COMENIUS UNIVERSITY IN BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS

SPECTRAL DECOMPOSITIONS OF MATRICES AND THEIR RANDOM SUBMATRICES

BACHELOR THESIS

Bratislava 2015 **Šimon Váry**

COMENIUS UNIVERSITY IN BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS
DEPARTMENT OF APPLIED MATHEMATICS AND STATISTICS

SPECTRAL DECOMPOSITIONS OF MATRICES AND THEIR RANDOM SUBMATRICES

BACHELOR THESIS

Study programme: Economic and Financial Mathematics
Branch of study: 9.1.9. Applied mathematics (1114)
Supervisor: Mgr. Martin Niepel, PhD.

Bratislava 2015 **Šimon Váry**

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY
KATEDRA APLIKOVANEJ MATEMATIKY A ŠTATISTIKY

SPEKTRÁLNE ROZKLADY MATÍC A ICH NÁHODNÝCH PODMATÍC

BAKALÁRSKA PRÁCA

Študijný program: Ekonomická a finančná matematika
Študijný odbor: 9.1.9. Aplikovaná matematika (1114)
Vedúci práce: Mgr. Martin Niepel, PhD.

Bratislava 2015 **Šimon Váry**

Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

THESIS ASSIGNMENT

Name and Surname:	Šimon Váry
Study programme:	Economic and Financial Mathematics (Single degree study, bachelor I. deg., full time form)
Field of Study:	9.1.9. Applied Mathematics
Type of Thesis:	Bachelor´s thesis
Language of Thesis:	English
Secondary language:	Slovak

Title: Spectral decompositions of matrices and their random submatrices

Aim: Investigation of the relationship between eigenvalues and eigenvectors of a symmetric matrix and a matrix obtained by a projection on a random subspace. To understand the concept of randomness for symmetric and orthogonal matrices, to become familiar with the classical probability distributions of random matrices.

Keywords: random matrices, spectral decompostition, projection, random subspace

Supervisor:	Mgr. Martin Niepel, PhD.
Department:	FMFI.KAMŠ - Department of Applied Mathematics and Statistics
Head of department:	prof. RNDr. Daniel Ševčovič, CSc.

Assigned: 23.10.2014

Approved: 23.11.2014 doc. RNDr. Margaréta Halická, CSc.
Guarantor of Study Programme

.....
Student

.....
Supervisor

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Šimon Váry
Študijný program: ekonomická a finančná matematika (Jednoodborové štúdium, bakalársky I. st., denná forma)
Študijný odbor: 9.1.9. aplikovaná matematika
Typ záverečnej práce: bakalárska
Jazyk záverečnej práce: slovenský
Sekundárny jazyk: anglický

Názov: Spektrálne rozklady matíc a ich náhodných podmatíc
Spectral decompositions of matrices and their random submatrices

Cieľ: Skúmať vzťahy medzi vlastnými hodnotami a vlastnými vektormi symetrickej matice a matice, ktorá sa z nej získa pomocou projekcie na náhodne vybraný podpriestor. Porozumenie konceptu náhodnosti pre symetrické a ortogonálne matice, oboznámenie sa s klasickými pravdepodobnostnými rozdeleniami náhodných matíc.

Kľúčové slová: nahodné matice, spektrálny rozklad, projekcia, náhodný podpriestor

Vedúci: Mgr. Martin Niepel, PhD.
Katedra: FMFI.KAMŠ - Katedra aplikovanej matematiky a štatistiky
Vedúci katedry: prof. RNDr. Daniel Ševčovič, CSc.

Dátum zadania: 23.10.2014

Dátum schválenia: 23.11.2014 doc. RNDr. Margaréta Halická, CSc.
garant študijného programu

.....
študent

.....
vedúci práce

Acknowledgement

I wish to express my sincere thanks to Mgr. Martin Niepel, PhD. for guidence, valuable advice and friendly approach throughout writing of this thesis. Many thanks also to my family and close friends for their support.

Poďakovanie

Rád by som vyjadril moje úprimné poďakovanie Mgr. Martinovi Niepelovi, PhD. za vedenie, cenné rady a priateľský prístup počas písania tejto práce. Ďakujem tiež mojej rodine a priateľom za ich podporu.

Abstract

Váry, Šimon: Spectral decompositions of matrices and their random submatrices (Bachelor thesis), Comenius University in Bratislava, Faculty of Mathematics, Physics and Informatics, Department of Applied Mathematics and Statistics; Supervisor: Mgr. Martin Niepel, PhD., Bratislava, 2015, 46p.

In thesis, we were concerned with spectral decompositions of random matrices and their random submatrices. We summarized some of the basic concepts of random matrix theory, with emphasis on random symmetric matrices. We outlined moment method of proof for Wigner's Semicircle Law. We examined relations between eigenvectors of random symmetric matrices and their random submatrices. We statistically analysed observed phenomena. In the final part, we provided geometric intuition for the observed phenomena, which allowed us to rephrase this problem and give new directons for further research.

Key words: random matrices, spectral decomposition, projection, Wigner's Semicircular Law, random subspace, random projection

Abstrakt

Váry, Šimon: Spektrálne rozklady matíc a ich náhodných podmatíc (Bakalárska práca), Univerzita Komenského v Bratislave, Fakulta matematiky, fyziky a informatiky, Katedra aplikovanej matematiky a štatistiky; vedúci práce: Mgr. Martin Niepel, PhD., Bratislava, 2015, 46s.

V práci sme sa zaoberali spektrálnymi rozkladmi náhodných matíc a ich náhodných podmatíc. Zhrnuli sme základné poznatky teórie náhodných matíc, s dôrazom na náhodné symetrické matice. Uviedli sme dôkaz Wignerovho polkružnicového zákona cez momenty. Skúmali sme vzťahy vlastných vektorov pre náhodne symetrické matice a ich náhodné podmatice. Pozorované vzťahy sme štatisticky analyzovali. V závere sme uviedli geometrickú interpretáciu pozorovaných javov, pomocou ktorých sme boli schopní refrázovať problém na iný a navrhnuť smer ďalšieho výskumu.

Kľúčové slová: náhodné matice, spektrálny rozklad, projekcia, Wignerov polkružnicový zákon, náhodný podpriestor, náhodná projekcia

Introduction

As sciences entered into an age where an increasingly large volume of complex data is generated through automated measuring devices, much current research in statistics focuses on high-dimensional problems and dimensionality reduction. There might be many variables and/or many observations of them [1]. If we think of each variable as an additional dimension, observation of many variables corresponds to a data point in a high dimensional space. The goal then might be to find a good approximation for the data in a lower dimensional space without loosing much of its explanatory power.

One method for dimensionality reduction is the principal component analysis (PCA) which uses spectral decomposition of a covariance matrix. Imagine a situation in which we have n observations of k random variables. PCA calculates eigenvalues and eigenvectors of the estimated covariance matrix $S \in \mathbb{R}^{k\times k}$ to find lower dimensional set of variables with a high explanatory power. In traditional statistics we would be interested in limiting behaviour as $n \to \infty$ for some fixed k. Thus the estimated covariance matrix would have the fixed size. This was in line with the needs of the times when the number of variables k was small to moderate.

One the other hand, nuclear physicists in 1950s' had to deal with models set for high energy levels, and so the number of variables in the matrix models was large. In quantum mechanics the energy levels of a system, are determined by the eigenvalues of a Hermitian matrix in Schrödinger formula: $H\psi_i = E_i\psi_i$. For the low-lying energy levels this calculation is possible, but at higher energy levels, the analysis becomes too complicated.

Wigner came with a different approach [2] and approximated Hermitian matrix $H \in \mathbb{C}^{n\times n}$ with a matrix whose entries are random variables. He then derived a formula known as Wigner's Semicircle Law which describes asymptotic distribution of spectrum of H as $n \to \infty$. Effectiveness of Wigner's model comes from the fact that instead of calculating concrete particle interactions, he makes an assumption that these are random and observes only overall structure of such system [3].

What if there is a large amount of observations as well as parameters on which we want to perform PCA? One would be then interested in an asymptotic behaviour of eigenvalues and eigenvectors of the estimated covariance matrix $S \in \mathbb{R}^{k \times k}$ as $n, k \to \infty$ while maintaining some fixed ratio: $\frac{n}{k} \in [1, \infty]$. If the null hypothesis assumes that the data come from some distribution, the covariance matrix becomes random with a growing size. In this context there is a need for understanding spectra of random matrices.

In our work we took a close look at relations of eigenvalues and eigenvectors of random matrices and their randomly selected submatrices. This relates to a case, when we leave some number of variables out of many and ask to what extent can we anticipate principal components to change. So in fact, we can ask a question: How robust is PCA in regard to omitting a certain percentage of observed variables? Relationship of eigenvalues can be well described. Cauchy's Interlacing theorem sets bounds for the spectrum of the submatrix. For large random matrices, with known eigenvalue distribution, we can even more accurately approximate the expected spectrum. Particular goal of this thesis was to disclose a similar relationship for the eigenvectors.

In the first two chapters, we introduce a reader to some basic results in random matrix theory. At the center of our attention are random symmetric matrices and their spectral properties.

In the third chapter we describe various observed phenomena between eigenvectors of random matrices and their submatrices, which were not proved yet. For this task, we sampled random symmetric matrices and statistically analysed properties of their eigenvectors.

In the last chapter we provide a reader with geometric intuition behind observed phenomena and show invariant alterations of the experiments.

It turns out that preservation of eigenvectors depends on their corresponding eigenvalues. Eigenvectors corresponding to eigenvalues located on the sparse part of the spectrum are typically more preserved than eigenvectors whose eigenvalues are on the dense part of the spectrum.

1 Preliminaries

This chapter is divided into two sections. In the first we introduce notation used in rest of the thesis, whilst in the second we provide basic theoretical background. This includes definition of *Dirac measure* and *empirical spectral distribution (ESD)*.

1.1 Notation

All *vectors* are assumed to be column vectors. *Matrices* will be represented by capital letters (i.e. $A, B, \dots$). Entries of those matrices will be denoted with corresponding small letters with subindices $\{a_{ij}\}_{i,j=1}^{n} = A$. *Eigenvalues* of a square matrix A will be typically denoted as $\lambda_i(A)$. For symmetric and Hermitian matrices, whose *spectrum* is real; they will be always ordered in descending order: $\lambda_1(A) \geq \lambda_2(A) \geq ... \geq \lambda_n(A)$, and can form a diagonal matrix Λ_A. We say that the spectrum of matrix A is *simple* if all eigenvalues have multiplicity one. *Eigenvectors* will be represented by $v_i(A)$ and may form a matrix:

$$V_A = \begin{pmatrix} | & | & & | \\ v_1(A) & v_2(A) & \dots & v_n(A) \\ | & | & & | \end{pmatrix}$$

For *normal matrices*, V_A will be denoted as U_A to emphasize that by the *spectral theorem* they form unitary basis: $U_A^{\mathsf{H}} U_A = I$. Similarly, for real symmetric matrices we denote V_A as Q_A.

Random variables will be represented by capital letters: $X, Y, Z, ...$, typically we use end of alphabet for random variables. *Probability density function* (PDF) of random variable X is denoted as f_X.

1.2 Definitions

Definition 1.2.1 (Trace)

Trace of a matrix $\mathrm{Tr}(A)$ *is the sum of its diagonal entries :* $\mathrm{Tr}(A) = \sum_i^n a_{ii}$.

