Quantitative Finance

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rquf20

A First Course in Random Matrix Theory for Physicists, Engineers and Data Scientists

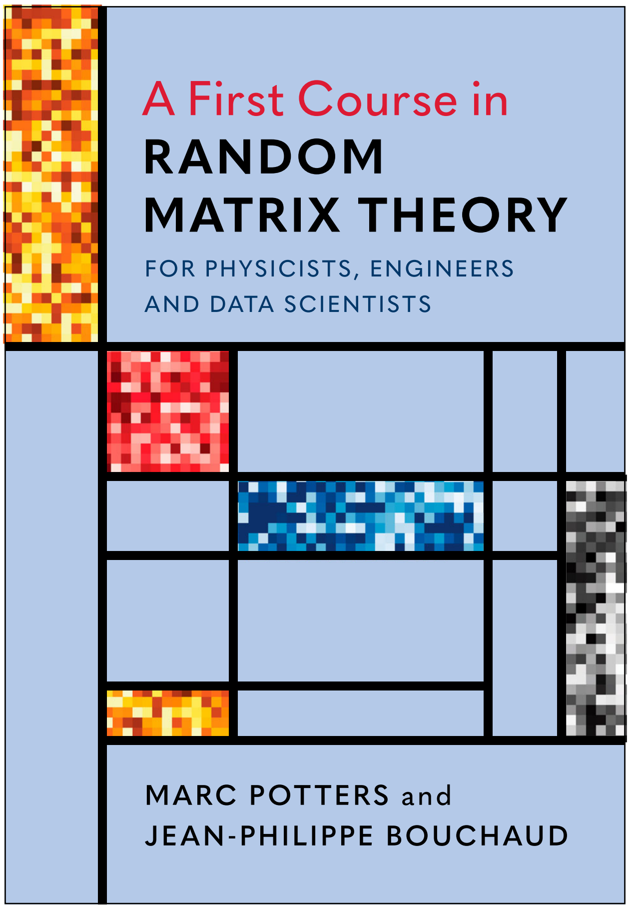
by Marc Potters and Jean-Philippe Bouchaud, Cambridge University Press (2021). Hardback. ISBN 9781108768900.

Louis-Pierre Arguin

Quantitative Finance, 2023
https://doi.org/10.1080/14697688.2023.2166867

Routledge
Taylor & Francis Group

Book review



A First Course in Random Matrix Theory for Physicists, Engineers and Data Scientists, by Marc Potters and Jean-Philippe Bouchaud, Cambridge University Press (2021). Hardback. ISBN 9781108768900.

Random matrix theory (RMT) was first developed in the physics community as an approximation to understand large systems in quantum mechanics. It has now become an integral part of data sciences since, by design, the theory is well equipped to analyze large data sets that include many data points as well as many variables. There is a vast literature on RMT in physics and in mathematics. However, there are few books on the market that introduce the subject to applied mathematicians and data scientists. As such, the book *A First Course in Random Matrix Theory for Physicists, Engineers and Data Scientists* is a welcome addition to the literature.

The book was first developed as a graduate-level course targeting physicists and applied mathematicians. The end product is an exhaustive introduction to the classical theory of random matrices, but also crucially, to recent developments in the field. It will appeal not only to graduate students and instructors, but also to researchers mainly in data sciences, applied mathematics, and physics who would like to deepen their understanding of specific tools and techniques of RMT. The style of the writing is well suited for researchers working on large data sets as the focus of the authors is not to prove theorems with the most general assumptions, but rather to showcase practical techniques and calculations using often well-known examples such as Wigner matrices and Wishart matrices. The material is presented in a very intuitive manner. An undergraduate background in linear algebra and probability is obviously needed. A familiarity with complex analysis is definitely necessary, since tools like the Stieltjes transform are used throughout the book. Knowledge of stochastic calculus is, however, not assumed and the necessary concepts are presented in Chapter 8.

The authors, Bouchaud and Potters, have achieved a real feat here in bringing many concepts of RMT to life and in presenting them in an introductory setting. Both authors are accomplished physicists and mathematicians that have made important contributions to the mathematical theory of random matrices and disordered systems at large. Most importantly, they are also successful financial practitioners. Bouchaud is co-founder and chairman of Capital Fund Management (CFM). He is also adjunct professor at École Normale Supérieure and co-director of the CFM-Imperial Institute of Quantitative Finance at Imperial College London. Potters is currently CIO of CFM and serves on the scientific committee of the CFM-Imperial Institute of Quantitative Finance.

The book contains twenty chapters and is divided into three parts: (I) *Classical Random Matrix Theory*, (II) *Sums and Products of Random Matrices*, and (III) *Applications*. Each chapter contains a fair number of exercises that are designed to develop a familiarity with the concepts introduced. Some of the exercises are numerical, which I found very useful to build the intuition of abstract techniques (such as inverting a Stieltjes transform). There are also detailed and explanatory bibliographical notes at the end of every chapter that guide the reader to recent papers on the material covered.

