

Mathematics Monograph Series **2**

**Spectral Analysis of  
Large Dimensional Random  
Matrices**

Zhidong Bai Jack W. Silverstein

(大维随机矩阵的谱分析)



SCIENCE PRESS  
Beijing

## Spectral Analysis of Large Dimensional Random Matrices

**T**he aim of the book is to introduce basic concepts, main results, and widely applied mathematical tools in the spectral analysis of large dimensional random matrices. In it we will introduce many of the fundamental results, such as the semicircular law of Wigner matrices, the Marchenko-Pastur law, the limiting spectral distribution of the multivariate  $F$  matrix, limits of extremal eigenvalues, spectrum separation theorems, convergence rates of empirical spectral distributions, central limit theorems of linear spectral statistics and the partial solution of the famous circular law. While deriving the main results, the book will simultaneously emphasize the ideas and methodologies of the fundamental mathematical tools, among them being: truncation techniques, matrix transformations, moment convergence theorems, and the Stieltjes transform. Thus, its treatment is especially fitting to the needs of mathematics and statistic graduate students, and beginning researchers, who can learn the basic methodologies and ideas to solve problems in this area. It may also serve as a detailed handbook on results of large dimensional random matrices for practical users.

ISBN 7-03-017766-5



9 787030 177667 >

ISBN 7-03-017766-5

定价: 80.00 元

Mathematics  
Monograph Series



# Spectral Analysis of Large Dimensional Random Matrices



in  **EN**  
Ecole Normale Supérieure

Mathematics Monograph Series 2

Zhidong Bai Jack W. Silverstein

# **Spectral Analysis of Large Dimensional Random Matrices**

(大维随机矩阵的谱分析)

*Responsible Editor:* Lü Hong

Copyright © 2006 by Science Press  
Published by Science Press  
<http://www.sciencep.com>  
16 Donghuangchenggen North Street  
Beijing 100717, China

Printed in Beijing, 2006

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior written permission of the copyright owner.

ISBN 7-03-017766-5 (Beijing)

To my great teacher Professor Yongquan Yin  
my wife Xicun Dan  
and my sons Li and Steve Gang

— Zhidong Bai

To my children Hila and Idan

— Jack W. Silverstein



## Preface

This monograph is an introductory book on the Theory of Random Matrices (RMT). The theory dates back to the early development of Quantum Mechanics in the 1940's and 50's. In an attempt to explain the complex organizational structure of heavy nuclei, E. Wigner, Professor of Mathematical Physics at Princeton University, argued that one should not compute energy levels from Schrödinger's equation. Instead, one should imagine the complex nuclei system as a black box described by  $n \times n$  Hamiltonian matrices with elements drawn from a probability distribution with only mild constraints dictated by symmetry considerations. Under these assumptions and a mild condition imposed on the probability measure in the space of matrices, one finds the joint probability density of the  $n$  eigenvalues. Based on this consideration, Wigner established the well-known semi-circular law. Since then, RMT has been developed into a big research area in mathematical physics and probability. Its rapid development can be seen from the following statistics from Mathscinet database under keyword Random Matrix on 10 June 2005 (See Table 0.1.)

1955–1964	1965–1974	1975–1984	1985–1994	1995–2004
23	138	249	635	1205

**Table 0.1.** Publication numbers on RMT in 10 year periods since 1955

Modern developments in computer science and computing facilities motivate ever widening applications of RMT to many areas.

In statistics, classical limit theorems have been found to be seriously inadequate in aiding in the analysis of very high dimensional data.

In the biological sciences, a DNA sequence can be as long as several billions. In finance research, the number of different stocks can be as large as tens of thousands.

In wireless communications, the number of users can be several millions.



All of these areas are challenging classical statistics. Based on these needs, the number of researchers on RMT is gradually increasing. The purpose of this monograph is to introduce the basic results and methodologies developed in RMT. We assume readers of this book are graduate students and beginning researchers who are interested in RMT. Thus, we are trying to provide the most advanced results with proofs using standard methods, as detailed as we can.