It can be proved that trace has following properties:

(i) Cyclic property of trace. Trace is invariant under cyclic permutations:

$$\mathrm{Tr}(ABCD) = \mathrm{Tr}(BCDA) = \mathrm{Tr}(CDAB) = \mathrm{Tr}(DABC).$$

(ii) Trace is similarity-invariant:

$$\mathrm{Tr}(A) = \mathrm{Tr}(S^{-1}BS) = \mathrm{Tr}(B\underbrace{SS^{-1}}_{I}) = \mathrm{Tr}(B).$$

(iii) If A is an $n \times n$ diagonalizable matrix then:

$$\mathrm{Tr}(A^k) = \mathrm{Tr}(Q^{-1}\Lambda^k Q) = \mathrm{Tr}(\Lambda^k) = \sum_{i=1}^{n} \lambda_i^k(A).$$

We will also use following two matrix norms:

Definition 1.2.2 (Frobenius norm)

Frobenius norm is a matrix norm of an $m \times n$ *matrix* A *defined as the square root of the sum of the absolute squares of its entries:*

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}\left|a_{ij}\right|^2} = \sqrt{\mathrm{Tr}(A^T A)}. \tag{1.2.1}$$

Due to the trace form and cyclicity of trace, we see that the Frobenius norm is invariant under orthogonal conjugation:

$$\|Q_1\, A\, Q_2\|_F = \mathrm{Tr}(Q_2^T\, A^T \underbrace{Q_1^T Q_1}_{I} A\, Q_2)^{\frac{1}{2}} = \mathrm{Tr}(A^T A\, \underbrace{Q_2 Q_2^T}_{I})^{\frac{1}{2}} = \|A\|_F.$$

Definition 1.2.3 (Operator norm)

Let M *be a* $n \times n$ *real matrix. Then its operator norm is:*

$$\|M\|_{op} = \sup_{x\in\mathbb{R}^n:\|x\|=1} \|Mx\| = \max_{i=1,..,n} |\lambda_i|. \tag{1.2.2}$$

In this thesis we will be concerned with interrelations between spectral properties of random matrices and their *principal submatrices*. We now define a notion of compression matrix:

Definition 1.2.4 (Compression of matrix)
Matrix $B \in \mathbb{R}^{k\times k}$ $(k < n)$ is said to be a compression of a matrix $A \in \mathbb{R}^{n\times n}$ if:

$$P^T AP = B, \tag{1.2.3}$$

where rectangular matrix $P \in \mathbb{R}^{n\times k}$ is said to be corresponding compression matrix. Additionally, we say that B is orthogonal compression of matrix A, using orthogonal compression matrix P, iff P suffices: $P^T P = I \in \mathbb{R}^{k\times k}$.

Note 1.2.1 (Orthogonal projection $\boldsymbol{PP^T}$)
If P is orthogonal compression matrix, then PP^T is orthogonal projection matrix with $rank(PP^T) = k$ on $\mathbb{R}^n$:

$$\left(PP^T\right)^2 = P\underbrace{P^T P}_{I}P^T = PP^T$$
$$\left(PP^T\right)^T = PP^T.$$

If PP^T is orthogonal projection matrix on first k vectors of standard basis, P has form:

$$P = \left(\begin{array}{c} I_{k\times k} \\ \hline \mathbf{0}_{(n-k)\times k} \end{array}\right), \tag{1.2.4}$$

and the compression B is top left principal submatrix of A.

To be able to observe spectrum of random matrices we use *empirical spectral distribution (ESD)*, which is a sum of Dirac measures on eigenvalues of the matrix.

Definition 1.2.5 (Dirac measure)
Dirac measure is a probability measure δ_x on $\mathbb{R}$, defined for a given $x \in \mathbb{R}$ and any measurable subset $M \subseteq \mathbb{R}$ by:

$$\delta_x(M) = \begin{cases} 1, & x \in M; \\ 0, & x \notin M. \end{cases}$$

Dirac measure is used to find whether a certain entry x is located in some set M. Our main concern is to be able to locate random matrix spectrum. We will use Dirac measures concentrated at eigenvalues of some symmetric matrix A: $\delta_{\lambda_i(A)}$. If we take sum of Dirac measures located in all eigenvalues of a matrix A, we are able to assess how many eigenvalues of A can be found in a certain interval $M \subseteq \mathbb{R}$.

Note 1.2.2 (Dirac measure on complex eigenvalues)
This principle can be used in general for complex eigenvalues/points as well. In this case we would have to define Dirac measure on complex measurable sets $M \subseteq \mathbb{C} = \mathbb{R} \oplus i\mathbb{R}$.

Definition 1.2.6 (Empirical spectral distribution measure)
Empirical spectral distribution (ESD) for a symmetric matrix $A \in \mathbb{R}^{n\times n}$ *is a probability measure* μ_A *on* $\mathbb{R}$*, defined for any measurable subset* $M \subseteq \mathbb{R}$ *by:*

$$\mu_A = \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i(A)}. \tag{1.2.5}$$

We may also omit A and write only μ_n if matrix A can be inferred from context. Notice that ESD is constructed in a way to resemble distribution of eigenvalues for large n.

When A is symmetric random matrix ensemble, then the ESD is random variable taking values in the space of probability measures - it is a random measure [4, p. 160].

To avoid working with a random measure, we can form a deterministic probability measure by taking expectation of the ESD:

$$\overline{\mu_A} = \mathbf{E}\,\mu_A = \mathbf{E}\,\frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i(A)}, \tag{1.2.6}$$

where integration by expectation of measure is defined as:

$$\int_{\mathbb{R}} \phi \, d\,\mathbf{E}\,\mu_A = \mathbf{E}\int_{\mathbb{R}} \phi \, d\mu_A$$

for every continuous compactly supported function $\phi \in C_C(\mathbb{R})$.

This will be used in proving Wigner's semicircle law. It will be easier to show the result for the expected ESD $\mathbf{E}\,\mu_A$. This is in general sufficient, as the difference:

$$\left| \int_{\mathbb{R}} \phi \, d\mu_A - \int_{\mathbb{R}} \phi \, d\,\mathbf{E}\,\mu_A \right|,$$

typically tends to 0 for every fixed $\phi \in C_C(\mathbb{R})$ as $n \to \infty$.

2 Random symmetric matrices

Random matrix is a matrix whose entries are random variables. There are many different classes of random matrices often called ensembles.

Two main subclasses of random matrix ensembles are *invariant ensembles* and *independent entry (iid) ensembles*. Invariant matrix ensembles are classes of random matrices which are invariant under certain group operation (e.g. conjugations by an orthogonal matrix), while iid ensembles have their entries independent [4].

Techniques in random matrix theory mostly rely on one or both of these assumptions [5]. In this chapter we firstly introduce Wigner random matrix (which is a subclass of iid matrix ensembles) and mention some of its properties. We then continue by strengthening conditions for Wigner matrices by asking that their entries possess the Gaussian distribution. It will be shown, that this assumption makes them orthogonally *invariant*.

Lastly we define another class of *invariant* random matrices called Wishart ensembles and mention Marchenko-Pastur Law.

2.1 Wigner matrix

Definition 2.1.1 (real symmetric Wigner matrix)

Let $\{Y_i\}$ and $\{Z_{ij}\}_{i<j}$ be two real-valued families of i.i.d random variables. Additionally we require that all entries have zero mean ($\mathbf{E}\, Y_i = \mathbf{E}\, Z_{ij} = 0$); unitary variance for the off-diagonal entries ($\mathbf{E}\, Z_{ij}^2 = 1$) and some constant variance for the diagonal entries ($\mathbf{E}\, Y_i^2 = \sigma^2$). Furthermore, suppose that their moments of all orders are finite. Then real $n \times n$ real symmetric Wigner matrix M_n is constructed as follows:

$$M = \begin{pmatrix} Y_1 & Z_{12} & \cdots & Z_{1n} \\ Z_{12} & Y_2 & \cdots & Z_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{1n} & Z_{2n} & \cdots & Y_n \end{pmatrix}_{n\times n} \quad ,where \quad \{M_n\}_{i,j} = \begin{cases} Y_i, & if\ i = j; \\ Z_{ij}, & if\ i < j; \\ Z_{ji}, & if\ i > j. \end{cases}$$

$$E(|Z_{ij}|^k) < \infty,$$
$$E(|Y_i|^k) < \infty.$$

2.2 Concentration of the operator norm

Results in random matrix theory hold asymptotically with respect to size of a matrices in the ensemble. Typically, the *operator norm* grows with increasing dimension of a matrix, and it is necessary to normalize our ensembles by dividing by the typical operator norm of $n \times n$ Wigner matrix.

It can be shown [4, p. 126] that typical size of operator norm of Wigner ensemble is $\mathcal{O}(\sqrt{n})$. Additionally, *Tracy-Widom law* [6] tells us not only where operator norm is asymptotically concentrated: $\|M_n\|_{op} \in [2\sqrt{n} - \mathcal{O}(n^{-1/6}), 2\sqrt{n} + \mathcal{O}(n^{-1/6})]$, but also what its distribution in this interval is. Therefore it is useful to normalize Wigner matrices $A_n = \frac{1}{2\sqrt{n}} M_n$ so the operator norm $\|A_n\|_{op}$ will be located around 1. Eigenvalues of A_n are going to be almost surely supported on the interval $[-1, 1]$.

2.3 Gaussian ensembles

If we specify that entries of Wigner matrix come from Gaussian distribution, we get Gaussian ensembles. It turns out that Wigner matrices with Gaussian entries have certain special properties, and it is easier to work with them.

Definition 2.3.1 (Gaussian orthogonal ensemble (GOE))
Let M be a $n \times n$ Wigner matrix. If its diagonal entries are $Y_i \sim \mathbb{N}(0, 2)_{\mathbb{R}}$ and off-diagonal entries are $Z_{ij} \sim \mathbb{N}(0, 1)_{\mathbb{R}}$, then M is Gaussian orthogonal ensemble (GOE).

It is important to note, that the "orthogonal" attribute comes from the fact that these matrices are invariant under conjugation by an orthogonal matrix. This means that if you conjugate matrix from *GOE* distribution by orthogonal matrix we get a new random matrix with same GOE distribution.

Remark 2.3.2 (GOE is orthogonally invariant)
We will show that joint probability density function of GOE ensembles is invariant under multiplication by orthogonal matrix from either side. Let entries of GOE matrix:

Y_i, Z_{ij} are two real-valued families of independent random variables:

$$Y_i \sim \mathbb{N}(0,2)_{\mathbb{R}}$$
$$Z_{ij} \sim \mathbb{N}(0,1)_{\mathbb{R}}$$

Joint density for some symmetric matrix $\{H\}_{i,j} = h_{ij}$ is:

$$P(H) = \prod_{i=1}^{n} f_Y(h_{ii}) \prod_{i=1}^{n-1} \prod_{j=i+1}^{n} f_Z(h_{ij})$$

$$= \left(\frac{1}{\sqrt{4\pi}}\right)^n \left(\frac{1}{\sqrt{2\pi}}\right)^{\frac{n(n-1)}{2}} \exp\left(-\frac{1}{4}\left(\sum_{i=1}^{n} h_{ii}^2 + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} h_{ij}^2\right)\right).$$

Since H is symmetric:

$$= 2^{-\frac{n}{2}} \left(\frac{1}{\sqrt{2\pi}}\right)^{\frac{n(n+1)}{2}} \exp\left(-\frac{1}{4}\sum_{i=1}^{n}\sum_{j=1}^{n} h_{ij}^2\right)$$
$$= 2^{-\frac{n}{2}} \left(\frac{1}{\sqrt{2\pi}}\right)^{\frac{n(n+1)}{2}} \exp\left(-\frac{1}{4}\|H\|_F^2\right).$$

As mentioned in 1.2.2, Frobenius norm is invariant under orthogonal conjugation, therefore GOE matrix ensemble is orthogonally invariant as well.

Main advantage of Gaussian ensembles comes from the fact, that we can express their probability function in terms of their spectra and that they are orthogonally invariant. This will be crucial in the upcoming section to be able to describe distribution of GOE's eigenvectors.

2.4 Spectrum of Wigner matrices

When $A \in \mathbb{R}^{n\times n}$ is a random symmetric matrix its spectrum is a set of n random points on real line; so it becomes a *point process*[1]. Spectrum of random matrices can be observed on two main levels of detail: macroscopic scope, when we are interested in overall behaviour of the whole spectrum; and microscopic scope, when we are interested in behaviour of individual eigenvalues [5]. While the microscopic scope is hard to

[1]Type of random process, for which one realization consists of a set of isolated points.

describe and there have been fewer results in this area, most progress in random matrix theory has been made on the macroscopic scope (including the famous *Wigner's Semicircle Law*). In this section we try to elaborate on two most known concepts in random matrix theory, first being macroscopic behaviour of spectra of Wigner matrices, and the second *eigenvalue repulsion*.

2.4.1 The semi-circular law

As an example, consider a GOE matrix $A \in \mathbb{R}^{3000\times 3000}$. **Figure 1** shows the histogram of all 3000 eigenvalues of one realization of A. Despite random fluctuations, it appears that the histogram follows a semicircular shape. This is in fact, universal property of all Wigner ensembles.

