Beyond classical random matrix ensembles: some results on deviation principles

Jens C. Sommerauer

Dissertation
zur Erlangung des Doktorgrades der Naturwissenschaften
an der Fakultät für Mathematik
der Ruhr-Universität Bochum

2007
# Contents

## Introduction

<table>
<thead>
<tr>
<th>Section</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Large and moderate deviation principles</td>
<td>13</td>
</tr>
<tr>
<td>1.1 Formal definition of large and moderate deviations</td>
<td>14</td>
</tr>
<tr>
<td>1.1.1 Cramér’s theorem</td>
<td>15</td>
</tr>
<tr>
<td>1.1.2 Weak LDP</td>
<td>16</td>
</tr>
<tr>
<td>1.1.3 Exponential tightness</td>
<td>17</td>
</tr>
<tr>
<td>1.2 Sanov’s theorem</td>
<td>18</td>
</tr>
<tr>
<td>1.3 Gärtner-Ellis approach</td>
<td>19</td>
</tr>
</tbody>
</table>

## Random Matrix Theory

<table>
<thead>
<tr>
<th>Section</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 One-matrix ensembles</td>
<td>23</td>
</tr>
<tr>
<td>2.1 The classical Gaussian ensembles</td>
<td>24</td>
</tr>
<tr>
<td>2.1.2 Chiral ensembles</td>
<td>28</td>
</tr>
<tr>
<td>2.1.3 BdG-ensembles</td>
<td>29</td>
</tr>
<tr>
<td>2.1.4 A system of disordered bosons</td>
<td>31</td>
</tr>
<tr>
<td>2.2 A Multi-Matrix model</td>
<td>33</td>
</tr>
</tbody>
</table>

## Large Deviation Principle for biorthogonal ensembles

<table>
<thead>
<tr>
<th>Section</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 Biorthogonal ensembles</td>
<td>39</td>
</tr>
<tr>
<td>3.1 Model assumptions and notations</td>
<td>40</td>
</tr>
<tr>
<td>3.2 Main Theorem</td>
<td>41</td>
</tr>
<tr>
<td>3.3 Preliminary Lemmas</td>
<td>42</td>
</tr>
<tr>
<td>3.4.1 Goodness of rate function</td>
<td>48</td>
</tr>
<tr>
<td>3.4.2 Exponential tightness</td>
<td>50</td>
</tr>
<tr>
<td>3.5 Proof of the main theorem</td>
<td>51</td>
</tr>
<tr>
<td>3.5.1 Obtaining a weak LDP</td>
<td>51</td>
</tr>
<tr>
<td>3.5.2 Extending the weak LDP into a full one</td>
<td>61</td>
</tr>
<tr>
<td>3.6 Examples</td>
<td>62</td>
</tr>
</tbody>
</table>
CONTENTS

4 Fluctuations in a Multi-Matrix model .................................................. 69
  4.1 Notations ......................................................................................... 70
    4.1.1 The measure under investigation ............................................. 70
    4.1.2 The potential $V$ .................................................................. 70
  4.2 Results ......................................................................................... 73
  4.3 Excursus: Map enumeration and matrix integrals ........................... 75
    4.3.1 Basic notions for map enumeration ....................................... 75
    4.3.2 One-matrix case .................................................................. 76
    4.3.3 Multi-matrix case ................................................................. 88
  4.4 Rate function ............................................................................... 92
  4.5 Proofs ......................................................................................... 93
  4.6 Generalization ........................................................................... 97
  4.7 Comparisons in the one-matrix model ........................................ 99
  4.8 Examples .................................................................................... 102

List of Symbols .................................................................................... 107

List of Figures .................................................................................... 111

Bibliography ....................................................................................... 111
Introduction

In probability theory, the theory of large deviations has evolved during the last decades as a powerful tool to assess unlikely events by quantifying small probabilities on an exponential scale. In the same period of time, random matrix theory was discovered by probabilists, who became active in that field, which was up to 1990 mainly occupied by researchers in theoretical physics. This thesis resides within the intersection of the theory of large deviations and random matrix theory.

The theory of large deviations

The formal cornerstone of large deviation theory is a monograph by Varadhan [47]. The theory was further developed from there, such that now, at least three textbooks ([13], [12], [11]) give an exposition of the theory and of different tools by which large deviation principles can be deduced.

A large deviation principle (LDP) specifies the asymptotic behavior of a sequence of distributions. The underlying space must be at least a topological one furnished with its Borel $\sigma$-algebra. If a LDP is in place for a sequence of probability measures $(Q_n)_n$ and if $\mathcal{A}$ denotes a measurable set, $\limsup_{n \to \infty} \frac{1}{a_n} \log Q_n(\mathcal{A})$ has an upper bound and $\liminf_{n \to \infty} \frac{1}{a_n} \log Q_n(\mathcal{A})$ can be bounded from below, while the bound is in both cases specified by the same object. That characterization of $Q_n(\mathcal{A})$ on an exponential scale is given in terms of a rate function $I$ and a speed $a_n$, where $I$ maps to $[0, \infty)$ and $(a_n)_n$ is a sequence of real numbers going to infinity for $n \to \infty$. We will give a formal definition of a LDP in chapter one, up to now it suffices to think of the following approximate behavior of $(Q_n)_n$ when obeying a LDP

$$Q_n(\mathcal{A}) \approx e^{-a_n I(\mathcal{A})},$$

where $I(\mathcal{A}) = \inf_{x \in \mathcal{A}} I(x)$. The speed determines the decay rate of unlikely events. The properties of the rate function $I$ play an important role as $I$ usually carries some information of the general setup, e.g. it may be the difference of the entropy and energy of the physical system, from which the analyzed sequence of distributions originated. As we will see, the rate function $I$ always has a minimizer and provided that the minimizer is unique (which is e.g. the
INTRODUCTION

case whenever $I$ is convex), a LDP immediately extends a weak law of large numbers into a strong one, thus the theory of large deviations may provide limit theorems. For a sequence of random variables, we also say that they fulfill a LDP, whenever a LDP holds for the sequence of their distributions.

In 1938, Cramér proved a LDP (he did not name it in that way as the theory was not developed at that time) for the sequence of empirical means of independent and identically (i.i.d.) normally distributed random variables. This theorem was extended to hold for the empirical mean of i.i.d. random variables $X_1, \ldots, X_n$, which have a finite moment generating function (sometimes also called Laplace transform) and this extension is now commonly called Cramér’s theorem. The speed is $n$ and the rate function is always the Legendre-Fenchel transform of the cumulant generating function (i.e. the logarithm of the moment generating function),

$$I(x) = \sup_{\lambda \in \mathbb{R}} \{\lambda x - \log \mathbb{E}[\exp\{\lambda X_1]\}\}.$$  

As the logarithmic moment generating function appears in $I$, the rate function carries the information about all moments.

Cramér’s theorem has been generalized by Gärtner and Ellis to yield a LDP for a sequence of random variables $(X_n)_n$, where the most important condition is the existence of the following limit for all $\lambda \in \mathbb{R}$,

$$\lim_{n \to \infty} \frac{1}{a_n} \log \mathbb{E}[\exp\{a_n\lambda X_n\}] = 0.$$  

It contains Cramér’s theorem as a special case, but its importance stems from the fact that it yields LDPs in situations, where some kind of dependence prevails. The Gärtner-Ellis approach can e.g. be used to derive a LDP for a special class of stationary Gaussian processes or for the empirical measure (which will be defined later on) of an irreducible Markov chain with finite state space. In chapter four, a MDP for traces of words in a multi-matrix model will be obtained via this approach.

For a sequence of random variables (or distributions) fulfilling a LDP and a central limit theorem (CLT), there is usually a $n$-pendent scaling involved in any case, which differs. For a chosen scaling, which lies between that of the CLT and the LDP, it might be possible to also deduce a LDP. Such principle is called a moderate deviation principle (MDP), and the MDP will hold for a range of possible scalings between the two exposed scalings. Usually it holds for the whole range, but in some cases, the range of possible scalings is further restricted, as for example in [16]. The rate function of the MDP may be different to that of the LDP.
and usually involves only the limiting variance of the corresponding random variables in that particular setting. The MDP for the empirical mean of i.i.d. random variables is presented in [11, Sect. 3.7] and in this setting, the rate function is the same as in Cramér’s theorem for Gaussian random variables.

Let \((X_n)_n\) be a sequence of i.i.d. real-valued random variables. The Glivenko-Cantelli theorem states that the uniform convergence of the empirical distribution \(F_n\) to the underlying true distribution function \(F\) of \(X_1\) holds almost surely, where

\[
F_n(x) = \sum_{i=1}^{n} \delta_{\{X_i \leq x\}}.
\]

The empirical distribution function is a special case of the empirical measure, which is defined as

\[
L_n(A) = \sum_{i=1}^{n} \delta_{X_i}(A),
\]

where \(A\) is a measurable set and \(\delta_{X_i}(A)\) is one, if \(X_i \in A\) and 0 otherwise. Note that \(L_n\) is a random measure. Sanov [44] obtained a LDP for the empirical measure on the space of probability measures on \(\mathbb{R}\), with speed \(n\) and the rate function turns out to be the famous relative entropy. Sanov-type results have recently been obtained for the eigenvalues of certain classes of matrices, see [4], [17], [31] and chapter three of this thesis.

**Random matrices**

During the thirties of the last century, large random matrices first appeared in statistics, when Wishart studied matrices of form \(Y = XX'\), and in the sixties physicists like Wigner and Dyson studied random matrices, but it was only in the nineties, when probabilists took a genuine interest in random matrix theory.

The enormous popularity, that the theory of random matrices gained during the last decades in the mathematical community is due to its appearance in many and diverse fields of mathematics. Classical results or conjectures like the famous Riemann-conjecture could be related to random matrices, since statistical properties of the zeros of the zeta-function were found to be similar to the one for eigenvalues of a special class of matrices, thus impacting number theory. But also recent developments in probability theory in a non-commutative setting, which is called free probability, could be related to random matrices. Some areas of statistics and combinatorics are also concerned with random matrices and the latter connection will be described in some more detail in the course of this thesis, when we explain how the enumeration of maps can be related to matrix integrals. As this dissertation deals with matrices,
INTRODUCTION

which emerged in theoretical physics and which are in the center of current research (e.g. in the SFB/TR 12), we will now be more detailed about the ‘first days’ of random matrix theory in physics.

In physics, more precise, in quantum mechanics, it was Wigner, who introduced random matrices in [49] and [50], when he suggested to approximate the Hamiltonian of a system by a large random matrix. The constraints of the system under investigation impose some condition(s) (usually some kind of symmetry) on the random matrix, which should be chosen as random as possible with respect to these limitations. This led Wigner to the study of a hermitian matrix with independent and identically normal distributed entries. By the method of moments he proved the convergence in probability of

$$\int f(x) L_n(dx) \rightarrow \int f(x) \sigma(dx)$$

for any continuous, bounded and real-valued function $f$, where $\sigma$ is the density of the semicircle distribution,

$$\sigma(x) = \frac{1}{2\pi \sqrt{4 - x^2}}1_{[-2,2]}.$$

This result is universal in the sense that it holds whenever the matrix entries are i.i.d. and follow a distribution that fulfills some moment conditions.

For independent centered Gaussian entries, the distribution of the matrix entries is proportional to

$$e^{-c \text{tr}(X^2)},$$

where $\text{tr}(X)$ denotes the trace of a matrix $X$ and the matrix $X$ is a hermitian matrix. The constant $c$ depends on the actual choice of the variance. Imposing some time-reversality on the system, it could be shown that the entries are no longer complex, but real, which results in $X$ being symmetric. And for yet another notion of time-reversality, quaternion entries are allowed, in which case $X$ has a certain block structure. If the matrix entries follow a normal distribution in the real and quaternion case, the distribution of the matrix entries is also of type (0.1) but with different constants $c$. Thus it can be written in one closed form, where only a newly introduced parameter $\beta$ takes different values, i.e. $\beta = 1$ for real entries, $\beta = 2$ for complex one and $\beta = 4$ in the last case.