With more than a half century's development of RMT, many different methodologies have been developed in the literature. Due to the limitation of our knowledge and length of the book, it is impossible to introduce all the procedures and results. What we shall introduce in this book are those results either obtained under moment restrictions using the moment convergence theorem, or the Stieltjes transform.

In an attempt at complementing the material presented in this book, we have listed some recent publications on RMT which we have not introduced.

The authors would like to express their appreciation to Professors Chen Mufa, Lin Qun, Shi Ningzhong and Ms. Lü Hong for their encouragement and help in the preparation of the manuscript. They would also like to thank Professors Zhang Baoxue, Lee Sungchul, Zheng Shurong, Zhou Wang and Hu Guorong for their valuable comments and suggestions.

Changchun, China  
Cary, North Carolina, USA

Zhidong Bai  
Jack W. Silverstein  
June 2006

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Large Dimensional Data Analysis	1
1.2	Random Matrix Theory	3
1.2.1	Spectral Analysis of Large Dimensional Random Matrices	4
1.2.2	Limits of Extreme Eigenvalues	5
1.2.3	Convergence Rate of ESD	6
1.2.4	Circular Law	6
1.2.5	CLT of Linear Spectral Statisticslinear spectral statistics	7
1.2.6	Limiting Distributions of Extreme Eigenvalues and Spacings	8
1.3	Methodologies	8
1.3.1	Moment Method	8
1.3.2	Stieltjes Transform	9
1.3.3	Orthogonal Polynomial Decomposition	10
<b>2</b>	<b>Wigner Matrices and Semicircular Law</b>	<b>13</b>
2.1	Semicircular Law by the Moment Method	14
2.1.1	Moments of the Semicircular Law	14
2.1.2	Some Lemmas of Combinatorics	14
2.1.3	Semicircular Law for iid Case	17
2.2	Generalizations to the Non-iid Case	23
2.2.1	Proof of Theorem 2.9	23
2.3	Semicircular Law by Stieltjes Transform	28
2.3.1	Stieltjes Transform of Semicircular Law	28
2.3.2	Proof of Theorem 2.9	29
<b>3</b>	<b>Sample Covariance Matrices, Marčenko-Pastur Law</b>	<b>35</b>
3.1	MP Law for iid Case	36
3.1.1	Moments of the MP Law	36

3.1.2	Some Lemmas on Graph Theory and Combinatorics ...	37
3.1.3	MP Law for iid Case .....	42
3.2	Generalization to the Non-iid Case .....	46
3.3	Proof of Theorem 3.9 by Stieltjes Transform .....	47
3.3.1	Stieltjes Transform of MP Law .....	47
3.3.2	Proof of Theorem 3.9 .....	48
4	<b>Product of Two Random Matrices</b> .....	53
4.1	Some Graph Theory and Combinatoric Results .....	54
4.2	Existence of the ESD of $\mathbf{S}_n \mathbf{T}_n$ .....	59
4.2.1	Truncation of the ESD of $\mathbf{T}_n$ .....	61
4.2.2	Truncation, Centralization and Rescaling of the X-variables .....	63
4.2.3	Proof of Theorem 4.3 .....	64
4.3	LSD of $F$ matrix .....	68
4.3.1	General Formula for the Product .....	68
4.3.2	LSD of Multivariate $F$ Matrices .....	69
4.4	Proof of Theorem 4.5 .....	72
4.4.1	Truncation and Centralization .....	72
4.4.2	Proof by Stieltjes Transform .....	75
5	<b>Limits of Extreme Eigenvalues</b> .....	83
5.1	Limit of Extreme Eigenvalues of the Wigner Matrix .....	84
5.1.1	Sufficiency of Conditions of Theorem 5.1 .....	85
5.1.2	Necessity of Conditions of Theorem 5.1 .....	93
5.2	Limits of Extreme Eigenvalues of Sample Covariance Matrix .....	96
5.2.1	Proof of Theorem 5.10 .....	98
5.2.2	The Proof of Theorem 5.11 .....	108
5.2.3	Necessity of the Conditions .....	109
5.3	Miscellanies .....	110
6	<b>Spectrum Separation</b> .....	111
6.1	What is Spectrum Separation .....	111
6.1.1	Mathematical Tools .....	117
6.2	Proof of (1) .....	119
6.2.1	Truncation and Some Simple Facts .....	119
6.2.2	A Preliminary Convergence Rate .....	120
6.2.3	Convergence of $s_n - Es_n$ .....	129
6.2.4	Convergence of Expected Value .....	134
6.2.5	Completing the Proof .....	138
6.3	Proof of (2) .....	139
6.4	Proof of (3) .....	141
6.4.1	Convergence of a Random Quadratic Form .....	141
6.4.2	Spread of Eigenvalues .....	143