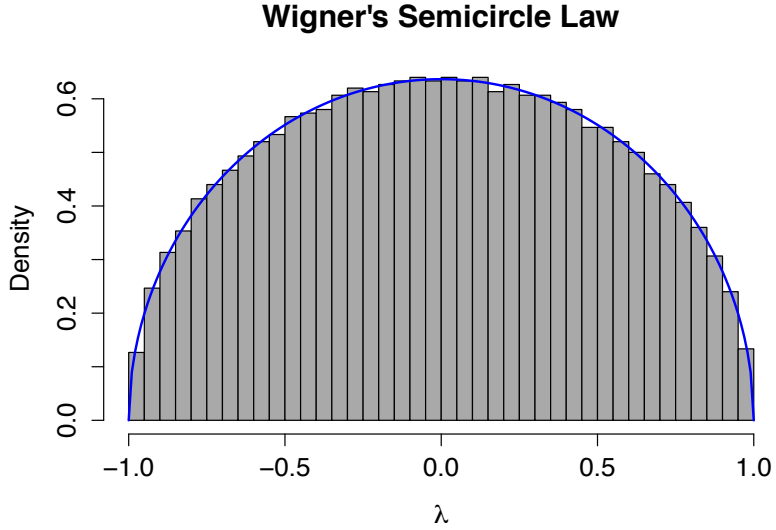


Figure 1: Wigner's Semicircle Law.

Theorem 2.4.1 (Wigner's Semicircle Law)

Let A_n be normalized real symmetric Wigner matrix. Then as n approaches infinity, ESD μ_{A_n} converges in probability to the Wigner semicircular distribution μ_{sc} as $n \to \infty$:

$$\mu_{sc} = \frac{2}{\pi}\sqrt{1-x^2}, \qquad x \in [-1,1]. \tag{2.4.1}$$

Precisely, this means that for any continuous bounded $f \in C(\mathbb{R})$ and every $\epsilon > 0$:

$$\lim_{n\to\infty} P\left(\left|\int f\, d\mu_{A_n} - \int f\, d\,\mu_{sc}\right| > \epsilon\right) = 0. \tag{2.4.2}$$

There are more ways one can prove Wigner's Semicircle Law. In this thesis we will outline shortened version of *the moment method*, which is the one Wigner proposed in his original paper (although he firstly proved this only for Bernoulli ensembles[2] [5]). We prove simplified theorem, in which we assume that f is a polynomial. In this case, it is going to be sufficient to calculate its moments $\int x^k \, d\mu_{sc}$ and $\int x^k \, d\mu_{A_n}$ and show that they converge to the same limit as n tends to infinity. Proof presented here is combined from several sources [7, 8, 9].

Lemma 2.4.2

Let μ_n be a sequence of probability measures with bounded support; furthermore,

$$\forall k \in \mathbb{N} : \qquad \int_{\mathbb{R}} x^k \, d\mu_n \to \int_{\mathbb{R}} x^k \, d\mu_{sc} \quad \text{as } n \to \infty. \tag{2.4.3}$$

Then μ_n converges to μ_{sc}.

Proof. It will be sufficient to show convergence of moments. First we calculate the integral for μ_{sc} (recall formula 2.4.1). Notice that by symmetry, the integral equals to 0 for every odd k.

$$I_{2k} = \int_{\mathbb{R}} x^{2k} \, d\mu_{sc} = \int_{-1}^{1} x^{2k} \left(\frac{2}{\pi} \sqrt{1 - x^2} \right) \, dx = \left| \begin{matrix} x = \sin t \\ dx = dt \, \cos t \end{matrix} \right| =$$

$$= \frac{2}{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^{2k} t \left(1 - \sin^2 t \right) \, dt = \frac{2}{\pi} \left(\underbrace{\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^{2k} t \, dt}_{I_A} - \underbrace{\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^{2k+2} t \, dt}_{I_B} \right)$$

By the $\int \sin^n(x)$ recurrence formula, we have:

$$\int \sin^n x \, dx = \frac{1}{n} \left(- \sin^{n-1} x \cos x + (n-1) \int \sin^{n-2} x \, dx \right), \tag{2.4.4}$$

which yields.

$$I_B = \frac{1}{2k+2} \left(\underbrace{\left[- \sin^{2k+1} t \cdot \cos t \right]_{-\frac{\pi}{2}}^{\frac{\pi}{2}}}_{=0} + (2k+1) \cdot \underbrace{\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^{2k} t \, dt}_{I_A} \right) = \frac{2k+1}{2k+2} \cdot I_A,$$

[2] Wigner matrices, whose entries come from Bernoulli distribution; $M_{ij} = \pm 1$, with equal probability.

$$I_{2k} = \frac{2}{\pi}\left(I_A - I_B\right) = \frac{2}{\pi}\left(I_A - \frac{2k+1}{2k+2}\cdot I_A\right) = \frac{2}{\pi}\left(\frac{1}{2k+2}\cdot I_A\right) =$$

$$= \frac{2}{\pi}\cdot\frac{1}{2k+2}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\sin^{2k}t\,dt = \frac{2}{\pi}\cdot\frac{1}{2k+2}\left(\frac{2k-1}{2k}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\sin^{2k-2}t\,dt\right) =$$

$$= \frac{2k-1}{2k+2}\left(\frac{2}{\pi}\cdot\frac{1}{2k}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\sin^{2k-2}t\,dt\right) = \frac{2k-1}{2k+2}I_{2(k-1)}.$$

Hence we get the recurrent rule for the integral: $I_{2k} = \frac{2k-1}{2k+2}I_{2(k-1)}$. Because μ_{sc} is a probability measure, we already know that the first integral has to be equal to 1: $I_0 = \int_{\mathbb{R}} 1\,d\mu_{sc} = 1$. Now we can derive explicit form of the integral sequence:

$$I_{2k} = \frac{(2k-1)\cdot(2k-3)\ldots3\cdot1}{(2k+2)\cdot(2k)\ldots6\cdot4} = \frac{(2k-1)!!\cdot2}{(2k+2)!!} =$$

using the double factorial identities:

$$= \frac{\frac{(2k)!}{2^k\,k!}\cdot2}{2^{k+1}\,(k+1)!} = \frac{1}{4^k}\frac{1}{k+1}\cdot\binom{2k}{k} = \tag{2.4.5}$$

$$= \left(\frac{1}{4}\right)^k\cdot C_k,$$

where C_k is known as *Catalan number*, and is defined as:

$$C_k = (k+1)^{-1}\cdot\binom{2k}{k}. \tag{2.4.6}$$

To summarize:

$$\int_{\mathbb{R}} x^k\,d\mu_{sc} = \begin{cases}\left(\frac{1}{4}\right)^k\cdot C_k, & \text{if k is odd;}\\ 0, & \text{if k is even.}\end{cases} \tag{2.4.7}$$

We need to show that we get the same result with the limit of the expected ESD of random Wigner matrix A_n: μ_{A_n}. We begin by writing the integral in its trace form:

$$\int_{\mathbb{R}} x^k\,d\,\mathbf{E}\,\mu_n = \mathbf{E}\int_{\mathbb{R}} x^k\,d\mu_n = \mathbf{E}\,\frac{1}{n}\sum_{i=1}^{n}\lambda_i^k(A_n) = \mathbf{E}\,\frac{1}{n}\,\mathrm{Tr}\left(A_n^k\right).$$

Our next goal is to show that:

$$\mathbf{E}\,\mathrm{Tr}\left(A_n^k\right) = \mathbf{E}\,\mathrm{Tr}\left(\left(\frac{1}{2\sqrt{n}}\,M_n\right)^k\right) = \begin{cases} \frac{1}{4^k}\,C_k, & \text{if k is odd;} \\ 0, & \text{if k is even.} \end{cases}$$

For the sake of clarity, in the upcoming section we note $M := M_n$. We begin by expanding M^k using the matrix multiplication formula:

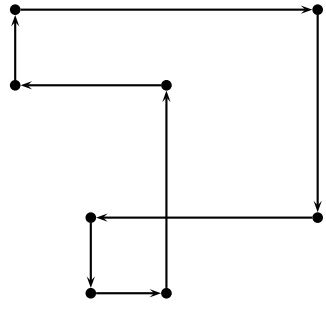
$$\begin{aligned}\mathrm{Tr}\left(M^k\right) &= \sum_{i_1=1}^{n}\left(M^k\right)_{i_1,i_1} = \sum_{i_1=1}^{n}\left(M\cdot M^{k-1}\right)_{i_1,i_1} = \sum_{i_1=1}^{n} M_{i_1*}\cdot\left(M^{k-1}\right)_{*i_1} = \\ &= \sum_{i_1=1}^{n}\sum_{i_2=1}^{n} M_{i_1,i_2}\cdot\left(M^{k-1}\right)_{i_2,i_1} = \sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{i_3=1}^{n} M_{i_1,i_2}\cdot M_{i_2,i_3}\cdot\left(M^{k-2}\right)_{i_3,i_2} = \cdots \\ &= \sum_{i_1,i_2,\ldots,i_k=1}^{n} M_{i_1,i_2}\cdot M_{i_2,i_3}\cdots M_{i_k,i_1}.\end{aligned}$$

Thus:

$$\mathbf{E}\,\mathrm{Tr}\left(M^k\right) = \sum_{i_1,i_2,\ldots,i_k=1}^{n} \mathbf{E}\left(M_{i_1,i_2}\cdot M_{i_2,i_3}\cdots M_{i_k,i_1}\right).$$

Trace of M^k can be understood as a sum of products of its entries in a cycle formed by k nodes as illustrated in 2.4.8 :

$$M = \left(\quad\right)_{n\times n} \tag{2.4.8}$$



Although, there are n^k combinations for the indices $i_1,\ldots,i_k$, contributions of many of these cycles are going to be in expectation equal to zero. This is caused by the zero mean assumption for the entries of Wigner matrices ($\mathbf{E}\,M_{ij} = 0$); therefore all cycles that contain some entries only once contribute to the sum with $0 = \mathbf{E}\left(M_{i_1,i_2}\cdot M_{i_2,i_3}\cdots M_{i_k,i_1}\right) = \underbrace{\mathbf{E}\left(M_{i_1,i_2}\right)}_{0}\mathbf{E}\left(M_{i_2,i_3}\cdots M_{i_k,i_1}\right)$, using also the other assumption that M_{ij} are independent. This means that, for the path to contribute to the sum, it needs to cross every entry at least twice.

Key point here is to divide the sum for paths that go through every entry exactly twice (thus forming $\mathbf{E}\, M_{ij}^2 = 1$ in process), as opposed to the other paths that go through some entries more than twice.

$$\mathbf{E}\,\frac{1}{n}\operatorname{Tr}\left(A_n^k\right) = \left(\frac{1}{4}\right)^{\frac{k}{2}}\left(\frac{1}{n}\right)^{\frac{k}{2}+1}\sum_{\text{double paths}}^{n}\mathbf{E}\left(M_{i_1,i_2}\cdot M_{i_2,i_3}\cdots M_{i_k,i_1}\right) + \left(\frac{1}{4}\right)^{\frac{k}{2}}\left(\frac{1}{n}\right)^{\frac{k}{2}+1}\sum_{\text{other paths}}^{n}\mathbf{E}\left(M_{i_1,i_2}\cdot M_{i_2,i_3}\cdots M_{i_k,i_1}\right).$$

Recall the assumption that moments of all orders are finite: $E(|M_{ij}|^m) < \infty$. This is necessary to show that although "other paths" can contribute to the sum with large (but finite) product of moments, number of paths of this type asymptotically tends to 0 as $n \to \infty$. "Other paths" can meet at most $s = \lfloor \frac{k-1}{2} \rfloor$ entries:

$$\left(\frac{1}{4}\right)^{\frac{k}{2}}\left(\frac{1}{n}\right)^{\frac{k}{2}+1}\sum_{\text{other paths}}^{n}\mathbf{E}\left(M_{i_1,i_2}\cdot M_{i_2,i_3}\cdots M_{i_k,i_1}\right) \le \left(\frac{1}{4}\right)^{\frac{k}{2}}\left(\frac{1}{n}\right)^{\frac{k}{2}+1} C_s\binom{n}{s}.$$

$\binom{n}{s}$ is a number of ways to pick the entries for the "other paths", while C_s is Catalan number that represents number of paths we can select through these entries once we have picked entries[3]. Now expressing the Catalan number as in 2.4.6 and using the upper bound for the binomial coefficient we get:

$$\begin{aligned}\left(\frac{1}{4}\right)^{\frac{k}{2}}\left(\frac{1}{n}\right)^{\frac{k}{2}+1} C_s\binom{n}{s} &= \left(\frac{1}{4}\right)^{\frac{k}{2}}\left(\frac{1}{n}\right)^{\frac{k}{2}+1}\frac{1}{s+1}\binom{2s}{s}\binom{n}{s}\\ &\le \left(\frac{1}{4}\right)^{\frac{k}{2}}\left(\frac{1}{n}\right)^{\frac{k}{2}+1}\frac{1}{s+1}\frac{2s!}{s}\frac{n^s}{s!}\end{aligned}$$

since k is constant and $n \to \infty$:

$$\frac{n^{\lfloor \frac{k-1}{2} \rfloor}}{n^{\frac{k+2}{2}}} \to 0,$$

which yields :

$$\int_{\mathbb{R}} x^k\, d\,\mathbf{E}\,\mu_n = \mathbf{E}\,\frac{1}{n}\operatorname{Tr}\left(A_n^k\right) = \left(\frac{1}{4}\right)^{\frac{k}{2}}\left(\frac{1}{n}\right)^{\frac{k}{2}+1}\sum_{\text{double paths}}^{n}\mathbf{E}\left(M_{i_1,i_2}\cdot M_{i_2,i_3}\cdots M_{i_k,i_1}\right).$$

Using the same combinatorial argument with Catalan numbers we calculate the sum for "double paths". To be able to pick each entry exactly twice we need k to be even. We continue with $2k$ for some $k \in \mathbb{N}$. There are exactly $\binom{n}{2k/2}$ ways to pick double entries and C_k ways to select path between them.