Dyson classified these three classes of matrices (which we will call ensembles from now on) in [15] based on invariance. The distribution of the matrix entries (0.1) for these ensembles was found to be invariant under the following transformation of the matrix $X$,

$$X \rightarrow UXU^{-1}.$$

For real entries, the matrix $U$ is orthogonal, while it is unitary resp. symplectic for complex resp. quaternion matrix entries. Due to the invariance under the specific group action,
the ensembles are called Gaussian orthogonal ensembles (GOE), Gaussian unitary ensembles (GUE) and Gaussian symplectic ensembles (GSE).

Moreover, Wigner derived the joint distribution of the eigenvalues (also referred to as eigenvalue distribution) in terms of specifying an explicit form of its density in the GUE case, whereas it can be written in a closed form embracing all three ensembles,

\[ \frac{1}{Z_{\beta,N}} \prod_{1 \leq j < k \leq N} |\lambda_k - \lambda_j|^{\beta} e^{-\beta^{\frac{4}{\beta}} \sum_{j=1}^{N} \lambda_j^2}, \]

where \( \beta \in \{1, 2, 4\} \) and \( Z_N \) is the partition function providing that the distribution of eigenvalues is a probability distribution.

The derivation of the eigenvalue distribution for these three ensembles had an enormous impact on quantum mechanics, because the energy states of a system are represented by the eigenvalues of the corresponding Hamiltonian, the energy gaps are described by the spacing distribution (of eigenvalues) and the ground states are in correspondence with the extremal eigenvalues. With a closed form of the joint distribution of the eigenvalues at hand, all these distributions or objects could be obtained resp. analyzed and a good exposition from a physical point of view can be found in [40], whereas such questions became popular in mathematics since the paper by Tracy and Widom [45].

Recent developments

During the last years, mesoscopic physics has been a source, from which new matrix ensembles other than the three classical ensembles have sprung up. These ensembles occurred by the analysis of disordered systems of fermions, such as Dirac operators in a random gauge field or disordered superconductors. Fermions are particles with half-integer spins, that obey Pauli’s exclusion principle, which says that two fermions cannot have the same quantum state simultaneously. In mathematical terms, the exclusion principle translates to having an antisymmetric wave function. In total, there are ten ensembles describing fermions, including the three classical ensembles. Each ensemble is also invariant with respect to some group action and these ten ensembles are in one-to-one correspondence with Riemannian symmetric spaces, see [29], which were classified by Cartan.

The LDP for the empirical measure of eigenvalues for any of the ten ensembles was obtained in [17].

In contrast to fermions, bosons are particles with an integer spin and for which the wave function is symmetric. Only recently, a matrix ensemble describing a system of disordered bosons
INTRODUCTION

was discovered in [37], where the eigenvalue distribution was obtained as well. This form is a special case of the eigenvalue distribution of biorthogonal ensembles that were studied by Borodin in [6]. Borodin assumed the eigenvalue distribution to be of a special form, in which weight functions appear, that are chosen according to the weight functions of biorthogonal polynomials and derived the correlation function. While his treatment assumed to have a eigenvalue distribution of a certain type, the authors of [37] gave an explicit model of disordered bosons, for which they derived the eigenvalue distribution. It corresponds to biorthogonal ensembles with Laguerre weight functions.

Instead of deriving a LDP for the empirical measure of eigenvalues of the system of disordered bosons only, the LDP can be obtained for the empirical measures of eigenvalues from any biorthogonal ensemble and this will be the content of the third chapter.

For the GUE, the density of the Lebesgue measure on the matrix entries is given by (0.1), where the constant \( c \) equals \( \frac{N}{2} \), but it can also be understood as a density for the Lebesgue measure on the space of \( N \times N \) hermitian matrices. Another development in random matrix theory was to study a more general situation than the GUE case. Replacing \( X^2 \) by \( X^2 + 2V(X) \), where \( V \) is a general polynomial leads to a measure of form

\[
\frac{1}{Z_N} e^{-\frac{N}{2} \text{tr}(X^2 + 2V(X))} dX,
\]

where \( dX \) denotes the Lebesgue measure on the space of hermitian \( N \times N \) matrices. In quantum field theory, such measures arise first in [5] and [7], where \( V(X) = tX^4 \), \( t \in \mathbb{R} \). One central object was the partition function \( Z_N \) for which an asymptotic expansion of \( \log Z_N \) was stated in [5] and it was given in terms of formal identities for non-converging power series. A rigorous proof of this expansion only appeared in 2003, [20], where general polynomials \( V \) were considered, whose coefficients lie in a neighborhood of the origin. By sophisticated Riemann-Hilbert techniques, the authors gave an asymptotic expansion of \( \log Z_N \) in powers of the inverse of the matrix size \( N \), along with a bound for the approximation error when only evaluating the expansion up to a fixed number of terms.

This model is a perturbed version of the GUE, where the perturbation is given by \( V \). But it can also be understood in terms of statistical mechanics by describing a system of particles with logarithmic interaction potential in the presence of an external field, that is given in terms of a potential \( V \). For this reason, we will often refer to \( V \) as potential.

A further generalization of the last model is the so-called multi-matrix model, which allows for \( m N \times N \) hermitian matrices. The potential \( V \) is now a function of \( m \) matrices. A monomial
\( q_i \) in \( m \) variables \( X_1, \ldots, X_m \) is of type

\[
q_i(X_1, \ldots, X_m) = X_{i_1} \cdots X_{i_{k_i}}, \text{ with } i_j \in \{1, \ldots, m\} \text{ for } j = 1, \ldots, k_i, \text{ and } k_i \geq 1.
\]

These monomials constitute \( V \), \( V(X_1, \ldots, X_m) = \sum_{i=1}^n t_i q_i(X_1, \ldots, X_m) \), where \( n \in \mathbb{N}, t_i \in \mathbb{C}, i = 1, \ldots, n \).

The measure, which now lives on the space of \( m \)-tuples of hermitian matrices, is

\[
\frac{1}{Z_N^V} e^{-\frac{N}{2} \text{tr}(\sum_{i=1}^m X_i^2 + 2V(X_1, \ldots, X_m))} dX.
\]

Because the matrices are hermitian and we have allowed for complex coefficients \( t_i \) it cannot a priori be ruled out that the expression \( \text{tr}(\sum_{i=1}^m X_i^2 + 2V(X_1, \ldots, X_m)) \) is complex. For this reason, the potential \( V \) has either to fulfill a condition that \( \text{tr}(V) \) is real or it is assumed to be self-adjoint, from which it can be deduced that \( \text{tr}(V) \) is real.

An expansion of the partition function \( Z_N^V \) in the multi-matrix model was recently obtained in [39], whenever \( V \) fulfills a certain convexity condition. This expansion is our starting point for calculating the cumulant generating function for a monomial \( q_i \) of the above type under the measure (0.2), which allows us to deduce a MDP for the centered trace of \( q_i \) in the fourth chapter.

This analysis is closely related to combinatorics. As in the asymptotic expansion of \( \log Z_N^V \) the coefficients for each \( N^{-2g}, g \in \mathbb{N} \), can be understood as generating functions of some special types of maps. Thus it touches on a problem of graph enumeration, while the potential \( V \) influences, which type of graphs we are counting. For the multi-matrix models, the edges of the graphs are colored, in contrast to the one-matrix model, where non-colored graphs are counted.

The thesis is structured as follows. In the first chapter, we review some facts about the theory of large deviations, where we concentrate on theorems and lemmas, which will be applied later on. Furthermore, we state theorems which are connected to objects, that we will also focus on, e.g. the empirical measure. In chapter two, a small account on the vast field of random matrix theory is given, by introducing the ten ensembles for disordered fermions, the ensemble for disordered bosons and the multi-matrix model in more detail.

The third chapter deals with biorthogonal ensembles. Based on the joint density of the eigenvalue distribution, a LDP for the empirical distribution of eigenvalues will be developed and proven and we explain how this LDP can be used to deduce a strong law of large numbers for the empirical measure. The chapter finishes by applying the LDP to three classes of biorthogonal ensembles, where one ensemble describes a system of disordered bosons. For this system of bosons, a weak law of large numbers for the mean empirical measure was deduced in [37] and our LDP extends that analysis by studying the fluctuations of the empirical measure.
around its limit, thus moving to the level of random measures. Moreover, we extend the weak law into a strong law of large numbers by applying the Borel-Cantelli lemma and the upper bound of the LDP.

Finally, the framework in chapter four is the multi-matrix model, and we prove a MDP for the normalized traces of monomials. According to our knowledge, the paper [10] is up to now the only one, in which moderate deviations in connection with random matrices have been studied. Along the lines of our proof, we see how we can give an alternative proof of a CLT for traces of words, that has first been proven in [25]. As the appearing rate function (for the MDP) and variance (for the CLT) involve generating functions for maps, we give an extensive excursus on map enumeration. We close by relating the MDP to certain multi-matrix models such as the random Ising model on random graphs or the $q$-Potts model.

I would like to express my deep gratitude to Prof. Dr. Peter Eichelsbacher for his great support throughout the last years. I highly appreciate his encouragement to study abroad and his excellent supervision during my doctoral studies. I am also grateful to Dr. Michael Stolz for many helpful discussions.

Moreover, I am also indebted to the “Studienstiftung des deutschen Volkes” for their dissertation grant and to Prof. Dr. Dr. h.c. mult. Alan Huckleberry for acting as my “Vertrauensdozent”. My fellow postgraduate students in our group Dipl.-Math. Katrin Hofmann-Credner, Dipl.-Math. Hanna Döring and Thorsten Mehlich deserve thanks for creating such a fantastic atmosphere on an academic as well as on a social level. Finally, I would like to thank my parents, my sister Dipl.-Ök. Britta Ricken and my girlfriend Christine for their continuous support and encouragement over time and distance.
1 Large and moderate deviation principles

Central limit theorems (CLTs) and laws of large numbers are well studied in probability theory, and they describe the limiting behavior of random variables in terms of specifying the limit resp. the limit distribution. In some situations convergence rates are available. But most of the time, all we know about ‘untypical’ or rare events, is that the probability of such events goes to zero. A large deviation principle (LDP) quantifies the asymptotics of probabilities on an exponential scale by a rate function and a speed, which determine e.g. the decay of probabilities of rare events.

The formal foundations of large deviation theory are due to Varadhan, who first laid them down in [46], while he presented some of the fruitful methods to obtain a LDP in his monograph [47], which also includes several examples and applications. From then on, the theory of large deviations became more prominent and all that is mentioned in this chapter can be found in the books [13], [12] and [11].

We will start this exposition by giving an example which is a special case of Cramér’s theorem, which was deduced 1938 in [9], and which fits in quite well in the framework of large deviations developed much later. This example serves to illustrate the idea behind a LDP and urges us to formally define what we mean by saying that a sequence of distributions or random variables obeys a LDP. From there, we move on to state Cramér’s theorem in its general version. We then turn to introduce notions and theorems that will be used in the following chapters so that this thesis might be as self-contained as possible. Therefore, we first define a weaker version of a LDP, which will be called weak LDP. To make the distinction between a weak LDP and a LDP more clear, the latter one will also be referred to as a full LDP. When dealing with a weak LDP one might ask, whether it is possible to also achieve a full LDP. For this reason we introduce the notion of exponential tightness, which allows the transition from a weak to a full principle.

We proceed by considering a theorem from Sanov, which provides a LDP for the empirical measure, an object that we will encounter in chapter three.