6.4.3	Dependence on $y$ .....	146
6.4.4	Completing the Proof of (3) .....	149
<b>7</b>	<b>Semicircle Law for Hadamard Products</b> .....	153
7.1	Renormalized Sample Covariance Matrix .....	153
7.2	Sparse Matrix and Hadamard Product .....	154
7.3	Proof of Theorem 7.4 .....	157
7.3.1	Truncation and Centralization .....	157
7.4	Proof of Theorem 7.4 by Moment Approach .....	161
<b>8</b>	<b>Convergence Rates of ESD</b> .....	169
8.1	Some Lemmas About Integrals of Stieltjes Transforms .....	169
8.2	Convergence Rates of Expected ESD of Wigner Matrices .....	171
8.2.1	Lemmas on Truncation, Centralization and Rescaling .....	172
8.2.2	Proof of Theorem 8.6 .....	174
8.2.3	Some Lemmas of Preliminary Calculation .....	179
8.3	Further Extensions .....	186
8.4	Convergence Rates of Expected ESD of Sample Covariance Matrices .....	186
8.4.1	Assumptions and Results .....	187
8.4.2	Truncation and Centralization .....	188
8.4.3	Proof of Theorem 8.16 .....	189
8.5	Some Elementary Calculus .....	194
8.5.1	Increment of M-P Density .....	194
8.5.2	Integral of Tail Probability .....	197
8.5.3	Bounds of Stieltjes Transforms of M-P Law .....	198
8.5.4	Bounds for $\tilde{b}_n$ .....	199
8.5.5	Integrals of Squared Absolute Values of Stieltjes Transforms .....	202
8.5.6	Higher Central Moments of Stieltjes Transforms .....	203
8.5.7	Integral of $\delta$ .....	207
8.6	Rates of Convergence in Probability and Almost Surely .....	209
<b>9</b>	<b>CLT for Linear Spectral Statistics</b> .....	213
9.1	Motivation and Strategy .....	213
9.2	CLT of LSS for Wigner Matrix .....	217
9.2.1	Strategy of the Proof .....	219
9.2.2	Truncation and Renormalization .....	221
9.2.3	Mean Function of $M_n$ .....	222
9.2.4	Proof of the Nonrandom Part of (9.2.13) for $j = l, r$ ...	228
9.3	Convergence of the Process $M_n - \mathbb{E}M_n$ .....	228
9.3.1	Finite-Dimensional Convergence of $M_n - \mathbb{E}M_n$ .....	228
9.3.2	Limit of $S_1$ .....	231
9.3.3	Completion of Proof of (9.2.13) for $j = l, r$ .....	238