[3]For more detail see [8, pp. 8-12].

$$\begin{aligned}\int_{\mathbb{R}} x^{2k}\, d\,\mathbf{E}\,\mu_n &= \left(\frac{1}{4}\right)^k \left(\frac{1}{n}\right)^{k+1} \sum_{\text{double paths}}^{n} \mathbf{E}\left(M_{i_1,i_2}\cdot M_{i_2,i_3}\cdots M_{i_{2k},i_1}\right)\\ &= \left(\frac{1}{4}\right)^k \left(\frac{1}{n}\right)^{k+1} C_k\, n(n-1)\dots(n-k-1)\\ &\xrightarrow{n\to\infty} \left(\frac{1}{4}\right)^k C_k.\end{aligned}$$

Therefore proving the Lemma 2.4.2:

$$\int_{\mathbb{R}} x^{2k}\, d\,\mathbf{E}\,\mu_n \xrightarrow{n\to\infty} \left(\frac{1}{4}\right)^k C_k \xleftarrow{n\to\infty} \int_{\mathbb{R}} x^{2k}\, d\mu_{sc}. \qquad \square$$

Now, one would continue to prove Wigner's Semicircle Law, by showing that both measures converge in probability. This can be done by using Chebyshev's Inequality by showing that variance of $\int_{\mathbb{R}} x^{2k}\, d\,\mathbf{E}\,\mu_n$ goes to 0. Rest of this proof as well as thorough explanation of combinatorial argument can be found in [8, pp. 5-17].

2.4.2 Repulsion of eigenvalues

One could wrongly assume that eigenvalues of symmetric Wigner matrices come from Wigner's semicircle distribution. This is not true, as individual eigenvalues of random matrices are not independent variables. Notable difference between independent picks from semicircular distribution and distribution of spectrum of Wigner matrix is eigenvalue repulsion, or how much individual points tend to cluster and how big are gaps between them. This behaviour causes that if we sample n independent observations using the semicircular distribution in 2.4.1, these points will typically cluster more, with larger gaps.

This can be explained by the *eigenvalue repulsion*; effect which causes eigenvalues to repel from each other. First argument, why spectra behaves in this way may be seen from codimension[4] of symmetric matrices with repeated eigenvalues relative to all symmetric matrices. Let us say that A is an $n\times n$ symmetric matrix with $dim(A) = \frac{n(n+1)}{2}$. By spectral decomposition $A = Q_A\Lambda_A Q_A^T$, where $dim(\Lambda_A) = n$ and $dim(Q_A) = \frac{n(n-1)}{2}$,

[4]Codimensionality is the dimension difference of variety and its subvariety.

we can say that codimension of symmetric matrices with at least one repeated eigenvalue ($dim(\Lambda_A) = n - 2$) is 2. On the contrary, n-tuples of real numbers with at least one repeating are of codimension 1. Because it is far more rare for eigenvalues to be equal, one would assume that there is "force" that "repels" eigenvalues from getting too close to each other [4, p. 56].

We continue by showing physical argument for eigenvalue repulsion stated on T. Tao's blog[10]. Suppose that A varies smoothly with respect to parameter t. We are then interested in dynamics of eigenvalues $\lambda_i(t) = \lambda_i\big(A(t)\big)$ and eigenvectors $v_i(t) = v_i\big(A(t)\big)$. If we differentiate eigenvalue equation $Ax_i = \lambda_i\, v_i$ with respect to t, using the product rule we have:

$$\begin{aligned} (A\, v_i)' &= (\lambda_i\, v_i)', \\ A'\, v_i + A\, v_i' &= \lambda_i'\, v_i + \lambda_i\, v_i'. \end{aligned} \tag{2.4.9}$$

We then continue by differentiating unitary condition for the eigenvectors $v_i^T v_i = 1$:

$$\begin{aligned} {v_i'}^T v_i + v_i^T v_i' &= 0, \\ {v_i'}^T v_i &= 0, \end{aligned} \tag{2.4.10}$$

thus eigenvector $v_i(t)$ is orthogonal to its derivative $v_i(t)'$. Next we take inner product of $v_i(t)$ with 2.4.9:

$$\begin{aligned} v_i^T A'\, v_i + \underbrace{v_i^T A\, v_i'}_{\lambda_i\, v_i^T v_i'} &= \lambda_i' \underbrace{v_i^T v_i}_{1} + \lambda_i\, v_i^T v_i', \\ \lambda_i' &= v_i^T A'\, v_i, \end{aligned} \tag{2.4.11}$$

where final equation in 2.4.11 is called *Hadamard first variation formula.* Inner product of $v_j(t)$, $j \neq i$ with 2.4.9 yields:

$$\begin{aligned} v_j^T A'\, v_i + \underbrace{v_j^T A\, v_i'}_{\lambda_j\, v_i^T v_i'} &= \lambda_i' \underbrace{v_j^T v_i}_{=0} + \lambda_i\, v_j^T v_i', \\ v_j^T A'\, v_i &= \big(\lambda_i - \lambda_j\big)\; v_j^T v_i', \end{aligned} \tag{2.4.12}$$

now using the fact that v_i forms orthonormal basis in $\mathbb{R}^n$, we can express v_i':

$$v_i' = \sum_{j \neq i} \frac{v_j^T A' v_i}{\lambda_i - \lambda_j} v_j\,. \tag{2.4.13}$$

We continue by taking second derivative of 2.4.9 and 2.4.10:

$$\begin{aligned} A''v_i + 2A'v_i' + Av_i'' &= \lambda_i''v_i + 2\lambda_i'v_i' + \lambda_i v_i'', \\ {v_i''}^T v_i + {v_i'}^T v_i' &= 0, \end{aligned}$$

and by taking inner product with v_i :

$$\begin{aligned} v_i^T A''v_i + 2v_i^T A'v_i' + \underbrace{v_i^T Av_i''}_{\lambda_i v_i^T v_i''} &= \lambda_i'' \underbrace{v_i^T v_i}_{1} + 2\lambda_i' \underbrace{v_i^T v_i'}_{0} + \lambda_i v_i^T v_i'' \\ \lambda_i'' &= v_i^T A''v_i + 2v_i^T A'v_i'. \end{aligned} \tag{2.4.14}$$

Using the 2.4.13, we conclude *Hadamard second variation formula*:

$$\lambda_i'' = v_i^T A''v_i + 2\sum_{j\neq i} \frac{\left(v_j^T A'v_i\right)^2}{\lambda_i - \lambda_j}. \tag{2.4.15}$$

One can view eigenvalues as points on real line influenced by various "forces". In this perspective 2.4.15 serves as an analogy to *Newton's second law of motion* for eigenvalues, where terms on the right hand side are "forces" acting on the eigenvalue λ_i. We see that initial acceleration of A'' provides such a force, while other eigenvalues provide repulsive forces. Each eigenvalue λ_j "repels" the other eigenvalue by force that is inversely proportional to their distance.

This however, should serve only as an intuition and curious property, not rigorous mathematical argumentation.

Tao and Vu proved that real symmetric random matrix A has simple spectrum (i.e. no repeated eigenvalues) with probability tending to 1 as n goes to infinity:

$$P\left(\text{spectrum of } A \text{ is simple}\right) \geq 1 - n^{-C}, \tag{2.4.16}$$

for some $C > 0$ [11].

2.5 Eigenvectors of Gaussian ensembles

Recall the invariant property of Gaussian ensembles. Now, using this fact we can show that eigenvectors of GOE are Haar distributed, i.e. it means they are uniformly distributed in a sphere.

Definition 2.5.1 (Haar distribution on orthogonal group)
Haar measure on the group $O(n)$ (i.e. $n \times n$ orthogonal matrices) is a probability measure P, which is left-translation invariant:

$$A \subset O(n), \quad \forall Q \in O(n): \quad P(A) = P(QA). \tag{2.5.1}$$

Haar measure on orthogonal matrices coincides with *uniform spherical measure*. It generates n random orthogonal vectors uniformly distributed in unit $\mathbb{R}^n$ sphere [12].

Definition 2.5.2 (Spherical measure)
We say that the set of k vectors $v_1, \ldots, v_k \in \mathbb{R}^n$ (forming matrix $V = [\vec{v_1}, \ldots, \vec{v_k}]_{n\times k}$) is orthogonally Haar distributed if :

$$\forall Q \in O(n): \quad P(V) = P(QV). \tag{2.5.2}$$

Remark 2.5.3 (GOE eigenvectors are Haar distributed)
Let H be GOE matrix, with spectral decomposition $H = Q\Lambda Q^T$ and R is some orthogonal matrix. From 2.3.2 we have:

$$P(Q\Lambda Q^T) = P\left((RQ)\Lambda(RQ)^T\right) \implies P(Q) = P(RQ).$$

Therefore, one of the properties of Gaussian ensembles is, that their eigenvectors are uniformly distributed in unit sphere. This however doesn't hold for general Wigner matrix. There are some universal results for eigenvectors of general Wigner matrices that can be found in the work of Tao and Vu [13].

2.6 Wishart matrix

We now briefly introduce other class of random symmetric matrix ensembles that are, orthogonaly *invariant*. Imagine a situation in which we have m observations with n variables. If these observations came from Gaussian distribution, than the sampled covariance of these observations represents random Wishart matrix.

Definition 2.6.1 (Wishart matrix)
Let entries of $n \times m$ matrix $X = \left\{X_{ij}\right\}_{i,j=1}^{n,m}$ be real-valued independent and identically distributed Gaussian random variables, with zero mean $\mathbf{E}\, X_{ij} = 0$ and unitary variance $\mathbf{E}\, X_{ij}^2 = 1$. Then $n \times n$ matrix $M = \frac{1}{m} X X^T$ is called real random Wishart matrix, with m degrees of freedom:

$$M = \frac{1}{m} \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1m} \\ X_{21} & X_{22} & \cdots & X_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{nm} \end{pmatrix} \cdot \begin{pmatrix} X_{11} & X_{21} & \cdots & X_{n1} \\ X_{12} & X_{22} & \cdots & X_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{1m} & X_{2m} & \cdots & X_{mn} \end{pmatrix}.$$

Note 2.6.1 (multivariate Chi-squared distribution)
For $n = 1$, we get Chi-squared distribution, so Wishart distribution is multivariate analogue of Chi-squared distribution.

Remark 2.6.2 (Wishart matrices are orthogonally invariant)
Since multivariate Gaussian distribution is orthogonally invariant, we get:

$$P(QX) = P(X),$$

therefore Wishart matrices are orthogonally invariant.