Furthermore, we look at some extensions of Cramér’s theorem, as it was carried out by Gärtner
1 LARGE AND MODERATE DEVIATION PRINCIPLES

and Ellis. This result will prepare the ground for what is to come in chapter four.
Let us now start with a nice but easy situation of considering a sequence \((X_i)_{i \in \mathbb{N}}\) of independently identically distributed (i.i.d.) random variables, where \(X_i\) has a Bernoulli distribution with \(P(X_i = 0) = \frac{1}{2} = P(X_i = 1)\). We consider the empirical mean \(\overline{S}_n = \frac{1}{n} \sum_{i=1}^{n} X_i\) and take a look at \(P(\overline{S}_n \geq a)\). When \(a > 1\), this probability is zero, so we first take \(a \in [\frac{1}{2}, 1]\). Since \(P(\overline{S}_n \geq a) = 2^{-n} \sum_{k \geq an} \binom{n}{k}\) the inequalities
\[
2^{-n} \max_{k \geq an} \binom{n}{k} \leq P(\overline{S}_n \geq a) \leq (n+1)2^{-n} \max_{k \geq an} \binom{n}{k},
\]
yield via Stirling’s formula \(n! = n^n e^{-n} \sqrt{2\pi n}(1 + o(1))\):
\[
\lim_{n \to \infty} \frac{1}{n} \log \max_{k \geq an} \binom{n}{k} = -a \log a - (1 - a) \log(1 - a).
\]
Thus, we have found that
\[
\lim_{n \to \infty} \frac{1}{n} \log P(\overline{S}_n \geq a) = -I(a), \tag{1.1}
\]
where
\[
I(a) = \begin{cases} 
\log 2 + a \log a + (1 - a) \log(1 - a) & \text{if } a \in [0, 1] \\
\infty & \text{otherwise}.
\end{cases}
\]
The symmetry relation \(I(a) = I(1 - a)\) provides that the same reasoning as above can be applied for \(a \in [0, \frac{1}{2}]\). Moreover, the (rate) function \(I(a)\) is convex and has a unique minimizer at \(a = \frac{1}{2}\).

1.1 Formal definition of large and moderate deviations

In the example above, the random variable \(\overline{S}_n\) was only allowed to fall in the interval \((-\infty, \frac{1}{2})\) or \((\frac{1}{2}, \infty)\) and the upper and lower bound of \(P(\overline{S}_n \geq a)\) merge on an exponential scale, which results in the limit (1.1). We will now state the definition of a large deviation principle, which covers many more cases than the specific situation considered so far. For all that is to come in this and the forthcoming chapters, we look at sequences, which are defined on \(\Gamma\), which is a Polish space (i.e. a complete and separable metric space). It is endowed with its Borel \(\sigma\)-algebra \(\mathcal{B}_\Gamma\).

Definition 1.1 A sequence of probability measures \((\mu_n)_{n \in \mathbb{N}}\) on a Polish space \(\Gamma\) obeys a large deviation principle with speed \(a_n\) and good rate function \(I(\cdot) : \Gamma \to \mathbb{R}_0^+\) if

- \(I\) is lower semi-continuous and has compact level sets \(N_L := \{x \in \Gamma : I(x) \leq L\}\), for every \(L \in [0, \infty)\).
1.1 FORMAL DEFINITION OF LARGE AND MODERATE DEVIATIONS

- For every open set $G \subseteq \Gamma$, it holds that

$$
\liminf_{n \to \infty} \frac{1}{a_n} \log \mu_n(G) \geq - \inf_{x \in G} I(x). \tag{1.2}
$$

- For every closed set $A \subseteq \Gamma$, it holds that

$$
\limsup_{n \to \infty} \frac{1}{a_n} \log \mu_n(A) \leq - \inf_{x \in A} I(x). \tag{1.3}
$$

Similarly, we will say that a sequence of random variables $(X_n)_{n \in \mathbb{N}}$ with values in a Polish space $\Gamma$ obeys a large deviation principle with speed $a_n$ and good rate function $I(\cdot) : \Gamma \to \mathbb{R}_0^+$ if the sequence of their distributions does.

Formally, a moderate deviation principle is nothing but a LDP. However, we will speak of a moderate deviation principle (MDP) for a sequence of distributions or random variables, whenever the scaling of the corresponding random variables is between that of an ordinary law of large numbers and that of a CLT. Thus, a MDP usually holds for a wide range of scalings and not just for one particular choice, as we will see in chapter four.

Note that the lower semi-continuity of $I$ could be omitted, since this property is equivalent for $I$ to have closed level sets. Following the folklore in large deviations, we yet included it. Moreover, due to the definition of the rate function and the upper bound, it follows that $\inf I = 0$ by taking $A = \Gamma$ in (1.3). Because of the compact level sets, there exists at least one $x \in \Gamma$ with $I(x) = 0$. If the minimizer is unique, the compactness of the level sets also provides a strong law of large numbers by applying Borel-Cantelli’s lemma, see [18, Thm. II B.3].

Finally, remember weak convergence and the Portmanteau theorem to see that a large deviation principle lifts the notion of weak convergence naturally on an exponential scale.

### 1.1.1 Cramér’s theorem

Now, we can state Cramér’s theorem in case of $\Gamma = \mathbb{R}^d$ and imbed it into our formal framework. Let $\langle \cdot, \cdot \rangle$ denote the usual scalar product on $\mathbb{R}^d$, $\langle \lambda, x \rangle = \lambda^t x$, where $\lambda, x \in \mathbb{R}^d$. For a real-valued function $f$ on $\mathbb{R}^d$, we write

$$
\mathcal{D}_f := \{ \lambda \in \mathbb{R}^d | f(\lambda) < \infty \}
$$

and denote the interior of $\mathcal{D}_f$ by $\mathcal{D}_f^o$. The theorem will be stated for the situation of $\Gamma = \mathbb{R}^d$. 

15
1 LARGE AND MODERATE DEVIATION PRINCIPLES

Theorem 1.2 (Cramér) Let \((X_i)_{i \in \mathbb{N}}\) be a sequence of i.i.d. random variables with values in \(\mathbb{R}^d\) with finite moment generating function around the origin, i.e. \(0 \in D^\circ_M\), with

\[
M(\lambda) := \mathbb{E} \left[ e^{\langle \lambda, X_1 \rangle} \right].
\]

Then, \(S_n = \frac{1}{n} \sum_{i=1}^n X_i\) obeys a LDP on \(\mathbb{R}^d\) with speed \(n\) and rate function

\[
I(x) = \sup_{\lambda \in \mathbb{R}^d} [\langle \lambda, x \rangle - \log M(\lambda)].
\]

The rate function \(I\) is the Legendre-Fenchel transform of the cumulant generating function \(\log M\), it is convex and the following duality result holds,

\[
\log M(\lambda) = \sup_{x \in \mathbb{R}^d} [\langle \lambda, x \rangle - I(x)].
\]

This allows us to deduce, that the distribution of a random variable \(X_1\) is uniquely determined by the rate function \(I\).

The proof of Theorem 1.2 can be found in [12] for \(\Gamma = \mathbb{R}\), whereas in [11] it is also given for \(\Gamma = \mathbb{R}, \mathbb{R}^d\) and a situation, where among other assumptions, \(\Gamma\) is a closed convex subset of a locally convex, Hausdorff, topological real vector space.

When proving Theorem 1.2 in the specific situation of \(\Gamma = \mathbb{R}\), one encounters two strategies, which are quite popular and helpful in the theory of large deviations. The upper bound is proved by an application of the Markov inequality with a function \(e^t\). The resulting inequality is also often called exponential Tschebyschev inequality and minimizing it in \(t\) leads to the appearance of the Legendre-Fenchel transform.

For the lower bound, an exponential measure transformation is carried out, such that the former untypical behavior becomes now typical and a central limit theorem (as in [12]) or a law of large numbers (as in [11]) becomes applicable.

1.1.2 Weak LDP

In some situations, the LDP is only fulfilled up to some restrictions, which leads to the following definition.

Definition 1.3 A sequence of probability measures \((\mu_n)_{n \in \mathbb{N}}\) on a Polish space \(\Gamma\) obeys a weak large deviation principle with speed \(a_n\) and rate function \(I(\cdot) : \Gamma \to \mathbb{R}_0^+\) if

1. \(I\) is lower semi-continuous and has closed level sets.
2. The lower bound holds as in (1.2).
• For every compact set $A \subseteq \Gamma$, the upper bound holds as in (1.3).

To distinguish between a LDP and a weak LDP we will sometimes speak of a full LDP in
contrast to a weak one and note that a rate function is only called good, when a full LDP
holds. On Polish spaces, the rate function is also unique if a weak LDP holds, see [11, Ex.
4.1.30].

One situation, in which only a weak LDP can be deduced is dealt with in the next theorem.

**Theorem 1.4 (Theorem 4.1.11 in [11])** Let $\mathcal{A}$ be a base of the topology of the Polish space
$\Gamma$. For every $A \in \mathcal{A}$ define

$$
\mathcal{L}_A := -\lim \inf_{n \to \infty} \frac{1}{a_n} \log \mu_n(A),
$$

and

$$
I(x) := \sup_{\{A \in \mathcal{A} : x \in A\}} \mathcal{L}_A.
$$

If for each $x \in \Gamma$

$$
I(x) := \sup_{\{A \in \mathcal{A} : x \in A\}} \left[ -\lim \sup_{n \to \infty} \frac{1}{a_n} \log \mu_n(A) \right],
$$

then $(\mu_n)_{n \in \mathbb{N}}$ fulfills a weak LDP on $\Gamma$ with rate function $I$.

In the third chapter we will see an application of Theorem 1.4 on the space of probability
measures, of which we first show, that it is Polish. It might be natural to ask, in which
situation it is possible to extend a weak LDP into a full one. The next paragraph describes
one condition which provides such an extension.

### 1.1.3 Exponential tightness

The notation of a tight probability measure will be lifted to an exponential scale in sense of
the following definition.

**Definition 1.5** A sequence of probability measures $(\mu_n)_{n \in \mathbb{N}}$ on $\Gamma$ is said to be exponentially
tight, if for every $M \geq 0$, there exists a compact set $K_M \subset \Gamma$, such that

$$
\lim \sup_{M \to \infty} \lim \sup_{n \to \infty} \frac{1}{a_n} \log \mu_n(K^c_M) = -\infty.
$$

Sometimes, an alternative definition can be found, which is easily seen to be equivalent to
Definition 1.5:

**Definition 1.6** A sequence of probability measures $(\mu_n)_{n \in \mathbb{N}}$ on $\Gamma$ is said to be exponentially
tight, if for every $L \geq 0$, there exists a compact set $K_L \subset \Gamma$, such that

$$
\lim \sup_{n \to \infty} \frac{1}{a_n} \log \mu_n(K^c_L) < -L.
$$

17
1 LARGE AND MODERATE DEVIATION PRINCIPLES

It turns out, that exponential tightness is one condition, which strengthens a weak LDP as is stated in the next lemma.

**Lemma 1.7** If \((\mu_n)_{n \in \mathbb{N}}\) satisfies a weak LDP and if it is exponentially tight, then a full LDP is in place.

In order to show the upper bound for closed sets, observe that for a measurable set \(A \subset \Gamma\) and a compact set \(K_L\) the following holds.

\[
\mu_n(A) \leq \mu_n(\overline{A} \cap K_L) + \mu_n(K_L^c).
\]

While the upper bound can be applied to the compact set \(\overline{A} \cap K_L\) (here, \(\overline{A}\) denotes the closure of \(A\)), \(\mu_n(K_L^c)\) is controlled by the exponential tightness and the upper bound for closed sets is established because of

\[
\limsup_{n \to \infty} \frac{1}{a_n} \log[\alpha_n + \beta_n] = \max \left\{ \limsup_{n \to \infty} \frac{1}{a_n} \log \alpha_n, \limsup_{n \to \infty} \frac{1}{a_n} \log \beta_n \right\}.
\]

The goodness of the rate function can be seen as follows. For a level set \(N_L\), (1.8) provides the existence of a compact set \(K_L\), such that an application of the lower bound to the open set \(K_L^c\) yields

\[
\inf_{x \in K_L^c} I(x) \overset{(1.2)}{=} -\liminf_{n \to \infty} \frac{1}{a_n} \log \mu_n(K_L^c) \geq -\limsup_{n \to \infty} \frac{1}{a_n} \log \mu_n(K_L^c) \overset{(1.8)}{=} L.
\]

Hence, \(N_L \subset K_L\) and therefore \(N_L\) is compact.

