Y \in \mathrm{L}(\mathcal{X})\}$, a fact that is used quite a bit later on. The first section of the chapter also covers purifications of exchangable (i.e., symmetric) states[9] and, using a theorem of von Neumann, proves a fundamental result that characterizes those operators that commute with all permutation-invariant operators: $X \in \mathrm{L}(\mathcal{X}^{\otimes n})$ commutes with all permutation-invariant operators if and only if X is a linear combination of the operators W_π for $\pi \in S_n$.

The next section describes uniform measures on the unit sphere in a space $\mathcal{X}$ (the set of unit vectors in $\mathcal{X}$), as well as the Haar measure on $\mathrm{U}(\mathcal{X})$. Both measures are distinguished by being invariant under unitary transformations. The section then gives three applications of integrating over these measures—the quantum de Finetti theorem, a theorem of Werner giving quantitative limits on cloning quantum states (generalizing the well-known No Cloning theorem), and a proof that all unital channels sufficiently close to the completely depolarizing channel are mixed unitary.

The final section discusses measure concentration for the two measures discussed previously, as well as its applications. Measure concentration says that any well-behaved (Lipschitz) function defined on a continuous probability space sticks close to its mean (or median) with high probability. Measure concentration is used in proving results in quantum information via the *probabilitic method*, which can prove that an object with a certain property exists—not by constructing such an object explicitly (which may not be possible) but rather showing that a randomly chosen object must satisfy the property with positive probability. After proving Lévy's lemma and Dvoretsky's theorem, this section proves that most bipartite pure states are highly

[9] In physics, these might also be called Bosonic states.

entangled. It then uses the probabilistic method to prove an important theorem of Hastings showing that the minimum output entropy of a channel is not additive in general: there exist channels Φ and Ψ of the same type such that

$$\mathrm{H}_{\min}(\Phi \otimes \Psi) < \mathrm{H}_{\min}(\Phi) + \mathrm{H}_{\min}(\Psi) ,$$

where $\mathrm{H}_{\min}(\Xi)$ for a channel Ξ is defined as the minimum value of $\mathrm{H}(\Xi(\rho))$ over all possible input states ρ. This theorem is a key ingredient in the next chapter regarding the non-additivity of a certain type of channel capacity. And indeed ...

Chapter 8 is all about quantum channel capacities. This topic is naturally motivated by Shannon's Noisy Coding theorems of the last century. It turns out, however, that the quantum case is more complicated; there are several inequivalent but nonetheless interesting types of capacity for a quantum channel. The information carried by the channel may be classical or quantum, and the channel may or may not be assisted by prior entanglement shared between sender and receiver. The first section deals with transmitting classical information, with or without shared prior entanglement. It proves the Holevo-Schumacher-Westmoreland theorem, which characterizes the classical capacity of a channel (without shared entanglement) in terms of another quantity known as its *Holevo capacity*. (Roughly, the Holevo capacity $\chi(\Phi)$ of a channel Φ is the maximum possible Holevo information obtained for any probabilistic ensemble of input states sent through Φ.) Then the entanglement-assisted classical capacity theorem is proved, which characterizes this kind of capacity in terms of the analogous entanglement-assisted Holevo capacity. Besides the two Holevo capacities, another useful ancillary quantity—the maximum coherent information—is also discussed.

The next section talks about quantum information through quantum channels. It defines the (entanglement non-assisted) quantum capacity and the entanglement generation capacity of a channel, and proves that the two quantities are equal. It next defines the entanglement-assisted quantum capacity and proves that it is exactly half of the entanglement-assisted classical capacity. The relationship between the two is reminiscent of the "2 classical bits = one quantum bit" relationship obtained by teleportation and dense coding. Then, after several lemmas, the quantum capacity theorem is proved, which equates the quantum capacity with the regularized maximum coherent information of the channel (here, "regularized" means one takes the limit as $n \to \infty$ of the average of the quantity for n repetitions of the channel).

Finally, the third section address non-additivity and super-activation. It proves a remarkable result of Hastings that refutes the long-standing *additivity conjecture* by showing that the Holevo channel capacity can be super-additive: there exists a channel Φ such that $\chi(\Phi \otimes \Phi) > 2\chi(\Phi)$. Before Hastings's result, the additivity conjecture was extensively studied and shown to have a number of equivalent formulations. It is refuted here based on the minimum output entropy result of the previous chapter. The proof is nonconstructive, and no explicit Φ is currently known. By contrast, one can construct a channel which by itself has zero quantum capacity but its tensor product with itself has positive quantum capacity. Such a phenomenon is called *super-activation*. The chapter (and the book) ends with a proof of super-activation, as well as discussing the need for regularization in defining the various capacities of quantum channels.

Topics not covered in the book

As I said earlier, Watrous's book is not meant to be broad or encyclopedic, and there are some topics that are related to quantum information (or informed by it) that nonetheless are not covered in the book. For example, Watrous does not cover quantum error correction or fault tolerance in any depth. He also does not discuss informational aspects of quantum cryptography, e.g., the security of the quantum key exchange protocol of

Bennett and Brassard (commonly referred to as BB84). Some of these topics are included in the earlier[10] book of Wilde (M. M. Wilde, *Quantum Information Theory*, Cambridge, 2013). Wilde's book, which I am much less familiar with, also contains a lot of high-level motivation and intuition. Both books are thus useful in complementary ways—Wilde's book emphasizing broad-based conceptual intuition; Watrous's book emphasizing precision and nuts-and-bolts mathematical techniques.

Overall opinion

This is a great book that fulfills a vital need: a unified, precise, and complete presentation of the most important topics in quantum information. As I mentioned before, after Chapter 1, all theorem-like statements (theorems, propositions, lemmas, and corollaries), except for the most trivial ones, have complete proofs. I found this to be extremely useful while going through the book. The book can be used for reference, self study, or as the primary text for a graduate-level course in quantum information. Reading the book in detail (verifying the proofs all the while) gave me a solid, confident understanding of the techniques used by those in the field—an understanding that I am eager to apply in the future.

[10] First edition reviewed in this column, SIGACT News **47**(3), 2016, pages 12-14. There is a second edition, copyright 2017.