Theorem 2.6.1 (Marchenko-Pastur Law [14])
Let A_n be normalized real symmetric Wishart matrix, with m degrees of freedom. Then as n approaches infinity and $\frac{n}{m} = q$ stays fixed for some $q \in (0, 1]$; ESD μ_{A_n} converges in probability to the Marchenko-Pastur distribution μ_{mp} as $n \to \infty$:

$$\mu_{sc} = \frac{1}{2\pi x q} \sqrt{(b - x)(x - a)}, \qquad x \in [a, b], \tag{2.6.1}$$

where $a = (1 - \sqrt{q})^2$ *and* $b = (1 + \sqrt{q})^2$. *Precisely, this means that for any continuous bounded* $f \in C(\mathbb{R})$ *and every* $\epsilon > 0$*:*

$$\lim_{n\to\infty} \Pr\left(\left|\int f \, d\mu_{A_n} - \int f \, d\,\mu_{mc}\right| > \epsilon\right) = 0. \tag{2.6.2}$$

Note 2.6.2

Marchenko-Pastur Law holds even for more general case, when entries of matrix X *do not come from Gaussian distribution; it is sufficient that they are independent, zero centered* $\mathbf{E}\, X_{ij} = 0$ *and have unitary variance* $\mathbf{E}\, X_{ij}^2 = 1$.

The moment method of proof for Marchenko-Pastur Law is similar to the moment method for Wigner's Semicircle Law outlined earlier and is based on similar combinatorial arguments [15]. Here we present graphical illustration of the law only, with the histogram of spectrum of one 1000×1000 Wishart matrix realization:

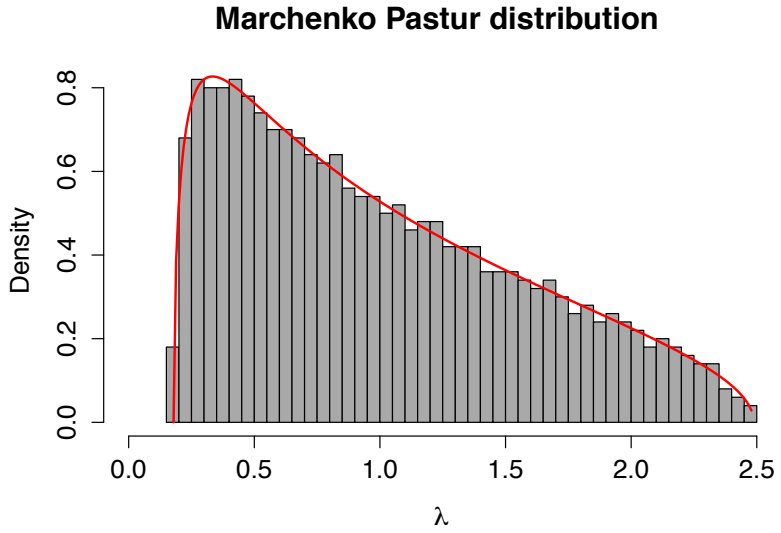


Figure 2: Illustration of Marchenko-Pastur Law for $n = 1000$, $m = 3000$, $q = \frac{1}{3}$.

3 Spectral decompositions of random matrices

In this chapter we introduce the main topic of this thesis. We investigate relations between spectral properties of random Wigner matrices and their randomly picked submatrices.

We consider the *original matrix* $A \in \mathbb{R}^{n\times n}$ and its *principal submatrix* $B \in \mathbb{R}^{k\times k}$ ($k < n$). The submatrix B can be formed by using *orthogonal compression matrix* $P \in \mathbb{R}^{n\times k}$. The orthogonal compression matrix P consists of k basis vectors of $\mathbb{R}^n$ as columns as stated in 1.2.4:

$$B = P^T AP. \tag{3.0.3}$$

In this thesis we consider symmetric real matrices A and B. Due to the spectral theorem, we can write the following chain of equations:

$$P^T AP = (P^T Q_A)\Lambda_A (P^T Q_A)^T = Q_B^T \Lambda_B Q_B = B. \tag{3.0.4}$$

Relations between matrices in 3.0.4 (under various conditions for A and P) are of particular interest in rest of this thesis. We show, that spectrum of B depends on the spectrum of A, and therefore one can anticipate similar connection between their eigenvectors encoded in matrices Q_A and Q_B as well.

3.1 Relation of spectra

Cauchy's Interlacing Theorem, holds for any Hermitian matrix and gives us boundaries for eigenvalues of submatrix in terms of eigenvalues of the original matrix.

Theorem 3.1.1 (Cauchy's Interlacing Theorem)
Let $B \in \mathbb{C}^{n\times n}$ be a Hermitian matrix and $B \in \mathbb{C}^{k\times k}$ be a compression of A. Then the following equalities hold:

$$\forall i \in \{1, \dots, k\} : \qquad \lambda_i(A) \le \lambda_i(B) \le \lambda_{n-k+i}(A) \tag{3.1.1}$$

Proof with geometrical interpretation can be found in [16]. For $k = n - 1$ we get: $\lambda_i(A) \le \lambda_i(B) \le \lambda_{i+1}$, hence the name interlacing theorem. Note that when we pick the orthogonal compression P using first k vectors of standard basis (so that the compression B is principal submatrix), we get information about the spectrum of principal

submatrix.

Note 3.1.1 (Illustration of Cauchy's Interlacing Theorem)

Here we offer a graphical illustration for real 7×7 *matrix with eigenvalues uniformly distributed in* $[-1, 1]$. *Eigenvalues in every row, determine bounds for the eigenvalues beneath them.*

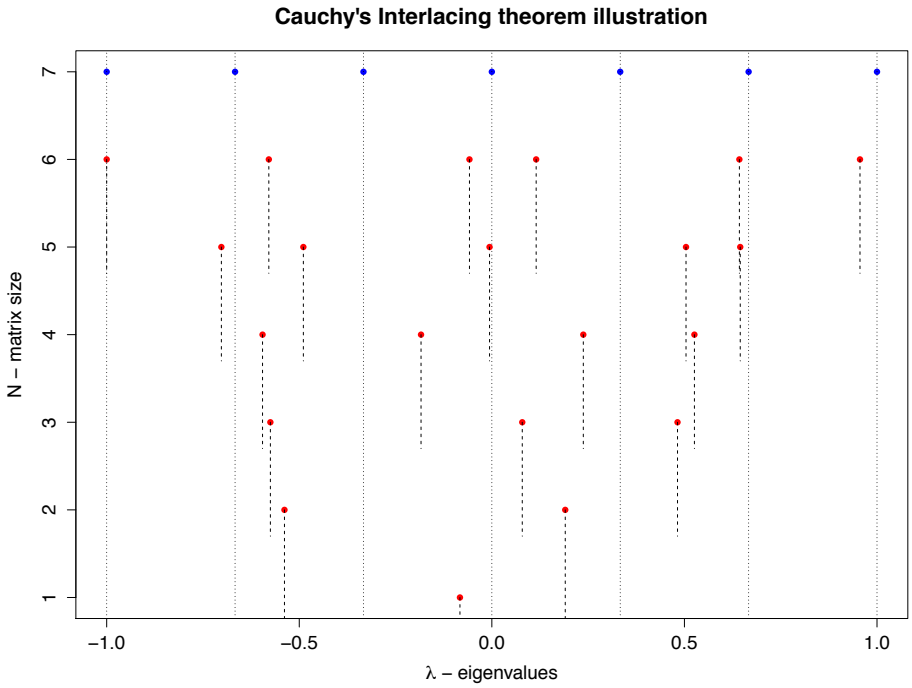


Figure 3: Illustration of Cauchy's Interlacing Theorem: blue marks represent spectrum of the original matrix A, following rows were created by recursive series of orthogonal compressions. Dotted lines demonstrate bounds for spectrum of the following compression matrix.

Therefore we can predict location of the compressed matrix eigenvalues, in terms of the original spectrum. In case of random Wigner matrices, this prediction can be taken even further. Since by taking principal submatrix of A we maintain its symmetricity and distribution of entries, B is random Wigner matrix as well. Only difference is a scale of $\lambda_i(B)$; while the eigenvalues of the former matrix $\lambda_i(A)$ are almost surely located in $[-1, 1]$, eigenvalues $\lambda_i(B)$ are almost surely in $\left[-\sqrt{\frac{k}{n}}, \sqrt{\frac{k}{n}}\right]$. Using the Wigner's Semicircle Law, we can then asymptotically approximate Λ_A as well as Λ_B by semicircle.

However this is not true in case of Wishart matrices when, by compressing the matrix we shift parameter $q : \frac{n}{m} \to \frac{k}{m}$. While in case of Wigner matrices it is sufficient to linearly rescale semicircle distribution; this can not be done for Wishart matrices and Marchenko-Pastur Law, since the distribution changes according to q as illustrated in **Figure 4**.

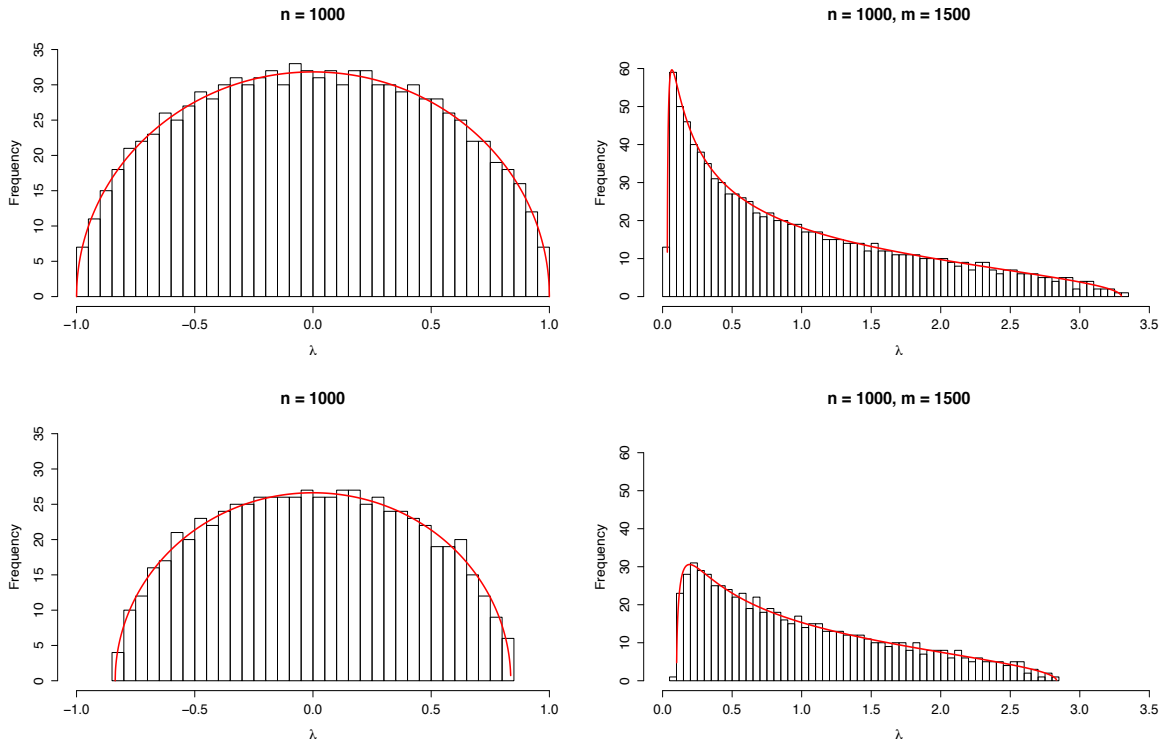


Figure 4: Illustration of the different effect of orthogonal compression on spectra of Wigner and Wishart matrix.

3.2 Relation between eigenvectors

Let assume that we have a random real Wigner matrix A, and we take its random compression B using random compression matrix P as described above. Asymptotically we know Λ_A and Λ_B will behave according to the semicircular distribution. Exact form of P is not important as it will not effect spectrum of B in any direct way. We are then interested in relation between eigenvectors of A and B. Disclosure of such relation would allow us to estimate Q_B in terms of spectral properties of the original matrix A and the distribution of P.