Although Definition 1.5 (resp. Definition 1.6) and Lemma 1.7 have already been known and used before, it was in [13], where they explicitly found its entrance into the literature.

### 1.2 Sanov’s theorem

In this section we will state a theorem that goes back to Sanov and contains a LDP for the empirical measure of random variables.

Let \((X_i)_{i \in \mathbb{N}}\) be a sequence of random variables on a probability space \((\Omega_n, \mathcal{F}, P_n)\) with values in a Polish space \(\Gamma\), denote the Borel \(\sigma\)-algebra of \(\Gamma\) by \(\mathcal{B}_\Gamma\) and the law of \(X_1\) by \(\mu\), \(\mathcal{L}(X_1) = \mu\). The empirical measure \(L_n\) is defined by

\[
L_n(\omega, B) := \frac{1}{n} \sum_{i=1}^n \delta_{X_i(\omega)}(B), \quad \omega \in \Omega_n, B \in \mathcal{B}_\Gamma,
\]

(1.9)
1.3 Gärtner-Ellis Approach

where

\[ \delta_{X_k}(\omega)(B) = \begin{cases} 1 & X_k(\omega) \in B \\ 0 & \text{otherwise.} \end{cases} \]

Working on a Polish space \( \Gamma \) guarantees that the space of probability measures on \( \Gamma, \mathcal{M}_1(\Gamma) \), is Polish as well, when it is furnished with the weak topology, which is induced by weak convergence. (For topological considerations of this kind, we refer to [42].) Now, \( L_n := L_n(\omega, \cdot) \) is in \( \mathcal{M}_1(\Gamma) \) and the following theorem was obtained by Sanov [44] for the case \( \Gamma = \mathbb{R} \), while Donsker and Varadhan [14] extended it to general Polish spaces \( \Gamma \).

**Theorem 1.8 (Sanov)** The sequence of empirical measures \( (L_n)_{n \in \mathbb{N}} \) of a sequence of i.i.d. random variables \( (X_i)_{i \in \mathbb{N}} \) with values in a Polish space \( \Gamma \) and law \( \mathcal{L}(X_1) = \mu \) satisfies a LDP on \( \mathcal{M}_1(\Gamma) \) with speed \( n \) and rate function

\[ I_\mu(\nu) = H(\nu|\mu) = \begin{cases} \int \log \frac{d\nu}{d\mu} \, d\nu & \text{if } \nu \ll \mu, \\ \infty & \text{otherwise.} \end{cases} \quad (1.10) \]

The rate function \( H(\nu|\mu) \) is the famous relative entropy and is sometimes also called Kullback-Leibler information of \( \nu \) with respect to \( \mu \).

Concerning the proof of Sanov’s theorem, the proof of the lower bound again involves the application of exponential measure transformations and laws of large numbers, while one crucial step in proving the upper bound is the exponential Tschebyschev inequality. Moreover, the strategy is to first establish a weak LDP and to show the exponential tightness of \( L_n \).

Although the technical details are more complex, there is some resemblance to the proof of Cramér’s theorem. Indeed, a connection between these theorems can be established via the contraction principle. That principle basically says that LDPs carry over under continuous transformations, but under less mild assumptions (as e.g. in Theorem 1.2), since it is assumed that the moment generating function exists on the whole real line, \( M(\lambda) < \infty, \forall \lambda \in \mathbb{R}^d \), cf. [11, Ex. 6.2.21].

Vice versa, the random variables \( \delta_{X_1}, \delta_{X_2}, \ldots \) are i.i.d. on \( \mathcal{M}_1(\Gamma) \) and \( L_n \) is its empirical mean and Theorem 1.8 could also be deduced from Theorem 1.2, when the rate function \( H(\nu|\mu) \) is regarded more generally as a Legendre-Fenchel transform of the cumulant generating function of \( \delta_{X_1} \), see e.g. [11, Chap. 6].

### 1.3 Gärtner-Ellis approach

This section generalizes Cramér’s theorem, where only i.i.d. random variables were allowed. We will in detail state this theorem in case of \( \Gamma = \mathbb{R} \), as this is the situation in which the theorem is applied in the fourth chapter.
Let \( (X_n)_{n \in \mathbb{N}} \) now be a sequence of random variables on \( \mathbb{R} \) with distributions \( (\mu_n)_{n \in \mathbb{N}} \). The starting point is the logarithmic moment generating function

\[
\Lambda_n(\lambda) = \log \mathbb{E} \left[ e^{\lambda X_n} \right], \quad \lambda \in \mathbb{R}.
\]

Next, we introduce a scaling via a sequence \( a_n \), and define for each \( \lambda \in \mathbb{R} \),

\[
\Lambda(\lambda) = \lim_{n \to \infty} \frac{1}{a_n} \Lambda_n(a_n \lambda), \tag{1.11}
\]

and remember

\[
\mathcal{D}_\Lambda = \{ \lambda \in \mathbb{R} | \Lambda(\lambda) < \infty \}.
\]

The Legendre-Fenchel transform of \( \Lambda(\lambda) \) is

\[
\Lambda^*(x) = \sup_{\lambda \in \mathbb{R}} \{ \lambda x - \Lambda(\lambda) \}.
\]

Additionally, for a real-valued function \( f \) on \( \mathbb{R}^d \), we call a \( y \in \mathbb{R}^d \) exposed point of \( f \), if there exists a \( \lambda \in \mathbb{R}^d \), such that for all \( x \neq y \),

\[
\langle \lambda, y \rangle - f(y) > \langle \lambda, x \rangle - f(x),
\]

where such a \( \lambda \) is called an exposing hyperplane.

The Gärtner-Ellis theorem allows to deduce a LDP from the existence of the limit (1.11).

**Theorem 1.9 (Gärtner-Ellis)** If the limit (1.11) exists as an extended real number for all \( \lambda \in \mathbb{R} \) and if \( 0 \in \mathcal{D}_\Lambda^0 \) (where \( \mathcal{D}_\Lambda^0 \) denotes the interior of \( \mathcal{D}_\Lambda \)), the following holds:

(I) For any closed set \( A \subset \mathbb{R} \),

\[
\limsup_{n \to \infty} \frac{1}{a_n} \log \mu_n(A) \leq - \inf_{x \in A} \Lambda^*(x).
\]

(II) For any open set \( G \subset \mathbb{R} \),

\[
\liminf_{n \to \infty} \frac{1}{a_n} \log \mu_n(F) \geq - \inf_{x \in G \cap F} \Lambda^*(x),
\]

where \( F \) is the set of all exposed points of \( \Lambda^* \) whose exposing hyperplanes belong to \( \mathcal{D}_\Lambda^0 \).

(III) If in addition,

1. (a) \( \Lambda \) is lower semi-continuous,
2. (b) \( \mathcal{D}_\Lambda^0 \) is non-empty,
3. (c) $\Lambda$ is differentiable in $\mathcal{D}_\Lambda^o$ and

4. (d) $\lim_{n \to \infty} |\Lambda'(\lambda_n)| = \infty$ for all sequences $(\lambda_n)_{n \in \mathbb{N}}$ in $\mathcal{D}_\Lambda^o$ which converge to a boundary point of $\mathcal{D}_\Lambda^o$,

then $(\mu_n)_{n \in \mathbb{N}}$ fulfills a LDP with speed $a_n$ and rate function $\Lambda^*$.

Concerning the proof of Theorem 1.9, the upper bound is again established by an application of an exponential Tschebyschev inequality. While for the lower bound, an exponential change of measure is carried out and since there is no law of large numbers in place a priori, the upper bound is used to derive the lower bound.

While a more abstract version of Theorem 1.9 exists, it is clear how to employ this theorem in a real setting: Ensure the existence of (1.11) first and then check whether the conditions gathered under (III) are fulfilled. That route will be taken on in the last chapter.
1 LARGE AND MODERATE DEVIATION PRINCIPLES
2 Random Matrix Theory

In random matrix theory, the classical situation is to consider sequences \((X_n)_{n \in \mathbb{N}}\) of matrix-valued random variables \(X_n\). Some recent work has been done concerning the distribution of an \(m\)-tuple of random matrices \((X_1, \ldots, X_m)\). This chapter deals with both situations.

2.1 One-matrix ensembles

Random matrices were first studied by Wishart [51] and Wigner [49, 50]. Wishart’s motivation was the multivariate analysis of the covariance matrix \(Y^{N,M} = X^{N,M}(X^{N,M})^*\), where \(X^{N,M}\) is a \(N \times M\) matrix with complex random entries, \(N\) and \(M\) are quite large and where \(X^*\) denotes the adjoint matrix of \(X\), \(X^* = X^t\). The matrix \(X^{N,M}\) consists of \(N\) independent vectors of observations \(X_1, \ldots, X_N, X_i \in \mathbb{C}^M\), and all vectors are identically distributed. The goal was to find a system, which encodes most of the variations of the data, but which is of much lesser dimension then the matrix \(Y^{N,M}\). Such a system can be found by performing a principal component analysis, what requires the study of the eigenvalues and eigenvectors of \(Y^{N,M}\). Thus, the spectrum of matrices of type \(X^{N,M}(X^{N,M})^*\) (which are called Wishart matrices) is of interest.

In quantum mechanics, the energy states of a system is represented by the eigenvalues of the Hamiltonian. When Wigner studied highly excited nuclei, exact calculations were impossible and he suggested a statistical approach by approximating the Hamiltonians by large random matrices. Hence, the spectrum of any infinite-dimensional operator could be modelled by the spectrum of a large (but finite) random matrix. The system itself imposes some symmetry conditions on the matrices, and the most random model fulfilling these condition(s) is chosen. Again, the eigenvalues of such matrices are of importance as is the spacing distribution, i.e. the distribution of the distances of neighboring eigenvalues.

In this section, we present eleven matrix ensembles. All these ensembles are studied in mesoscopic physics. The first ten ensembles describe disordered systems for fermions. These are the three classical ensembles, first studied by Wigner and Dyson, three chiral ensembles and four BdG-ensembles. The eleventh ensemble describes a model for bosons and was recently introduced in [37]. We will focus on this ensemble, since it will be studied in the next chapter.
2 RANDOM MATRIX THEORY

in even greater generality. As the density of the joint distribution of eigenvalues in the bosonic setting is a special case of the one for biorthogonal ensembles, which were defined in [6], we can obtain a result for all biorthogonal ensembles, which is analogous to the one obtained in [17] for the ten fermionic ensembles. The actual result and its proof will be the main content of the third chapter.

The literature, that has been the basis for this exposition, are the books by Mehta [40] and Forrester [21] and the papers [37], [6].

2.1.1 The classical Gaussian ensembles

In his threefold way [15], Dyson classified three classes of matrices (ensembles) due to the invariance of their laws under some specific group action.

Gaussian orthogonal ensemble (GOE)

**Definition 2.1** A matrix \(X^N = (X^N_{jk})_{1 \leq j,k \leq N}\) of dimension \(N \times N\) with real entries is said to belong to the GOE, if it is symmetric and if its entries are independent and normally distributed with zero mean. The variance of diagonal entries is \(\frac{2}{N}\), while it is \(\frac{1}{N}\) for all other entries.

This definition is not tied to any fixed \(N\), but the GOE is rather the collection \((X^N, N \in \mathbb{N})\).