Part I covers the classical RMT. Chapter 1 reviews the necessary concepts of linear algebra dealing with deterministic matrices. Chapter 2 introduces the Wigner matrices, which is arguably the simplest distribution of random matrices. These are symmetric matrices whose (upper diagonal) entries are independent and identically distributed. For simplicity, the authors assume in most of the book that the entries are Gaussian-distributed. They also introduce the Stieltjes transform as a tool to understand the eigenvalue density of random matrices, and compute it in the case of Wigner matrices to obtain the famous semi-circle law. Another important ensemble (i.e., a distribution of a random matrix) is the Wishart ensemble. This ensemble is a simple model of a *sample covariance matrix*, and hence is extremely relevant for applications to data sciences. The eigenvalue distribution for this ensemble is derived in Chapter 4, this is the so-called *Marchenko-Pastur law*. In Chapter 5, the eigenvalues of the random matrices are seen as charged particles in a potential or *Coulomb gas*. This is a powerful point-of-view, as it generalizes to other ensembles, and gives information on the fluctuations of the eigenvalues around their most likely configuration. Chapter 6 and 7 discuss the fundamental relations between random matrices and orthogonal polynomials of *Hermite*, *Laguerre* and *Jacobi* types.

Part II is without a doubt the most abstract part of the book. The concepts are, however, very important for applications in physics and in data sciences as it tackles the question of the effect of additive and multiplicative noise to a signal. In RMT, this problem translates into the behavior of the spectrum when large random matrices are added or multiplied with each other. Chapter 8 introduces the basic notion of stochastic calculus that is needed to define *Dyson Brownian motion* in Chapter 9. Dyson Brownian motion describes the evolution of eigenvalues of random matrices as the variance increases. It is central in the modern RMT as it is well suited to understand the universality of the eigenvalue distribution as well as the effect of adding two random matrices together. The latter is the problem investigated in Chapter 10, where the notion of *R-transform* is also introduced. This naturally leads to the notion of *freeness* discussed in Chapter 11 and 12. Free probability is a fairly abstract concept, and it is refreshing to see it discussed in an introductory book with an intuitive focus. In a loose sense, freeness generalizes the notion of independence for non-commutative objects, such as random matrices. It turns out that large random matrices are *free*, since the matrix of their eigenvectors tends to be uniformly distributed on the orthogonal group. Their distribution is thus rotationally invariant. The techniques of free probability (such as the R-transform) can then be applied to understand the effect of adding two random matrices. Chapter 13 might be one of my favorites as it shows how to calculate the Stieltjes transform (and thus get information on the spectrum) using the *Replica Method*, that was first developed to compute the free energy of disordered systems. The replica method is a powerful computational tool that is rarely presented in an introductory book. Finally, Chapter 14 discusses the occurrence of *outlier eigenvalue and eigenvector* that may or may not appear when a rank-one matrix is added to a random one.

Part III is really where this book stands out, as it presents applications of RMT to many areas. The authors have included explicit computations and worked-out examples that researchers and students will likely find useful. Chapter 15 goes through the examples of addition and multiplication of Wishart and Wigner matrices. Chapter 16 focuses on the product of many matrices and the relation to Dyson Brownian motion. Sample covariance matrices (for which Wishart matrices is the canonical example) with different types of correlations (spatial and temporal) and their effects on the spectrum are discussed in Chapter 17. The problems of estimating a vector and a sample covariance matrix in the presence of additive or multiplicative noise are tackled in Chapter 18 using a Bayesian approach. In particular, the *Ridge* and the *Lasso* techniques are introduced. In my opinion, one of the highlights of Part III is section 19.5 that shows how to construct the transforms (Stieltjes, R-transform, etc.) using real data with a parametric fit or a non-parametric approximation of the sample eigenvalues. More generally, Chapter 19 introduces the notion of *Rotationally Invariant Estimator* to estimate a signal matrix C knowing a noisy version E. Applications to finance are discussed in the final chapter. The focus is on portfolio theory. In particular, the authors are interested in minimizing the risk of a portfolio of assets under the constraints on the gains. When the covariance matrix between the assets is deterministic, this is the classical Markowitz problem. The authors generalize this setup when the knowledge of C is noisy. One might have to build an estimator for C and the tools of the previous chapters come handy.

I recommend this book for researchers and students in data sciences and applied mathematics that want to get acquainted with the basic RMT. It is exhaustive enough (including recent developments) to appeal to seasoned research scientists that are looking for hands-on RMT techniques and computations to be applied to their own particular data problems. Another introductory book on the market is *Introduction to Random Matrices: Theory and Practice* by Livan, Novaes and Vivo. This is a shorter and more introductory book in its scope.

The present book covers more material and more recent techniques that might be more suited to financial practitioners and researchers.

Louis-Pierre Arguin
Department of Mathematics,
Baruch College & Graduate Center,
City University of New York, New York, NY, USA

https://orcid.org/0000-0003-4704-2622

Louis-Pierre Arguin is an Associate Professor in the Department of Mathematics of Baruch College and of the Graduate Center, City University of New York, since 2014. His research interests are disordered and complex systems, branching processes and number theory. He is the author of 'A First Course in Stochastic Calculus' published by the AMS in 2021.