To be able to compare n eigenvectors in $\mathbb{R}^n$ with k eigenvectors in $\mathbb{R}^k$ we need to project the former n-tuple into k dimensional, or insert the k-tuple into n dimensional space. We are therefore interested in examining relations between $P^T v_i(A)$ and $v_j(B)$ using the inner product[5]. By forming the $n \times k$ matrix $V = (P^T Q_A)^T Q_B$ we compute inner products between all eigenvector pairs (i.e. $V_{ij} = \langle P^T v_i(A), v_j(B)\rangle = \langle v_i(A), P v_j(B)\rangle$). Matrix V is of particular interest of this thesis. Since :

$$\frac{a^T b}{\|a\| \cdot \|b\|} = cos\, \phi, \tag{3.2.1}$$

where ϕ is angle between vectors a and b, we can transform matrix V to be a matrix of corresponding angles.

One has to also deal with eigenvalue and eigenvector ambiguity in a sense that we have freedom to rotate each $v_i(A)$ by a unit phase $e^{i\phi}$ in complex Hermitian case or to multiply each $v_i(A)$ by ± 1 in real symmetric case. To remove this ambiguity we will consider absolute values or squares of inner products (or absolute deviation angles). Let $V = \left\{v_{ij}\right\}_{i,j=1}^{n,k}$ then we will use notation $|V| = \left\{\left|v_{ij}\right|\right\}_{i,j=1}^{n,k}$.

To observe asymptotic typical behaviour, we calculate V for N realizations of random matrix A and random orthogonal compression P. For this tasks we wrote collection of scripts in R for generating matrices from various random ensembles and methods for their analysis[6]. We then follow by calculating mean of absolute values of

[5]Note that $\langle P^T v_i(A), v_j(B)\rangle = \langle v_i(A), P v_j(B)\rangle$, so both methods yield same results.

[6]All scripts are available at https://github.com/vary2/rmt-thesis.

matrices $\overline{V} = N^{-1} \sum_1^N |V_i|$. In **Figure 5**, we present perspective plot generated using the *rgl* [17] package in R, of $\overline{V}$ when A is 100×100 GOE matrix and P is an orthogonal compression on 70 random basis vectors.

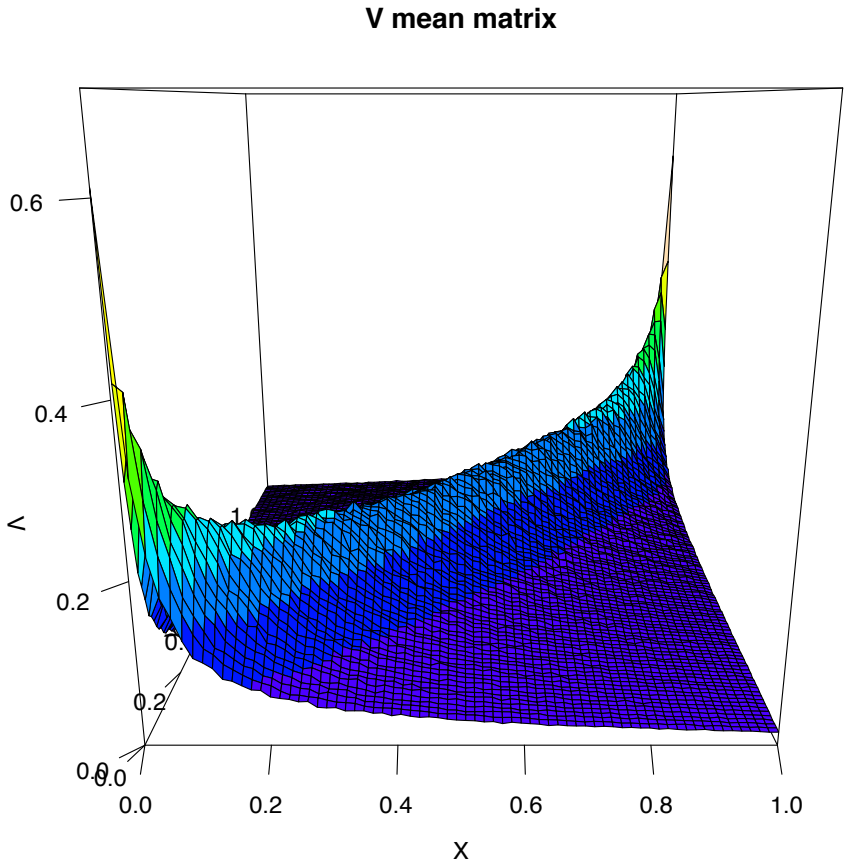


Figure 5: Results $\overline{V}$ for $N = 1000, n = 100, k = 70$. Eigenvalues are ordered in descending order, therefore $(0,0)$ point corresponds to the $\langle P^T v_1(A), v_1(B) \rangle$ and $(1,1)$ to $\langle P^T v_{100}(A), v_{70}(B) \rangle$.

Notice, that $\overline{V}$ has raised entries around the diagonal. Because $\overline{V}$ is a rectangular matrix, we can not look at the exact diagonal, but as $n \to \infty$ and $k \to \infty$, matrix $\overline{V}$ starts to behave continuously. In this manner we define diagonal entries for large rectangular $n \times k$ matrix V as $\left\{ v_{j\frac{n}{k},j} \right\}_{j=1}^{k}$, where value $j\frac{n}{k}$ is rounded to the closest integer. Entries around diagonal represent inner products of eigenvectors whose eigenvalues are close, while the off-diagonal entries represent inner products of eigenvectors whose eigenvalues are distant.

One can examine many different statistics of V. We can study distribution of certain entry of the matrix. For this purpose, we sampled 10 000, $V \in \mathbb{R}^{100\times 70}$ matrices and observed histograms of the individual matrix entries. Entries concentrated around the diagonal tend to have higher variance and mean. As we move farther from the diagonal, entries tend to get more stable and are typically lower. Moreover notice in **Figure 6**,

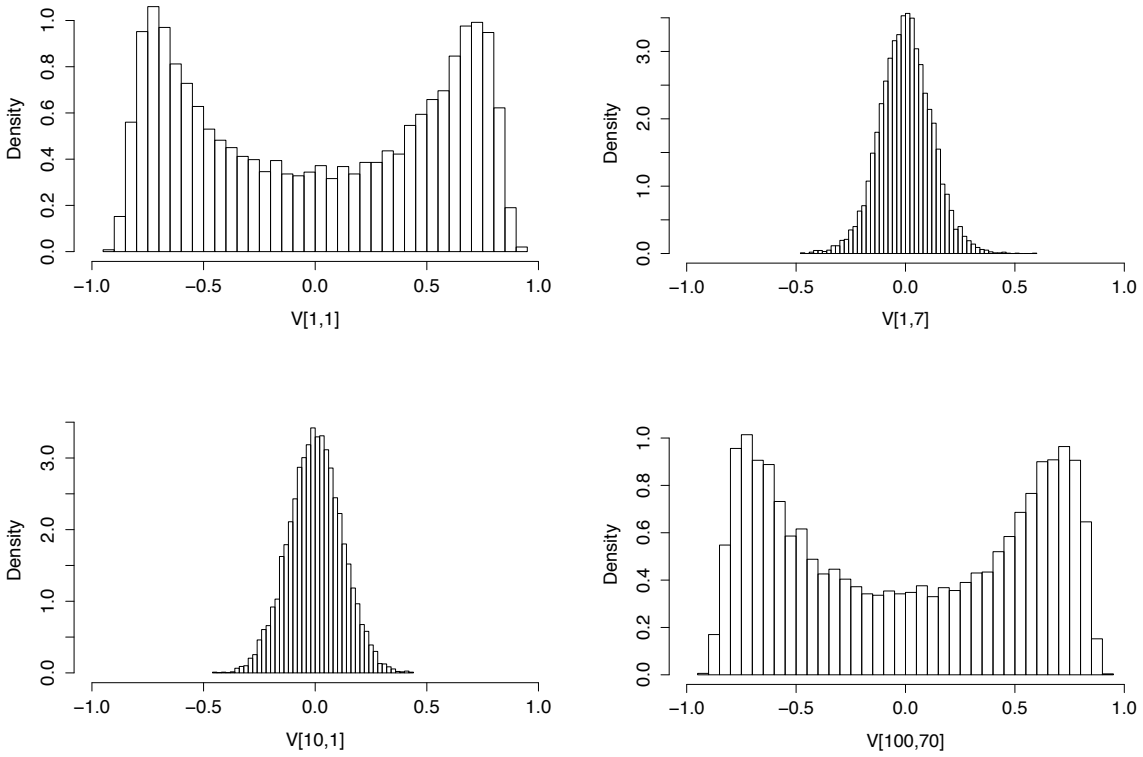


Figure 6: Distribution of $V_{1,1}, V_{10,1}, V_{1,7}$ and $V_{100,70}$ for $N = 10000, n = 100, k = 70$.

that "symmetric" entries seem to be identically distributed across both "diagonals".

To statistically verify this assumption we designed following test. Firstly, we implemented two-sample Kolmogorov-Smirnov test and tested each entry $\left\{v_{ij}\right\}_{i,j=1}^{n,k}$, with its "symmetric" (across the main "diagonal") counterpart, i.e. $v_{jq,i/q}$, where $q = \frac{n}{k}$ and $jq, i/q$ are rounded to be integers. This produced $P_1 \in \mathbb{R}^{n\times k}$ matrix of 7 000 *p-values*, each made from pair of 10 000 independent samples. In the second step, we repeated this procedure for the "other diagonal" (i.e. v_{ij} with $v_{j(k-j+1),i/q}$), producing

$P_2 \in \mathbb{R}^{n\times k}$. Results of these tests can be seen in **Table 1**. We assume that higher than expected percentage of failed tests are due to the inaccuracies caused by rounding of "symmetric" coordinates.

p-value	P_1	P_2
≤ 0.05	15.14%	11.03%
≤ 0.01	6.37%	4.06%

Table 1: Percentage of tests with *p-value* bellow 0.05 (resp. 0.01).

Other interesting property of V is the distribution of the corner entries $v_{1,1}$ and $v_{100,70}$. From histograms in **Figure 6** we see that they are most frequently located around $\approx 0.7 = \frac{70}{100}$. To systematically approximate the most frequent value, we divided values into 40 bins (centered at $0.025 + 0.05k$; $k \in [-20, 19] \subset \mathbb{Z}$) and found the most often realized bin. We repeated this process for fixed $n = 100$ and for different k. Results can be seen in **Table 2**. We see that the most frequent value v_{11} or $v_{n,k}$ is typically $\pm\frac{k}{n}$. This however is not true, for very small k.

k	$\mathbf{v_{11}}$	$\mathbf{v_{100,k}}$	k	$\mathbf{v_{11}}$	$\mathbf{v_{100,k}}$
40	0.225	−0.225	70	0.725	−0.725
45	0.225	−0.175	75	0.775	0.775
50	0.375	−0.375	80	−0.825	−0.825
55	0.525	−0.575	85	−0.875	−0.875
60	−0.625	−0.675	90	0.925	−0.925
65	−0.675	−0.675	95	0.975	−0.975

Table 2: Most frequent values of v_{11} and $v_{100,k}$, depending on k.

In the following chain of equations we use abbreviations: $x_j = v_j(B)$, and $y_i = v_i(A)$:

$$\begin{aligned}\lambda_j^2(B) = \|Bx_j\|^2 = x_j^T\, P^T APP^T AP\, x_j \;\approx\; x_j^T\, P^T A^2 P\, x_j \\ = \langle A\, Px_j, A\, Px_j\rangle = \|A\, Px_j\|^2,\end{aligned} \tag{3.2.2}$$

where the level of approximation in third step is subject to the orthogonal compression $P \in \mathbb{R}^{n\times k}$. We know that $P^TP = I^{k\times k}$, however it is unclear what the random orthogonal projection PP^T should be. Note that $PP^T \xrightarrow{k\to n} I^{n\times n}$, so the approximation as well as equation 3.2.2 is more accurate for lower dimensionality reduction (higher k). Now we can express $Px_j \in \mathbb{R}^n$ as linear combination of columns of Q_A: $\exists \vec{c_j} = (c_{j1}, \dots c_{jn})^T \in R^n : \quad Px_j = \sum_{i=1}^n c_{ji} y_i = Q_A c_j$ and revisit the 3.2.2:

$$\begin{aligned}\lambda_j^2(B) \;\approx\; \|A\, Px_j\|^2 = \|A\, Q_A c_j\|^2 = c_j^T\, Q_A^T A^2 Q_A\, c_j \\ \approx\; c_j^T \Lambda_A^2 c_j = \sum_{i=1}^n c_{ji}^2 \lambda_i(A)^2.\end{aligned} \tag{3.2.3}$$

Also, since Q_A is orthonormal basis and c_j must be of unit length we get $\sum_{i=1}^n c_{ji}^2 = 1$. Thus:

$$0 \;\approx\; \sum_{i=1}^n c_{ji}^2 (\lambda_i^2(A) - \lambda_j^2(B)). \tag{3.2.4}$$

Recall the construction of $V \in \mathbb{R}^{n\times k}$ in current notation: $V_{ij} = \left(Px_j\right)^T \cdot y_i$ and the above introduction of $c_i : Q_A c_j = Px_j \Rightarrow c_j = \left(P^T Q_A\right)^T x_j$. So in fact c_j is the j^{th} column of V, $V_{ij} = c_{ji}$ and columns of V have unitary length.