From now on and for the rest of this chapter, we will drop the superscript and simply write \(X\) and \(X_{jk}\) instead of \(X^N\) and \(X^N_{jk}\) for this ensemble and all that are to follow.

It follows from Definition 2.1 that the density of the joint distribution of the matrix entries is

\[
P(X) = \prod_{j=1}^{N} \frac{N}{\sqrt{4\pi}} e^{-\frac{N}{4} X_{jj}} \prod_{1 \leq j < k \leq N} \frac{N}{\sqrt{2\pi}} e^{-\frac{N}{2} X_{jk}} = C_N \prod_{j,k=1}^{N} e^{-\frac{N}{4} X_{jk}},
\]

where all the non-exponential terms are absorbed in \(C_N\). Denoting the trace of a matrix \(X\) by \(\text{tr}(X)\), and observing that \(\text{tr}(X^2) = \sum_{j,k=1}^{N} X^2_{jk}\), we can write (2.1) as

\[
P(X) = C_N e^{-\frac{N}{4} \sum_{j,k=1}^{N} X^2_{jk}} = C_N e^{-\frac{N}{4} \text{tr}(X^2)}.
\]

From (2.2), we can see that the joint distribution of the matrix entries is invariant under the transformation

\[
X \rightarrow UXU^t,
\]

where \(U\) is an orthogonal matrix (i.e. \(U^t = U^{-1}\)), since \(\text{tr}(UXU^t) = \text{tr}(U^tUX) = \text{tr}(X)\). This is the reason for calling this ensemble Gaussian orthogonal ensemble.

As \(X\) is symmetric, we also know that it has a spectral decomposition

\[
X = \Lambda \Delta \Lambda^{-1},
\]
where $D$ is an orthogonal matrix and $\Lambda$ is a diagonal matrix containing all $N$ eigenvalues of $X$, $\lambda_1, \ldots, \lambda_N$, which provides that

$$\text{tr}(X^2) = \text{tr}(DD^{-1}D\Lambda^{-1}) = \text{tr}(\Lambda^2) = \sum_{j=1}^{N} \lambda_j^2.$$ 

One might suspect, that the information about the distribution of the matrix entries encodes some information about the eigenvalues. Indeed, it is possible to deduce from (2.1) the joint probability distribution of the eigenvalues. In order to do this, one must carry out a change of variable from the $N(N + 1)/2$ independent variables to the $N$ eigenvalues $\lambda_1, \ldots, \lambda_N$ and $N(N - 1)/2$ variables $y_1, \ldots, y_{N(N-1)/2}$, which afterwards must be integrated out. The main task in that calculation is the calculation of the Jacobian for the change of variables, and we will hint as in [30] at how it is proved, whereas the detailed calculations can be found in [21] or [40].

Let $dX = (dx_{jk})_{1 \leq j,k \leq N}$ denote the matrix of differentials of a symmetric matrix $X$. The following identities can be shown to hold, where the spectral decomposition (2.3) for the symmetric matrix $X$ is used, and the usual product rule for differentials, $d(XY) = dX \cdot Y + X \cdot dY$,

$$dX = d(DD^t) = dD \cdot \Lambda \cdot D^t + D \cdot d\Lambda \cdot D^t + D \cdot \Lambda \cdot dD^t$$

$$= D \cdot (D^t \cdot dD \cdot \Lambda + d\Lambda + \Lambda \cdot dD^t \cdot D) \cdot D^t$$

$$= D^t \cdot dD \cdot \Lambda + d\Lambda + \Lambda \cdot dD^t \cdot D$$

$$= d\Lambda + D^t \cdot dD \cdot \Lambda - \Lambda \cdot D^t \cdot dD.$$ 

Here, we used that $dX$ is invariant under orthogonal transformations and $dD^t \cdot D = -D^t \cdot dD$, because of $0 = d1_{N \times N} = d(D^t D) = dD^t \cdot D + D^t \cdot dD$. In particular, the difference $D^t \cdot dD \cdot \Lambda - \Lambda \cdot D^t \cdot dD$ is responsible for the presence of the factor $\prod_{1 \leq j < k \leq N}(\lambda_k - \lambda_j)$, which can be interpreted as level repulsion of the eigenvalues.

Thus, it turns out, that the joint probability distribution has the following density function,

$$q_N(\lambda_1, \ldots, \lambda_N) = \frac{1}{Z_{1,N}} \prod_{1 \leq j < k \leq N} |\lambda_k - \lambda_j| e^{-\frac{N}{4} \sum_{j=1}^{N} \lambda_j^2}, \quad (2.4)$$

where $Z_{1,N}$ is the partition function and $\prod_{1 \leq j < k \leq N} |\lambda_k - \lambda_j|$ is the Vandermonde determinant, which can also be written as

$$\Delta(\lambda) := \prod_{1 \leq j < k \leq N} |\lambda_k - \lambda_j| = \text{det}[A],$$

where $A$ is a matrix.
where
\[
A = \begin{pmatrix}
1 & \lambda_1 & \lambda_1^2 & \cdots & \lambda_1^{N-1} \\
1 & \lambda_2 & \lambda_2^2 & \cdots & \lambda_2^{N-1} \\
& \vdots & & & \\
1 & \lambda_N & \lambda_N^2 & \cdots & \lambda_N^{N-1}
\end{pmatrix}
\]

As stated above, in Quantum mechanics large but finite \(N \times N\) hermitian random matrices are used in order to model the discrete part of the spectrum of a quantum system. When imposing the condition of displaying some time-reversal and rotational symmetry on the system, it can be shown that the matrix describing the system can only stem from the GOE.

When the time-reversal symmetry is violated (usually due to the presence of magnetic interactions) we encounter the next classical ensemble.

**Gaussian unitary ensemble (GUE)**

**Definition 2.2** A matrix \(X = (X_{jk})_{1 \leq j,k \leq N}\) of dimension \(N \times N\) with complex entries is said to belong to the GUE, if it is hermitian \((X = X^*,\) where \(X^* = X^t\) is the adjoint matrix of \(X\)) and if the real and imaginary parts of its entries are independent and normally distributed with zero mean. The variance of the \(N\) diagonal entries, which are real due to the matrix being hermitian, is \(\frac{1}{N}\), while the real and imaginary part of the off-diagonal entries follow a centered normal distribution with variance \(\frac{1}{2N}\).

Thus, the joint probability distribution of the matrix entries is specified by the distribution of the entries of the upper-diagonal entries and its density is as follows,

\[
P(X) = \prod_{j=1}^{N} \sqrt{\frac{N}{2\pi}} e^{-\frac{N}{2} X_{jj}^2} \prod_{1 \leq j < k \leq N} \frac{N}{\pi} e^{-N(Re(X_{jk})^2 + Im(X_{jk})^2)}
\]

\[
= \prod_{j=1}^{N} \sqrt{\frac{N}{2\pi}} e^{-\frac{N}{2} X_{jj}^2} \prod_{1 \leq j < k \leq N} \frac{N}{\pi} e^{-N|X_{jk}|^2} = C_N \prod_{j,k=1}^{N} e^{-\frac{N}{2}|X_{jk}|^2},
\]

where \(C_N\) collects all the non-random factors. As \(X\) is hermitian, we find that

\[
\sum_{j,k=1}^{N} |X_{jk}|^2 = \text{tr}(XX^*) = \text{tr}(X^2),
\]

and therefore,

\[
P(X) = C_N e^{-\frac{N}{2} \sum_{j,k=1}^{N} |X_{jk}|^2} = C_N e^{-\frac{N}{2} \text{tr}(X^2)} \tag{2.5}
\]

Obviously, the distribution of the matrix entries is invariant under a unitary transformation \(X \rightarrow UXU^*\), where \(U\) is a unitary matrix (i.e. \(U^* = U^{-1}\)), which explains, why it is called
2.1 ONE-MATRIX ENSEMBLES

GUE.

As for the GOE, it is possible, see [21] or [40], to obtain from (2.5) the density of the joint probability distribution of the eigenvectors, which is

\[ q_N(\lambda_1, \ldots, \lambda_N) = \frac{1}{Z_{2,N}} \prod_{1 \leq j < k \leq N} |\lambda_k - \lambda_j|^2 e^{-\frac{N}{2} \sum_{j=1}^{N} \lambda_j^2} = \frac{1}{Z_{2,N}} \Delta(\lambda)^2 e^{-\frac{N}{2} \sum_{j=1}^{N} \lambda_j^2}. \]  

(2.6)

Gaussian symplectic ensemble (GSE)

We are left with the case of modelling a quantum system with time-reversal symmetry but without rotational symmetry.

The matrix fitting in this context consists of quaternions and from a ‘constructive’ point of view can be obtained as follows: Let \( N \) be even and consider the matrix \( X = (X_{jk})_{1 \leq j, k \leq \frac{N}{2}} \), with

\[ X_{jk} = \frac{\sum_{l=1}^{4} g_{jk}^l e^l}{\sqrt{4N}}, \quad 1 \leq j < k \leq N, \]

and

\[ X_{kk} = \sqrt{\frac{1}{2N}} g_{kk} e^1, \quad 1 \leq k \leq N, \]

where \((g_{jk}^l, j \leq k, l \in \{1, 2, 3, 4\})\) is a family of independent standard normal random variables.

Here, each \( X_{jk} \) is a \( 2 \times 2 \) matrix, since the appearing \((e^l)_{1 \leq l \leq 4}\) are the Pauli matrices

\[ e^1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad e^2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad e^3 = \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix}, \quad e^4 = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}. \]

Thus, each matrix entry can be written as \( q = c_0 e^1 + c_1 e^2 + c_2 e^3 + c_3 e^4 \) and the dual is defined as \( \overline{q} = c_0 e^1 - c_1 e^2 - c_2 e^3 - c_3 e^4 \), which provides \( |q|^2 := q\overline{q} = (c_0^2 + c_1^2 + c_2^2 + c_3^2)e^1 \).

The probability distribution of the matrix entries turns out to be

\[ P(X) = \prod_{j=1}^{N} \sqrt{\frac{N}{\pi}} e^{-N X_{jj}^2} \prod_{1 \leq j < k \leq N} \frac{4N^2}{\pi^2} e^{-2N|X_{jk}|^2} = C_N e^{-N \text{tr} X^2}. \]

By a change of variables, the joint probability distribution of the eigenvalues can be computed, see [21] or [40], which has the following density,

\[ q_N(\lambda_1, \ldots, \lambda_N) = \frac{1}{Z_{4,N}} \Delta(\lambda)^4 e^{-N \sum_{j=1}^{N} \lambda_j^2}. \]  

(2.7)

Comparing (2.4), (2.6) and (2.7), we see that the joint eigenvalue distributions can be brought together under the following expression

\[ Q_{\beta,N}(d\lambda_1, \ldots, d\lambda_N) = \frac{1}{Z_{\beta,N}} \Delta(\lambda)^\beta e^{-\frac{4}{\beta} N \sum_{j=1}^{N} \lambda_j^2} \prod_{j=1}^{N} d\lambda_j, \]  

(2.8)
where $\beta \in \{1, 2, 4\}$ yields the corresponding ensembles ($\beta = 1$: GOE, $\beta = 2$: GUE and $\beta = 4$: GSE). Moreover, the nomenclature of the ensemble is according to the group, which leaves the distribution of the matrix entries invariant under its action. These are the classical ensembles, as presented by Dyson in [15]. These ensembles were introduced by specifying the distribution of each matrix entry and demanding some kind of matrix structure. From now on, classes of matrices obtained by such an approach will be called Wigner ensembles. As we have seen, the three ensembles possess some invariance regarding some specific group action and it can be shown, that by demanding such an invariance (without specifying the distribution of the matrix entries), only the three ensembles would show up. Hence, postulating some kind of invariance may also be an approach to specify matrix ensembles and such ensembles will be called invariant ensembles. But our approach will always be to take a Wigner matrices with Gaussian entries. For this reason, the following treatment neglects group actions and rather concentrates on the specific structure, that a matrix displays in each ensemble.