9.3.4	Tightness of the Process $M_n(z) - EM_n(z)$ .....	239
9.4	Computation of the Mean and Covariance Function of $G(f)$ .....	241
9.4.1	Mean Function .....	241
9.4.2	Covariance Function .....	242
9.5	Application to Linear Spectral Statistics and Related Results .....	244
9.5.1	Tchebychev Polynomials .....	244
9.6	Technical Lemmas .....	245
9.7	CLT of LSS for Sample Covariance Matrices .....	246
9.7.1	Truncation .....	248
9.8	Convergence of Stieltjes Transforms .....	250
9.9	Convergence of Finite Dimensional Distributions .....	255
9.10	Tightness of $M_n^1(z)$ .....	266
9.11	Convergence of $M_n^2(z)$ .....	271
9.12	Some Derivations and Calculations .....	277
9.12.1	Verification of (9.8.8) .....	277
9.12.2	Verification of (9.8.9) .....	280
9.12.3	Derivation of Quantities in Example (1.1) .....	281
9.12.4	Verification of Quantities in Jonsson's Results .....	283
9.12.5	Verification of (9.7.8) and (9.7.9) .....	285
10	<b>Circular Law</b> .....	291
10.1	The Problem and Difficulty .....	291
10.1.1	Failure of Techniques Dealing with Hermitian Matrices .....	292
10.1.2	Revisit of Stieltjes Transformation .....	293
10.2	A Theorem Establishing a Partial Answer to the Circular Law .....	295
10.3	Lemmas on Integral Range Reduction .....	296
10.4	Characterization of the Circular Law .....	300
10.5	A Rough Rate on the Convergence of $\nu_n(x, z)$ .....	308
10.5.1	Truncation and Centralization .....	308
10.5.2	A Convergence Rate of the Stieltjes Transform of $\nu_n(\cdot, z)$ .....	309
10.6	Proofs of (10.2.3) and (10.2.4) .....	319
10.7	Proof of Theorem 10.3 .....	322
10.8	Comments and Extensions .....	324
10.8.1	Relaxation of Conditions Assumed in Theorem 10.3 .....	324
10.9	Some Elementary Mathematics .....	327
11	<b>Appendix A. Some Results in Linear Algebra</b> .....	329
11.1	Inverse Matrices and Resolvent .....	329
11.1.1	Inverse Matrix Formula .....	329
11.1.2	Holing a Matrix .....	330

11.1.3 Trace of Inverse Matrix .....	330
11.1.4 Difference of Traces of a Matrix $\mathbf{A}$ and Its Major Submatrices .....	331
11.1.5 Inverse Matrix of Complex Matrices .....	332
11.2 Inequalities Involving Spectral Distributions .....	332
11.2.1 Singular Value Inequalities .....	333
11.3 Hadamard Product and Odot Product .....	338
11.4 Extensions of Singular Value Inequalities .....	342
11.4.1 Definitions and Properties .....	342
11.4.2 Graph-Associated Multiple Matrices .....	344
11.4.3 Fundamental Theorem on Graph-Associated MM.....	347
11.5 Perturbation Inequalities .....	355
11.6 Rank Inequalities .....	361
11.7 A Norm Inequality .....	363
<b>12 Appendix B. Moment Convergence Theorem and Stieltjes Transform .....</b>	<b>365</b>
12.1 Moment Convergence Theorem .....	365
12.2 Stieltjes Transform .....	372
12.2.1 Preliminary Properties .....	372
12.2.2 Inequalities of Distance between Distributions in Terms of Their Stieltjes Transforms .....	375
12.2.3 Lemmas Concerning Levy Distance .....	378
<b>References .....</b>	<b>383</b>
<b>Index .....</b>	<b>391</b>



## Introduction

### 1.1 Large Dimensional Data Analysis

The aim of this book is to investigate the spectral properties of random matrices (RM) when their dimensions tend to infinity. All classical limiting theorems in statistics are under the assumption that the dimension of data is fixed. Then, it is natural to ask why the dimension needs to be considered large and whether there are any differences between the results for fixed dimension and those for large dimension.