We continue by estimating the level of required approximation:

$$\begin{aligned}\lambda_j^2(B) = \|Bx_j\|_{\mathbb{R}^k}^2 = \|P^T APx_j\|_{\mathbb{R}^k}^2 = \|PP^T APx_j\|_{\mathbb{R}^n}^2 \\ \leq \|PP^T\|_{op}\, \|APx_j\|_{\mathbb{R}^n}^2.\end{aligned} \tag{3.2.5}$$

We were able to substitute the level of approximation with $\|PP^T\|_{op}$, which isn't very useful, since operator norm of any orthogonal projection with $rank(PP^T) \geq 1$ is 1. Our upper bound can be however strengthened for the typical case using corollary of the Johnson Lindenstrauss lemma [18]:

Note 3.2.1 (Corollary of Johnson Lindenstrauss lemma)
If $y \in \mathbb{R}^n$ is the projection of $x \in \mathbb{R}^n$ onto a random at least $k = \frac{c \log n}{\epsilon^2}$ dimensional subspace, then with probability $1 - \frac{1}{n^{c-2}}$ holds:

$$\left| \sqrt{\frac{k}{n}} \|x\|^2 - \|y\|^2 \right| \leq \epsilon \sqrt{\frac{k}{n}} \|x\|^2. \tag{3.2.6}$$

Using this corollary and 3.2.4, we can write with certain probability p tending to 1, for large n and in respect to k, following chain of equations:

$$\begin{aligned} \lambda_j^2(B) = \|PP^T APx_j\|_{\mathbb{R}^n}^2 &\overset{p}{=} \sqrt{\frac{k}{n}} \|APx_j\|_{\mathbb{R}^n}^2 \\ 0 &\overset{p}{=} \sum_{i=1}^{n} c_{ji}^2 \left(\sqrt{\frac{k}{n}} \lambda_i^2(A) - \lambda_j^2(B) \right). \end{aligned} \tag{3.2.7}$$

Based on 3.2.7 we hypothesize that V is dependent on distribution of $\lambda_i(A)$ and $\lambda_j(B)$. To get some insight, whether V also depends on distribution of eigenvectors $v_i(A)$ and $v_j(B)$, we propose the following test. Firstly, we sample 10 000 $V \in \mathbb{R}^{n \times k}$ matrices, for A from each of three different categories of Wigner ensembles: GOE (V_G), Bernoulli[7] (V_B) and Uniform[8] (V_U) ensembles. We then applied Kolmogorov-Smirnov two sample test on data for each pair V_G, V_B, V_U. This produced three matrices $P_{GB}, P_{GU}, P_{BU} \in \mathbb{R}^{100 \times 70}$ of 7 000 *p-values*, each made from pair of 10 000 independent samples. Percentage of failed tests on individual levels of significance can be found in **Table 3**. Results suggest that V is similarly distributed for different types of Wigner matrices, with various distributions of $v_i(A)$ and $v_j(B)$.

p-value	P_{GB}	P_{GU}	P_{BU}
≤ 0.05	4.77%	4.93%	4.63%
≤ 0.1	9.6 %	9.49%	9.26%

Table 3: Percentage of tests with *p-value* bellow 0.05 (resp 0.01).

This is an evidence that there exists certain universality for V similar to semicircular law.

[7] A_{ij} equals to ± 1 with equal probability.
[8] A_{ij} are uniformly distributed around 0, with unitary variance.

We continue by extracting nk three-dimensional points from $\overline{V}_{GUE}$, where $x = \overline{\lambda}_j(B)$, $y = \overline{\lambda}_i(A)$ [9] and $z = \left(\overline{v_{ij}}^2\right)^{-1}$. In **Figure 7** we see, that points: $\left(\overline{\lambda}_j(B),\, \overline{\lambda}_i(A),\, \left(\overline{v_{ij}}^2\right)^{-1}\right)_{i,j}^{n,k}$ form roughly quadratic shape.

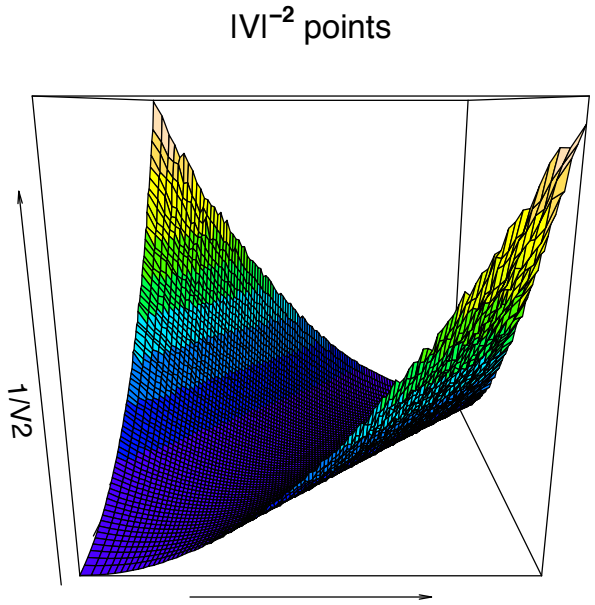


Figure 7: Plot of 100×70 points extracted from $\left(\overline{V}_{GUE}\right)$.

It turns out, this shape is well approximated off-diagonal (i.e. $x \neq y$) by following equation:

$$f(x, y) = \left(x - \sqrt{\frac{k}{n}}y\right)^2. \tag{3.2.8}$$

If we render V for 100×100 Wishart matrices, with 150 degrees of freedom, $\overline{V}$ has different shape, with asymmetrically raised diagonal. Recall, that spectra of Wishart matrices behave according to Marchenko-Pastur Law, we described in 2.6.1. It seems that eigenvectors whose eigenvalues are located on sparse part of spectrum tend to preserve inner products at higher rate than eigenvectors whose eigenvalues are located on dense parts of the spectrum.

[9] For clarification; this is mean of 10 000 j^{th}/i^{th} matrix eigenvalues, not mean made from the whole spectra of a single matrix.

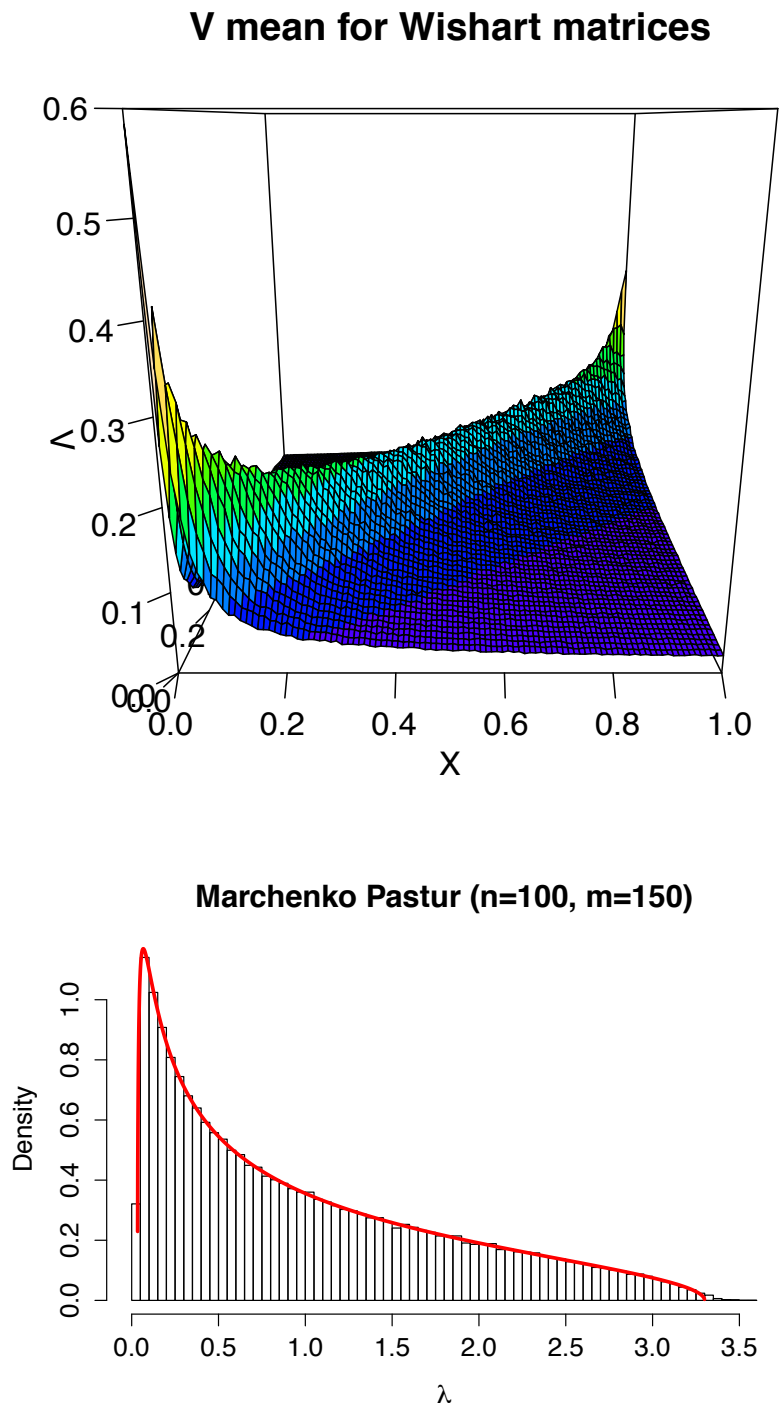


Figure 8: Results $\overline{V}$ for Wishart matrix with 150 degrees of freedom $N = 1000, n = 100, k = 70$, with underlying eigenvalue distribution.

4 Random ellipsoids

Every matrix (especially rectangular) defines several subspaces in $\mathbb{R}^n$ like column or row space, therefore random matrix represents randomly picked subspace, using certain distribution. In this manner, we can say that if we take some k columns of Q_A, where A is $n \times n$ GOE matrix, subspace generated by these vectors, so called k-frame, will be randomly uniformly rotated.

On the other hand square matrices also represent linear transformation $\mathbb{R}^n \to \mathbb{R}^n$, whose properties are defined by their eigenvalues and eigenvectors. Random matrix can be then viewed as an random linear transformation, with asymptotic properties based on underlying asymptotic behaviour of its eigenvalues and eigenvectors. We now define important matrix quantity, related to the matrix operator.