### 2.1.2 Chiral ensembles

The next ensembles evolved from the study of random Dirac operators. The appearing Hamiltonian in this context can be described by a $(N + M) \times (N + M)$ matrix $H$, with $N, M \in \mathbb{N}$, which has a block structure and is made up by a $N \times M$ matrix $X$, 

\[
H = \begin{pmatrix}
0_{N \times N} & X \\
X^* & 0_{M \times M}
\end{pmatrix},
\]

where the matrix entries could as above be real, complex or real quaternion. The corresponding three ensembles are called chiral ensembles. The matrix $H$ has $N \wedge M := \min\{N, M\}$ eigenvalues and the eigenvalues appear in pairs $\pm \lambda_j$, since if

\[
\begin{pmatrix}
0 & X \\
X^* & 0
\end{pmatrix}
\begin{pmatrix}
a \\
b
\end{pmatrix} = \lambda
\begin{pmatrix}
a \\
b
\end{pmatrix}
\]

is fulfilled, then it also holds that

\[
\begin{pmatrix}
0 & X \\
X^* & 0
\end{pmatrix}
\begin{pmatrix}
a \\
-b
\end{pmatrix} = -\lambda
\begin{pmatrix}
a \\
-b
\end{pmatrix}.
\]

As in the classical setting, the matrix entries of $X$ are now chosen to be i.i.d. normal random variables with variance as in the three cases above (e.g. in the real setting $\frac{1}{N}$ for off-diagonal
entries and $\frac{2}{\beta}$ for diagonal ones), which provides that the joint distribution of the eigenvalues for these ensembles can be obtained and it is
\[
Q_{\beta,N,M}(d\lambda_1, \ldots, d\lambda_{N\times M}) = \frac{1}{Z_{\beta,N,M}} \prod_{1 \leq j < k \leq N\times M} |\lambda_k^2 - \lambda_j^2|^{\beta} \prod_{j=1}^{N\times M} \lambda_j^{\beta|N-M|+\beta-1} e^{-\frac{\beta}{4} \lambda_j^2} \prod_{j=1}^{N\times M} d\lambda_j
\]
\[
= \frac{1}{Z_{\beta,N,M}} \Delta(\lambda^2)^\beta \prod_{j=1}^{N\times M} \lambda_j^{\beta|N-M|+\beta-1} e^{-\frac{\beta}{4} \lambda_j^2} \prod_{j=1}^{N\times M} d\lambda_j, \quad (2.9)
\]
where $Z_{\beta,N,M}$ is the partition function. Detailed calculations (which is in essence a change of variables similar to the one for the classical ensembles) leading to (2.9) can be found in [21]. The parameter $\beta$ gives information about the matrix entries, by being one for real entries, two for complex entries and four for quaternions.

2.1.3 BdG-ensembles

When considering disordered superconductors, the Hamiltonian of the system can be described by a block matrix of type
\[
H = \begin{pmatrix}
X_1 & X_2 \\
X_2^T & -X_1^T
\end{pmatrix}, \quad (2.10)
\]
where $X_i, i = 1, 2,$ are $N \times N$ matrices and it must always hold that $X_1 = X_1^*$. The matrix entries are either real, complex or real quaternion and additionally, the matrix $X_2$ may be symmetric or skew-symmetric. The resulting ensembles are called Bogoliubov-de Gennes (BdG) ensembles, since the phenomenon of superconductance can be expressed via a Bogoliubov-de Gennes equation.

Sometimes, they are also referred to as superconductor ensembles and the following four cases occur:

(1) $\beta = 1, X_2^* = X_2$:

As the entries are real, we find $X_1 = X_1^* = X_1^T$ and
\[
H = \begin{pmatrix}
X_1 & X_2 \\
X_2 & -X_1
\end{pmatrix}, \quad (2.11)
\]

(2) $\beta = 2, X_2^* = X_2$:

Now, the entries are complex and we have $X_1 = X_1^*, X_2^* = \overline{X_2} = X_2$ resulting in
\[
H = \begin{pmatrix}
X_1 & X_2 \\
\overline{X_2} & -\overline{X_1}
\end{pmatrix}. \quad (2.12)
\]
(3) $\beta = 2$, $X_2^t = -X_2$: When $X_2$ is skew-symmetric and $H$ has complex entries, it turns out that

$$H = \begin{pmatrix} X_1 & X_2 \\ -\bar{X}_2 & -\bar{X}_1 \end{pmatrix}.$$  \hfill (2.13)

This matrix ensemble is often represented as a skew-hermitian matrix with pure complex entries (i.e. has no real part). This can be seen by undertaking the transformation $H \rightarrow \tilde{H} := XHX^1$, where

$$U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1_{N \times N} & 1_{N \times N} \\ i1_{N \times N} & -i1_{N \times N} \end{pmatrix} \text{ and therefore } U^{-1} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1_{N \times N} & -i1_{N \times N} \\ 1_{N \times N} & i1_{N \times N} \end{pmatrix}.$$  \hfill (2.14)

Because we find (after some matrix calculus), $\tilde{H} = -\tilde{H} = \tilde{H}^t$, this ensemble has the mentioned structure.

(4) $\beta = 4$, $X_2^t = -X_2$: With real quaternion matrix entries, $X_1 = X_1^*$ and skew-symmetric $X_2$, it can be shown that

$$H = \begin{pmatrix} X_1 & X_2 \\ X_2^* & -X_1^t \end{pmatrix}$$

can as well be represented by

$$H = \begin{pmatrix} X_1 & X_2 \\ X_2 & -X_1 \end{pmatrix},$$  \hfill (2.15)

where $X_i, i = 1, 2$ have purely complex entries and are skew symmetric. We won’t show it here as it is not enlightening for our purposes (see [21, Chap. 2.7.2] for details). It can be shown similar to the chiral ensemble, that in any of the four cases, the eigenvalues appear in pairs $\pm \lambda_j$ and when the entries of each block matrix $X_i, i = 1, 2$, are chosen to be i.i.d. normal random variables the density of the joint distribution of the eigenvalues is given by

$$Q_{\beta,p(N)}(d\lambda_1, \ldots, d\lambda_{p(N)}) = \frac{1}{Z_{\beta,p(N)}} \Delta(\lambda^2)^{\beta} \prod_{j=1}^{p(N)} \lambda_j^2 e^{-\frac{N}{\beta} \lambda_j^2} \prod_{j=1}^{p(N)} d\lambda_j,$$  \hfill (2.15)

and again, the parameter $\beta$ carries information about the matrix entries (e.g. $\beta = 1$ for real entries). The parameter $\alpha$ equals 1 in case (1), in case (2) we find $\alpha = 2$, case (3) corresponds to $\alpha = 0$, while in case (4) a distinction is made between even and uneven $N$: if $N$ is even, we have $\alpha = 1$, whereas it is 5 for uneven $N$. Moreover, $p(N) = N$ for cases (1)-(3) and $p(N)$ equals $\lfloor \frac{N}{2} \rfloor$ in case (4).
The matrix ensembles considered so far (GOE, GUE, GSE, chiral ensembles and BdG ensembles) play an important role in mesoscopic physics, where they model disordered systems of fermions, and they were classified in [29]. The densities of the joint eigenvalue distributions can be subsumed under the following general form, whenever the matrix entries are i.i.d. normally distributed,

\[ q_n(x_1, \ldots, x_{p(n)}) = \frac{1}{Z_n} \prod_{1 \leq j < k \leq p(n)} |x_k^\gamma - x_j^\gamma|^{\beta} \prod_{j=1}^{p(n)} w_n(x_j)^n 1_{\Sigma^p}(x_1, \ldots, x_{p(n)}), \]

where \( Z_n \) bestows the normalization upon the distribution, \( \gamma, \beta, p(n) \in \mathbb{N}, \Sigma \subset \mathbb{R} \) and the weight functions \( w_n \) include all other appearing terms.

Regarding the three classical ensembles, many results have been obtained by now, e.g. Wigner’s semicircle law for the empirical distribution [49, 50], the spacing distribution and correlation functions of the eigenvalues, where we refer to [40] for a presentation of these results and for further references to the original papers, and a LDP for the largest eigenvalue [3]. But some results hold for all ten fermionic ensembles, such as a LDP for the empirical measures of eigenvalues [17]. And only recently, analogs to Wigner’s semi-circle law for the eigenvalue distribution have been proven for the other seven fermionic ensembles in [32], where also universality (in this context, universality means, that the result holds for any distribution of the matrix entries) under some moment conditions could be established.

In contrast to fermions, disordered systems of bosons are less studied. As far as we know, there is up to now only one physical model which describes a disordered system of bosons, see [37]. In that paper, the authors deduced the distribution of eigenvalues for such a system and it is a special case of the one given in [6]. Besides that, none of the results mentioned for the fermionic ensembles have been obtained in the bosonic setting.

### 2.1.4 A system of disordered bosons

The random matrix model for a system of disordered bosons was introduced and examined in [37]. The authors argue that this model is most closely analogous to the classical Gaussian ensembles in the fermionic world although the matrix entries of the Hamiltonian belonging to a bosonic system are not chosen independently and identically distributed, as is often the case when looking at fermions.

These systems of disordered bosons appear, when modelling vibrational modes of a solid, spin waves in a magnet, electromagnetic modes in an optical medium or oscillations of the superfluid density of a Bose-Einstein condensate.
Within this framework, the matrix
\[ X = \begin{pmatrix} A & -B \\ C & -A^t \end{pmatrix}, \quad (2.16) \]
is considered, where \( A, B, C \in \mathbb{R}^{N \times N} \) with \( B = B^t, C = C^t \). In general, \( X \) may not be diagonalizable (e.g. take \( A = B = 0 \) and \( C = 1_{N \times N} \)), but even if it is, the eigenvalues of \( X \) might be complex. For this reason, a symmetric stability matrix \( h \in \mathbb{R}^{2N \times 2N} \) is introduced and the stability criterion is that \( h \) is positive definite, where \( h \) is extracted from \( X \) as follows:
\[ h = \begin{pmatrix} B & A \\ A^t & C \end{pmatrix}. \quad (2.17) \]
It is obvious that \( h \) is real symmetric. The requirement that all eigenvalues of \( h \) are positive, provides that the eigenvalues of \( X \) come as imaginary pairs \( \pm i\lambda_j \), where \( \lambda_j > 0, \ j = 1, \ldots, N \). These are called simple-boson energies of the small-amplitude motion.

Hence, the constructing principle behind this setting is not any symmetry, but a stability condition.

The stability matrix is a positive definite real symmetric matrix \( h \) and such matrices can naturally be constructed by adding a sufficient number of rank-one projectors with positive weights. This is equivalent to put
\[ h_{ij} = \sum_{k=1}^{M} v_{ik} v_{jk}, \ M \in \mathbb{N}, \ i, j = 1, \ldots, 2N, \quad (2.18) \]
where the real numbers \( v_{ik} \) can be chosen to be independent and normally distributed with mean 0 and variance \( \tau^{-1} \). This choice enabled the calculation of a probability distribution \( d\mu(h) \) for \( h \), which is
\[ d\mu(h) = \frac{1}{Z_n} e^{-\frac{1}{2}\tau \text{Tr}(h)} \det(h)^{\frac{1}{2}(l-1)} \prod_{i \leq j} dh_{ij}, \text{ where } l = M - 2N \geq 0, \]
for all positive matrices \( h \). In [37] \( d\mu(h) \) was reduced to a probability distribution on the space of characteristic frequencies \( x_1, \ldots, x_n \) (i.e. the positive eigenvalues of \( iX \)) to be
\[ d\mu_{n,\alpha}(x_1, \ldots, x_n) = \frac{1}{Z_n} \prod_{k=1}^{n} x_k^{\alpha} e^{-\tau x_k} \prod_{1 \leq i < j \leq n} |x_i - x_j| |x_i^2 - x_j^2| 1_{\Sigma_n}(x_1, \ldots, x_n) dx_k. \quad (2.19) \]
As we have seen, there is a nice physical motivation to study ensembles, for which the density of its eigenvalues distribution is of type (2.19). In [6], Borodin introduced an eigenvalue distribution with the following density,
\[ q_n(x_1, \ldots, x_n) = \frac{1}{Z_n} \prod_{i=1}^{n} w(x_i) \prod_{1 \leq i < j \leq n} |x_i - x_j| |x_i^\theta - x_j^\theta|, \quad (2.20) \]
where \( \theta \in \mathbb{N} \). Obviously, (2.20) includes (2.19) as a special case. The more general version (2.20) will be the starting point for our analysis in chapter three.