In the past three or four decades, a significant and constant advancement in the world has been in the rapid development and wide application of computer science. Computing speed and storage capability have increased a thousand fold. This has enabled one to collect, store and analyze data sets of very high dimension. These computational developments have had strong impact on every branch of science. For example, R. A. Fisher's resampling theory had been silent for more than three decades due to the lack of efficient random number generators, until Efron proposed his renowned bootstrap in the late 1970's; the minimum  $L_1$  norm estimation had been ignored for centuries since it was proposed by Laplace, until Huber revived it and further extended it to robust estimation in the early 1970's. It is difficult to imagine that these advanced areas in statistics would have gotten such deep stages of development if there were no such assistance from the present day computer.

Although modern computer technology helps us in so many aspects, it also brings a new and urgent task to the statisticians, that is, whether the classical limit theorems (i.e., those assuming fixed dimension) are still valid for analyzing high dimensional data and how to remedy them if they are not.

Basically, there are two kinds of limiting results in multivariate analysis: those for fixed dimension (classical limit theorems) and those for large dimension (large dimensional limit theorems). The problem turns out to be which kind of results is closer to reality? As argued in Huber (1973), some statisticians might say that five samples for each parameter in average are enough for using asymptotic results. Now, suppose there are  $p = 20$  parameters and



we have a sample of size  $n = 100$ . We may consider the case as  $p = 20$  being fixed and  $n$  tending to infinity, or  $p = 2\sqrt{n}$ , or  $p = 0.2n$ . So, we have at least three different options to choose for an asymptotic setup. A natural question is then, which setup is the best choice among the three? Huber strongly suggested to study the situation of increasing dimension together with the sample size in linear regression analysis.

This situation occurs in many cases. In parameter estimation for a structured covariance matrix, simulation results show that parameter estimation becomes very poor when the number of parameters is more than 4. Also, it is found that in linear regression analysis, if the covariates are random (or having measurement errors) and the number of covariates is larger than six, the behavior of the estimates departs far away from the theoretic values, unless the sample size is very large. In signal processing, when the number of signals is two or three and the number of sensors is more than 10, the traditional MUSIC (MULTivariate SIGNAL Classification) approach provides very poor estimation of the number of signals, unless the sample size is larger than 1000. Paradoxically, if we use only half of the data set, namely, we use the data set collected by only five sensors, the signal number estimation is almost hundred-percent correct if the sample size is larger than 200. Why this paradox would happen? Now, if the number of sensors (the dimension of data) is  $p$ , then one has to estimate  $p^2$  parameters ( $\frac{1}{2}p(p+1)$  real parts and  $\frac{1}{2}p(p-1)$  imaginary parts of the covariance matrix). Therefore, when  $p$  increases, the number of parameters to be estimated increases proportional to  $p^2$  while the number ( $2np$ ) of observations increases proportional to  $p$ . This is the underlying reason of this paradox. This suggests that one has to revise the traditional MUSIC method if the sensor number is large.  $p \geq 4$

An interesting problem was discussed by Bai and Saranadasa (1996) who theoretically proved that when testing the difference of means of two high dimensional populations, Dempster's (1959) non-exact test is more powerful than Hotelling's  $T^2$  test even when the  $T^2$ -statistic is well defined.

It is well known that statistical efficiency will be significantly reduced when the dimension of data or number of parameters becomes large. Thus, several techniques of dimension reduction were developed in multivariate statistical analysis. As an example, let us consider a problem in principal component analysis. If the data dimension is 10, one may select 3 principal components so that more than 80% of the information is reserved in the principal components. However, if the data dimension is 1000 and 300 principal components are selected, one would still have to face a high dimensional problem. If one only chooses 3 principal components, he would have lost 90% or even more of the information carried in the original data set. Now, let us consider another example.

**Example 1.1.** Let  $X_{ij}$  be iid standard normal variables. Write