Definition 4.0.1 (Rayleigh quotient)
The Rayleigh quotient of the matrix $A \in \mathbb{R}^{n\times n}$ for the nonzero vector $x \in \mathbb{R}^n$ is the scalar:

$$R(A, x) = \frac{x^T A x}{x^T x} \in \mathbb{R}. \tag{4.0.9}$$

Note 4.0.2
Let $y = v_i(A)$ be an eigenvector of A. Then Rayleigh quotient is equal to the corresponding eigenvalue :

$$\frac{y^T A y}{y^T y} = \frac{y^T(\lambda_i(A)y)}{y^T y} = \lambda_i(A).$$

Rayleigh quotient is a generalization of eigenvalues; it gives us information how vector y is stretched using the transformation matrix A. Notice that if we express $y \in \mathbb{R}^n$ in terms of eigenvectors $y = c_1 v_1(A) + \ldots + c_n v_n(A) = Q_A\,\vec{c}$, then:

$$\begin{aligned} \frac{y^T A y}{y^T y} &= \frac{c^T \overbrace{Q_A^T A Q_A}^{\Lambda_A} c}{c^{\mathsf{H}} Q_A^{\mathsf{H}} Q_A\, c} = \frac{c^T \Lambda_A c}{c^T c} \\ &= \frac{\lambda_1 |c_1|^2 + \ldots + \lambda_n |c_n|^2}{|c_1|^2 + \ldots + |c_n|^2}, \end{aligned} \tag{4.0.10}$$

where $c_1, \dots c_n$ are coefficients determining contribution of the individual eigenvectors in y. Rayleigh quotient for real symmetric A is therefore located inside its spectrum:

$$\lambda_n = \min_{y \in \mathbb{R}^n} \frac{y^T A y}{y^T y} \quad \leq \quad \frac{y^T A y}{y^T y} \quad \leq \quad \max_{y \in \mathbb{R}^n} \frac{y^T A y}{y^T y} = \lambda_1.$$

Definition 4.0.2 (Rayleigh transformation)

Let $A \in \mathbb{R}^{n \times n}$ be some symmetric matrix. Then Rayleigh transformation $T_A : \mathbb{R}^n \to \mathbb{R}^n$ is defined for $x \in \mathbb{R}^n$ as follows:

$$T_A(x) = R(A, x) \cdot x = \frac{x^T A x}{x^T x} \cdot x. \tag{4.0.11}$$

If A is $n \times n$ symmetric matrix and we transform unitary vector $x = \sum_1^n c_i v_i(A) = Q_A \vec{c}$ we get:

$$T_A(x) = \frac{x^T A x}{\underbrace{x^T x}_{1}} x = \left(\sum_{i=1}^{n} \lambda_i(A) c_i^2 \right) \cdot x, \tag{4.0.12}$$

where $\sum_{i=1}^n c_i^2 = \|c\| = 1$. Thus, T_A transforms unitary sphere $S_n = \{x : \|x\| = 1\}$ into an ellipsoid with semi-principal axes determined by eigenvectors $v_i(A)$ and their lengths determined by absolutes of eigenvalues $|\lambda_i(A)|$ as illustrated in **Figure 9**. Moreover, due to the 4.0.10, $T_A(x)$ has a following property:

$$T_A\left(v_i(A)\right) = \lambda_i(A)\, v_i(A). \tag{4.0.13}$$

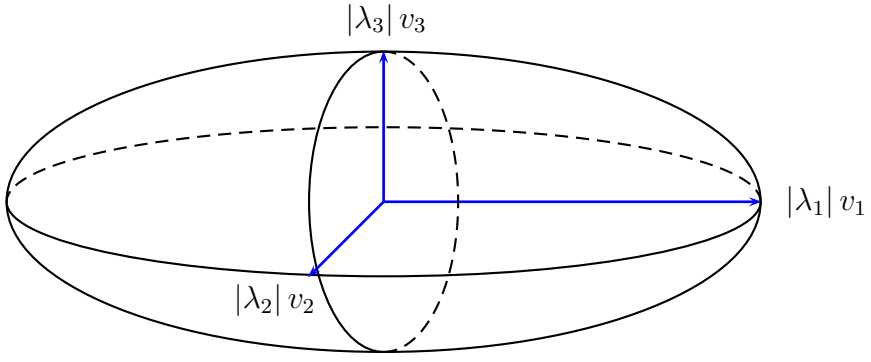


Figure 9: Illustration of ellipsoid created by $T_A(S_3)$.

We now turn our attention to the matrix A and some principal submatrix $B =$

P^TAP, made by the orthogonal compression matrix $P \in \mathbb{R}^{n\times k}$. Let $x \in \mathbb{R}^k$, $\|x\|_{\mathbb{R}^k} = 1$:

$$T_A(Px) = \frac{(Px)^T A\,(Px)}{(Px)^T\,(Px)} \cdot (Px) = \frac{x^T \overbrace{P^TAP}^{B} x}{x^T \underbrace{P^TP}_{I_{k\times k}} x} \cdot (Px) \tag{4.0.14}$$

$$= x^TBx \cdot (Px) = P\,T_B(x).$$

Notice that $Px \in \mathbb{R}^n$ corresponds to n dimensional vector located in k dimensional subspace, determined by columns of P (i.e. k random vectors from standard basis). Therefore if $T_A(S_n)$ represents rotated ellipsoid, $\{T_A(Px) : \|x\| = 1\}$ is going to represent cut through ellipsoid $T_A(S_n)$. This cut is also an ellipsoid determined by $P\,T_B(x)$, with semi-principal axes determined by eigenvectors $Pv_j(B)$ and their lengths determined by absolute values $\big|\lambda_i(A)\,\|v_j(B)\|\big|$.

Recall construction of matrix V in **Chapter 3**. We were interested in inner products $\langle P^Tv_i(A), v_j(B)\rangle = \langle v_i(A), Pv_j(B)\rangle$ for random symmetric matrix $A \in \mathbb{R}^{n\times n}$ and orthogonal compression matrix $P \in \mathbb{R}^{n\times k}$ whose columns were k vectors randomly selected from standard basis.

So in fact, $v_i(A)$ and $v_j(B)$ were semi-principal axes of some n dimensional ellipsoid determined by the symmetric matrix A and semi-principal axes of ellipsoid created by k dimensional cut. Graphical illustration can be seen in **Figure 10**.

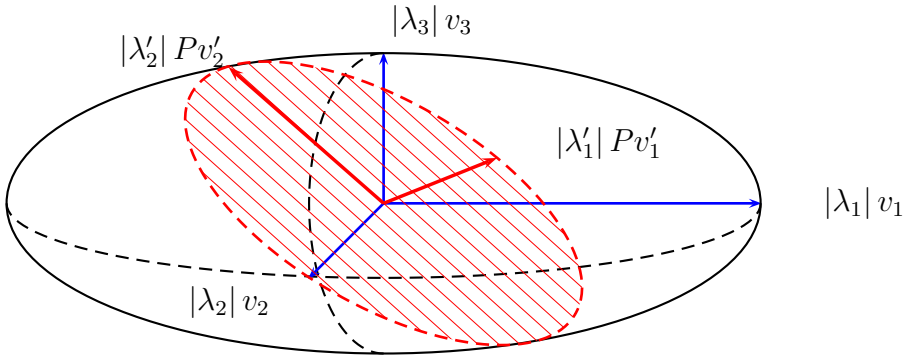


Figure 10: Illustration of $T_A(S_3)$ and cut $P\,T_B(S_2)$; $v_i = v_i(A)$, $\lambda_i = \lambda_i(A)$, $v_j' = v_j(B)$, $\lambda_j' = \lambda_j(B)$.

When A is random, we can use asymptotic spectral properties of random matrices to specify properties of these ellipsoids. First of all, we know that the eigenvalues are

going to be distributed according to the Wigner's Semicircle Law, therefore we know the typical lengths of the semi-principal axes for both ellipsoids. Additionally, we know that eigenvectors of GOE matrices are Haar distributed, i.e. uniformly distributed in the unit sphere. So in fact, GOE matrix, determines uniformly rotated ellipsoid, with lengths of semi-principal axes distributed according to the Wigner's Semicircle Law. We then continue by cutting through this ellipsoid using k vectors from standard basis.

Using this geometric intuition we can rephrase the problem. As opposed to the uniformly rotated ellipsoid and non-rotated cut, we can consider non-rotated ellipsoid and cut through it using uniformly rotated k-frame. This would correspond to A only with diagonal entries (eigenvectors of A will be formed by standard basis) distributed according to the semicircular law and orthogonal compression P consisting of k orthogonal vectors from Haar distribution.

Moreover by using the invariance of Haar measure, we can rotate both, the ellipsoid and the k-frame used to cut through this ellipsoid and get the same results. We were able to validate this intuition by experiments.

Using the mentioned geometric interpretation, we can also provide some intuitive explanation for the shape of V matrix in **Figure 5** in **Chapter 3**. Asymmetricity on the main diagonal of V most likely originate from the fact that cut through the ellipsoid by uniformly rotated k-frame, tends to have lesser effect on longer semi-principal axes.

5 Conclusion

There were two main aims for this thesis: to introduce reader to random matrix theory, and to observe relationship between spectral properties of a random symmetric matrices and their random submatrices.

In the third chapter, we defined the main goal of this thesis in mathematical terms. First of all, we showed two ways for prediction of spectra for random symmetric submatrices; one using Cauchy's Interlacing Theorem and the other using asymptotic distributions of eigenvalues of random symmetric matrices.

We observed several interesting phenomena between the inner products of eigenvectors of random symmetric matrices and their random submatrices. Where it was possible, we used statistical testing for their verification. Here we also supply a reader with scripts, we were used for a generation of various random matrices, and their statistical analysis, respectively. In addition, we were able to derive the equation 3.2.7, giving some information about the inner products of eigenvectors in terms of eigenvalues. Conclusion of this chapter was a hypothesis that observed phenomena happens regardless of eigenvector distribution of the original matrix. We postulated another hypothesis, that eigenvectors corresponding to the eigenvalues on the sparse part of the spectra, tends to be typically less affected by a random orthogonal projection of the matrix.

In the last chapter we try to provide a reader with a geometrical interpretation of the phenomena. We introduce a transformations based on Rayleigh quotients and describe ellipsoids they form. It is showed that random submatrices correspond to randomly rotated cuts through these ellipsoids. This allows us to rephrase this problem and give new directions in this research.

There are still many further unanswered questions about this topic and statistical tests one should examine. Many of these effects however, can not be easily explained, as thorough understanding of the random matrix theory requires a good understanding of many mathematical areas, including probability theory, complex analysis and linear algebra. Random matrices usually proved a useful insight into the underlying problem.

References

[1] I. M. Johnstone. High dimensional statistical inference and random matrices. In *Proceedings of International Conngress of Mathematicians*, pages 1–28, 2006.

[2] E. P. Wigner. Random Matrices in Physics. *SIAM Review*, 9:1–23, 1967.

[3] F. J. Dyson. Statistical theory of the energy levels of complex systems. *Mathematical Phys.*, 3:140–175, 1962.

[4] T. Tao. Topics in random matrix theory. http://terrytao.files.wordpress.com/2011/02/matrix-book.pdf, 2011. [Online; accessed 10-November-2014].

[5] T. Tao. Princeton Minerva Lectures: Universality for Wigner random matrices. University Lecture. https://youtube.com/watch?v=tihxQGGrMcc, 2013. [Online; accessed 19-April-2015].

[6] C. A. Tracy, H. Widom. On Orthogonal and Symplectic Matrix Ensembles. *Communications in Mathematical Physics*, 177:727–754, 1996.

[7] A. Feier. Methods of proof in random matrix theory. Bachelor thesis, Harvard University, 2012.

[8] T. Kemp. Introduction to Random Matrix Theory: Lecture Notes. http://www.math.ucsd.edu/~tkemp/247A/247A.Notes.pdf, 2013. [Online; accessed 12-April-2015].

[9] B. Virág. Wigners Semicircle Law: Lecture 1 Notes. http://wiki.math.toronto.edu/TorontoMathWiki/images/b/be/RMT_Lecture1.pdf, 2009. [Online; accessed 26-May-2015].

[10] T. Tao. When are eigenvalues stable? https://terrytao.wordpress.com/2008/10/28/when-are-eigenvalues-stable/, 2008. [Online; accessed 23-May-2015].

[11] T. Tao, V. Vu. Random matrices have simple spectrum. 2006. [Preprint; arXiv:1412.1438].

[12] P. Diaconis. Patterns in eigenvalues. *Bulletin of the American Mathematical Society*, 40(2):155–178, 2003.

[13] T. Tao, V. Vu. Random Matrices: Universal properties of eigenvectors. 2011. [Preprint; arXiv:1103.2801].

[14] V. A. Marchenko, L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457, 1967.

[15] B. Valkó. Marchenko-Pastur Law: Lectures 6-7 Notes. http://www.math.wisc.edu/%7Evalko/courses/833/2009f/lec_6_7.pdf, 2009. [Online; accessed 26-May-2015].

[16] V. Gregor. The Relations of Spectra of Matrix and its Submatrices. Bachelor thesis, Comenius University in Bratislava, 2013.

[17] D. Adler, D. Murdoch, and others. *rgl: 3D visualization device system (OpenGL)*, 2014. R package version 0.93.1098.

[18] U. V. Vazirani, S. Rao. Algorithms (CS270): Lectures 17 Notes. http://www.cs.berkeley.edu/~satishr/cs270/sp11/rough-notes/measure-concentration.pdf, 2011. [Online; accessed 29-May-2015].