### 2.2 A Multi-Matrix model

In this section, we look at a two-fold extension of a classical ensemble. First, we deal with \( m \) random matrices at once, where \( m \in \mathbb{N} \) is fixed.

The second extension is, that we look at hermitian matrices whose distribution is given by a small convex perturbation of the GUE. We know from (2.5), that the distribution of a matrix \( H \) belonging to the GUE is of form

\[
C_N e^{-\frac{N}{2} \text{tr}(H^2)},
\]

and we will now slightly deviate from \( H^2 \) by allowing for more general polynomials.

The real part of a complex number \( z \) will be denoted by \( \Re(z) \) and its imaginary part by \( \Im(z) \). Let \( \mathcal{H}_N(\mathbb{C})^m \) be the set of \( m \)-tuples \( H = (H_1, \ldots, H_m) \), \( H_i = (H_i(kl))_{1 \leq k, l \leq N} \), of \( N \times N \) hermitian matrices, such that \( \Re(H_i(kl)), k < l \), \( \Im(H_i(kl)), k < l \), \( 2^{-\frac{1}{2}} H_i(kk) \) is a family of independent real Gaussian variables of variance \( (2N)^{-1} \). We will consider the following perturbation of the GUE:

\[
\mu_N^V(dH) = \frac{1}{Z_V^N} \exp \left\{ -N \text{tr} \left( V(H_1, \ldots, H_m) \right) \right\} \mu_N^N(dH),
\]

where \( Z_V^N \) is the normalizing constant,

\[
V(H_1, \ldots, H_m) := V_t(H_1, \ldots, H_m) = \sum_{i=1}^n t_i q_i(H_1, \ldots, H_m)
\]

a polynomial potential with \( n \in \mathbb{N} \), \( t = (t_1, \ldots, t_n) \in \mathbb{C}^n \) and monomials \( (q_j)_{1 \leq j \leq n} \) fixed, and

\[
\mu_N^N(dH) = \frac{1}{Z_N^N} \exp \left\{ -\frac{N}{2} \text{tr} \left( \sum_{i=1}^m H_i^2 \right) \right\} d^N H,
\]

the law of the \( m \)-dimensional GUE with \( d^N H \) the Lebesgue measure on \( \mathcal{H}_N(\mathbb{C})^m \).

For \( m = 1 \) and \( V(H) = 0 \) we recover the classical GUE. But for one-matrix models different generalizations to the classical situation were studied. In [10], the potential \( V \) was still zero, but it was allowed to add a deterministic diagonal matrix \( D \), to the hermitian matrix \( H \). The only condition on \( D \) was, that its spectral measure converges to a compactly supported probability measure. In this setting, a MDP for the difference of the cumulative distribution function (cdf) of the semicircle distribution and the cdf of the empirical measure was proved.
in [10].

Another direction, in which the GUE was generalized, consists of allowing for non-zero potentials \( V(H) \) in (2.21), while still keeping \( m = 1 \). The LDP for the empirical measure of the eigenvalues was obtained [4], where \( V \) had to fulfill either a boundness criterion or had to exhibit some special behavior when the argument goes to infinity. For general polynomial potentials \( V \), Johansson [33] proved a CLT for \( \sum_{i=1}^{N} f(\lambda_i) \), where \( \lambda_i \) are the eigenvalues of the hermitian matrix \( H \) and \( f \) is a special type of function, and he also gives a complete characterization of those functions \( f \), for which the CLT holds.

One-matrix models with non-vanishing \( V \) have been under investigation intensively in physics and with the choice \( V(H) = t H^4, t \in \mathbb{C} \), (2.21) was a commonly studied object, see [5, 7].

What was really striking, is the connection their analysis established between matrix integrals like \( Z_N^N \) and map enumeration. Let us now briefly hint at how matrix integrals and map enumeration are connected, while we refer to the excursus given in the fourth chapter for a thoroughful treatment of that topic. For the time being, we understand that graphs consist of vertices and edges which connect two vertices. When the graph is connected we can dissect it along the edges. Provided that this results in sets that are homeomorphic to discs, we call these sets faces and the connected graph a map. When dealing with normally distributed random variables \( X_1, \ldots, X_{2n} \), the Wick formula tells us how to compute the expectation of the product \( X_1 \cdots X_{2n} \) in terms of summing products of expectations of two factors only, namely

\[
E[X_1 \cdots X_{2n}] = \sum_{1 \leq s_1 < s_2 \cdots < s_n \leq 2n} \prod_{j=1}^{n} E[X_{s_j}X_{r_j}].
\]

As the matrix entries in the GUE case follow a normal distribution, Wicks formula can be applied to calculate the expectation of traces of moments of matrices and denoting the measure (2.5) of the GUE by \( \mu^N(dH) \), we obtain

\[
\int \prod_{i=1}^{k} (\text{Ntr}(H^{p_i}))) \mu^N(dH) = \sum_{F \geq 0} N^{F+k-\frac{\sum p_i}{2}} G((p_i)_{1 \leq i \leq k}, F),
\]

(2.22)

with

\[
G((p_i)_{1 \leq i \leq k}, F) = \# \{ \text{oriented maps with } F \text{ faces and 1 vertex of degree } p_i, 1 \leq i \leq k \}.
\]

The number \( F + k - \frac{\sum p_i}{2} \) corresponds to \( 2 - 2g \), where \( g \) denotes the genus of the surface (‘number of holes’), on which a connected oriented graph with \( F \) faces and one vertex of
degree $p_i$ for $1 \leq i \leq k$ can be embedded. Thus,

$$\sum_{F \geq 0} N^{F+k-\sum p_i} G((p_i)_{1 \leq i \leq k}, F) = \sum_{g=0}^{\infty} N^{2-2g} G((p_i)_{1 \leq i \leq k}, F). \quad (2.23)$$

Regarding the matrix size $N$ as a parameter in (2.23), the expectation of the trace of moments of matrices can be interpreted as a generating function for the number of oriented graphs with a given genus and some vertices with given degree. Moving to Laplace transforms of traces of matrices belonging to the GUE should therefore result in the appearance of generating function for maps. Indeed, it was shown in [20], that for parameters $t = (t_1, \ldots, t_k)$ in the vicinity of the origin, the following holds,

$$\log Z_{V_{t}}^{N} = \log \int \exp \left\{ -\sum_{i=1}^{k} N t_i \text{tr}(H^{p_i}) \right\} \mu^{N}(dH) = \sum_{g=0}^{\infty} N^{2-2g} F_{g}(t), \quad (2.24)$$

where

$$F_{g}(t) := \sum_{(n_1, \ldots, n_k) \in \mathbb{N}^k} \frac{(-t_i)^{k_i}}{k_i!} M((p_i, n_i)_{1 \leq i \leq k}, g).$$

Hereby, $M((p_i, n_i)_{1 \leq i \leq k}, g)$ is the number of connected maps with genus $g$ and $n_i$ vertices of degree $p_i$, $i = 1, \ldots, k$. The logarithm is responsible for counting connected maps instead of unconnected graphs.

While (2.24) was ‘believed’ in physics after the papers [5, 7] gave much evidence for the expansion to hold, it was only in [20], when it was placed on solid mathematical grounds by Ercolani and McLaughlin. They proved their results via a Riemann-Hilbert approach. Notice, that the admissible region of parameters $t \in \mathbb{R}^k$ of the polynomial $V$ for which they obtained the genus expansion encompasses e.g. any polynomial $V$ of even degree with positive leading coefficient. In this case, it is a basic fact (see [5] and [28]) that the measure (2.21) induces a probability measure on the eigenvalues $\lambda = (\lambda_1, \ldots, \lambda_N)$ with density

$$\frac{1}{Z_{V}^{N}} \exp \left( -N^2 \left( \frac{1}{N} \sum_{j=1}^{N} V(\lambda_j) + \frac{1}{N^2} \sum_{j \neq l} \log |\lambda_j - \lambda_l| \right) \right) d^N \lambda,$$

where $Z_{V}^{N}$ is precisely the partition function expanded in [20, Thm. 1.1], and $V(x) = \frac{1}{2} x^2 + \sum_{k=1}^{\nu} t_k x^k$ with suitable chosen parameters $(t_1, \ldots, t_{\nu})$.

Since many interesting models like the Ising model on random graphs, the q-Potts model on random graphs and a chain model are of form (2.21) and have $m > 1$, the question arose, whether a genus expansion for the multi-matrix model with general polynomial potentials $V$
could be obtained. It turned out, that (2.22) could be generalized to multi-matrix models. Suppose that each edge of a map is colored, where each $H_i$ corresponds to a certain color. Furthermore, we introduce an orientation and for each vertex, we distinguish one outgoing edge. Given a monomial $q_i(H_1, \ldots, H_m) = H_{i_1} \cdots H_{i_k}$, we say that a vertex is of type $q_i$, provided that the first outgoing edge is of color corresponding to $H_{i_1}$ and the $l$-th outgoing edge following the imposed orientation is of color $H_{i_l}$, for $l = 2, \ldots, k$. Graphs are constructed from these vertices by gluing edges that are of the same color. Now, we can state the asymptotic expansion for moments of traces of non-commutative monomials $q_i(H_1, \ldots, H_m)$, which is as follows

$$
\int \prod_{i=1}^{k} (N \text{tr}(q_i(H_1, \ldots, H_m))) \prod_{i=1}^{m} \mu^N(d H_i) = \sum_{F \geq 0} N^{F+k-\sum_{i=1}^{k} \beta_i} G_c((q_i)_{1 \leq i \leq k}, F),
$$
(2.25)

where

$$
G_c((q_i)_{1 \leq i \leq k}, F) = \#\{\text{oriented maps with } F \text{ faces and one star of type } q_i\}.
$$

As in the one-matrix model, moving from moments to Laplace transforms, it was also found that (2.24) could be generalized to the the multi-matrix setting, resulting in

$$
\log Z_N(t) = \log \int \exp \left\{- \sum_{i=1}^{k} N t_i \text{tr}(q_i(H_1, \ldots, H_m)) \right\} \prod_{i=1}^{m} \mu^N(d H_i) = \sum_{g=0}^{\infty} N^{2-2g} \tilde{F}_g(t),
$$
(2.26)

with

$$
\tilde{F}_g(t) := \sum_{(n_1, \ldots, n_k) \in \mathbb{N}^k} \prod_{i=1}^{k} \frac{(-t_i)^{n_i}}{n_i!} \tilde{M}((q_i, n_i)_{1 \leq i \leq k}, g).
$$

Hereby, $\tilde{M}((q_i, n_i)_{1 \leq i \leq k}, g)$ is the number of connected maps with genus $g$ and $n_i$ vertices of type $q_i$, $i = 1, \ldots, k$. The first step towards (2.26) was a first order expansion ($g=0$) of the partition function obtained by Guionnet in [22], a paper that was based on [26]. It turned out that $\lim_{N \to \infty} \frac{1}{N^2} \log Z_N^V$ could be given given in terms of a variational problem. For the one-matrix model, a corresponding result had already been obtained in [4]. Guionnet and Maurel-Segala refined the expansion in (2.26) up to second order ($g=1$) in [25], whereas the route taken there was more along the lines of [24]. Finally, Maurel-Segala [39] gave the full genus expansion of (2.25) and (2.26) for the multi-matrix model of form (2.21). In [25], the authors also derived a CLT for $N$ times the difference of the empirical measure and its limit, the solution of a Schwinger-Dyson equation, and showed that the CLT also holds when the expectation of the empirical measure is used as a centering. Notice, that Cabanal-Duvillard [8] introduced a stochastic calculus approach and proved a CLT for traces of non-commutative polynomials of
2.2 A MULTI-MATRIX MODEL

Gaussian Wigner and Wishart matrices, as well as for traces of non-commutative polynomials of pairs \((m = 2)\) of independent Gaussian Wigner matrices.

In chapter four, we will contribute to the analysis of this multi-matrix model by deducing a MDP for the normalized trace of a non-commutative polynomial under \(\mu^N_{V}\) as in (2.21) and present an alternative proof of the CLT derived in [25].
2 RANDOM MATRIX THEORY
3 Large Deviation Principle for biorthogonal ensembles

In this chapter, we will bring together random matrix theory and the techniques of large deviations. It might be possible to derive a LDP for the empirical measure of eigenvalues of some matrix ensembles, if the density of the joint distribution of the eigenvalues is known. In the previous chapter, we have seen that for Wigner ensembles, that density of the eigenvalue distribution can be given explicitly, provided that the matrix entries follow a normal distribution.

For the three classical ensembles GOE, GUE and GSE as classified by Dyson in [15], such a LDP was obtained in [4]. But also, in the very same paper the authors generalized their result to also include other weight functions than the classical $\exp\{-\frac{N}{2}x^2\}$. Weight functions $\exp\{-\frac{N}{2}V(x)\}$ were allowed, provided that either there exists a $a \in \mathbb{R}$, such that $V(x) - ax^2$ is a bounded continuous function on $\mathbb{R}$, or $V(x)$ is a continuously differentiable function with some specific asymptotic behavior when $x$ goes to infinity.

Because Dyson’s threefold way was falling short of describing all relevant matrix ensembles in mesoscopic physics, it was superseded by the tenfold way of Heinzner, Huckleberry and Zirnbauer in [29]. The authors established a one-to-one correspondence between ten matrix ensembles (which are the three classical, BdG and chiral ensembles) to Riemannian symmetric spaces as classified by Cartan. Based on their work, Eichelsbacher and Stolz derived in [17] a closed form for the joint eigenvalue distribution of any of the ten ensembles, and this enabled them to also establish a LDP for the empirical measure of the eigenvalues for the ten ensembles in the same paper. Their treatment is that general, that it also incorporates the cases of Wishart matrices or Jacobi ensembles, where the comparable LDP was derived in [31]. In physics, the ten matrix classes model the behavior of fermions.

The starting point of our task to obtain a LDP for the empirical measure of eigenvalues for biorthogonal ensembles is the density of the joint eigenvalue distribution as stated in [6], of which the one given in [37] is a special case.
We will start by explaining why the ensembles are called biorthogonal. Up to now, all that is known concerning these ensembles is the limit of the correlation function, [6], and the limit of the mean spectral distribution, [37]. We will answer the open question of deriving a LDP for the empirical measure of eigenvalues of matrices from biorthogonal ensembles. Moreover, in special situations a strong law of large numbers for the empirical measure of the eigenvalues of such matrix ensembles can be obtained from this LDP. The main challenge when proving the LDP will be to control the many appearing singularities, which require special treatment. In addition, we broaden the scope for the considered weight functions \( w \), to allow for \( n \)-dependent weight functions \( w_n \).

The chapter will finish by spelling out the LDP for three classes mentioned in [6], in which the weight functions can be regarded as the biorthogonal versions of the weight functions of Jacobi, Laguerre and Hermite polynomials.

### 3.1 Biorthogonal ensembles

As we have seen in the second chapter, there are matrix ensembles, whose eigenvalues have a joint distribution with a density of the form

\[
q(x_1, \ldots, x_n) = \frac{1}{Z_n} \prod_{i=1}^{n} w(x_i) \prod_{i<j} (x_i - x_j)^2,
\]

where \( Z_n \) is the normalizing constant (also called partition function).

Matrix ensembles, which lead to an eigenvalue distribution with density (3.1) can be connected to classes of orthogonal polynomials by choosing the corresponding weight function \( w \); e.g. choosing the weight function according to the Hermite polynomials yields the distribution as in the GUE case.

We remember, that an orthogonal system on \( L^2 \), the space of square Lebesgue-integrable functions, is given by a positive weight function \( w(x) \) and corresponding orthogonal polynomials \( (p_i)_{i \in \mathbb{N}} \) defined on a domain \( \Sigma \), where \( p_i \) is a polynomial of degree \( i \) such that the following holds,

\[
\int_{\Sigma} p_n(x) p_m(x) w(x) dx = \delta_{nm}.
\]

When Muttalib [41] considered models described by matrix ensembles of the above type and in addition, introduced a special two-body interaction, he could obtain the density of the joint eigenvalue distribution of the resulting matrix ensembles, which is a generalization of (3.1),

\[
q_n(x_1, \ldots, x_n) = \frac{1}{Z_n} \prod_{i=1}^{n} w(x_i) \prod_{1 \leq i < j \leq n} |x_i - x_j| |x_i^\theta - x_j^\theta|,
\]

(3.2)
3.2 MODEL ASSUMPTIONS AND NOTATIONS

where $\theta \in \mathbb{N}$ and $Z_n$ the partition function.

From (3.2) it is clear that one additional parameter $\theta$ enters the stage, while for $\theta = 1$, we recover the classical situation (3.1). The analysis of these ensembles helps to understand disordered systems and such ensembles can be related to biorthogonal polynomials.

Biorthogonal polynomials are two distinct families of polynomials $(p_i(x))_{i \in \mathbb{N}}, (\tilde{p}_i(x))_{i \in \mathbb{N}}$, where the coefficients of each polynomial are such that they satisfy

$$\int_{\Sigma} p_n(x)\tilde{p}_m(x)w(x)dx = g_n \delta_{nm},$$

where a proper scaling achieves $g_n = 1$ for all $n \in \mathbb{N}$. Biorthogonal polynomials were first studied in [35] and [36]. Matrix ensembles with a joint density of the distribution of eigenvalues as in (3.2) were studied in [6] and since the calculations relied on biorthogonal polynomials, these matrix ensembles are called biorthogonal ensembles.

We will generalize (3.2) even further in two different ways. We not only look at $n$ eigenvalues, but at $p(n)$, where the relation between $n$ and $p(n)$ will be specified later on. Furthermore, we allow for a $n$-dependence in our weight function, using $w_n$ instead of $w$, as such $n$-dependent weight functions appear when looking at some concrete ensembles, see e.g. the examples in the last part of this chapter. For these reasons, the density $q_n$ of the joint eigenvalue distribution of matrices coming from biorthogonal ensembles is more general than the one stated in the second chapter (2.20) and it is as follows:

$$q_n(x_1, \ldots, x_{p(n)}) = \frac{1}{Z_n} \prod_{i=1}^{p(n)} w_n(x_i)^n \prod_{1 \leq i < j \leq p(n)} |x_i - x_j|^{\theta} |x_i^\theta - x_j^\theta| 1_{\Sigma^{p(n)}}(x_1, \ldots, x_{p(n)}),$$  

(3.3)

where $\theta \in \mathbb{N}, p(n) \in \mathbb{N}, w_n : \mathbb{R} \to \mathbb{R}^+_0$ are continuous weight functions and $\Sigma$ is a a closed subset of $\mathbb{R}$, on which these ensembles are defined. Note that because of the following treatment, the weight function appears as $w_n^n$ in (3.3).

3.2 Model assumptions and notations

Now, we introduce some notation that will be used throughout the whole chapter and we specify the assumptions, made for the model of biorthogonal ensembles, we are going to consider.

Let $X_n = (X_1^n, \ldots, X_{p(n)}^n)$ be a set of random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The joint distribution $Q_n$ with density $q_n$ of type (3.3) for $X_n$ is an element of $\mathcal{M}_1(\Sigma^{p(n)})$ and
will be denoted by \( Q_n = \mathbb{P} \circ X_n^{-1} \).

We assume that \( (p(n))_{n \in \mathbb{N}} \) is such that
\[
\lim_{n \to \infty} \frac{p(n)}{n} = \kappa, \quad \text{with} \quad \kappa \in (0, \infty).
\]

The parameter \( \theta \) is an integer and if \( \theta \) is odd, \( \Sigma \) is a closed subinterval of \( \mathbb{R} \), whereas for \( \theta \) even, \( \Sigma \) is a closed subinterval of \([0, \infty)\).

For a function \( f : \mathbb{R} \to \mathbb{R} \) let \( \mathcal{N}(f) \) be the set of zeros of \( f \),
\[
\mathcal{N}(f) = \{ x \in \mathbb{R} | f(x) = 0 \}.
\]

For the sequence of continuous weight functions \( (w_n)_{n} \) we demand that it satisfies the following assumptions:

(a1) there exists a continuous function \( w : \Sigma \to [0, \infty) \) such that
- \( \# \mathcal{N}(w) < \infty \), \( \mathcal{N}(w_n) \subseteq \mathcal{N}(w) \) for large \( n \). (a1.1)
- As \( n \to \infty \), \( w_n \) converges to \( w \), and \( \log w_n \to \log w \) uniformly on compact sets. (a1.2)
- \( \log w \) is Lipschitz on compact sets away from \( \mathcal{N}(w) \). (a1.3)

(a2) If \( \Sigma \) is unbounded, then there exists \( n_0 \in \mathbb{N} \) such that
\[
\lim_{x \to \pm \infty} |x|^{(\theta+1)(\kappa+\epsilon)} \sup_{n \geq n_0} w_n(x) = 0
\]
for some fixed \( \epsilon > 0 \).

A central object, that is going to be examined is the empirical distribution \( L_n(x) \) of \( x \) for \( x = (x_1, \ldots, x_{p(n)}) \) \( \in \Sigma^{p(n)} \), which is defined as
\[
L_n(x) := \frac{1}{p(n)} \sum_{j=1}^{p(n)} \delta_{x_j},
\]

The space of probability measures on the Borel sets of \( \Sigma \) is denoted by \( \mathcal{M}_1(\Sigma) \) and in the following it is always endowed with the weak topology.

### 3.3 Main Theorem

Now we can state the main result of this chapter, a large deviation principle on the space of probability measures for the empirical distribution of the eigenvalues of matrices from biorthogonal ensembles.
3.3 MAIN THEOREM

Theorem 3.1 The sequence \((P \circ (L_n \circ X_n)^{-1})_n = (Q_n \circ L_n^{-1})_n\) satisfies a LDP on \(M_1(\Sigma)\) with respect to the weak topology with speed \(n^2\) and good rate function

\[
I(\mu) = \frac{\kappa^2}{2} \int \int \left\{ \log |x^\theta - y^\theta|^{-1} + \log |x - y|^{-1} \right\} \mu(dx) \mu(dy) - \kappa \int \log w(x) \mu(dx) + c,
\]

where \(\mu \in M_1(\Sigma)\) and

\[
c := \lim_{n \to \infty} \frac{1}{n^2} \log Z_n = - \inf_{\mu \in M_1(\Sigma)} \left\{ \frac{\kappa^2}{2} \int \int \left\{ \log |x^\theta - y^\theta|^{-1} + \log |x - y|^{-1} \right\} \mu(dx) \mu(dy) \right\} < \infty.
\]

From Theorem 3.1 a strong law of large numbers for the empirical measure of eigenvalues might be deduced, which is a not known in the situation of biorthogonal ensembles. We start by remembering the GUE case. It is well known due to Wigner, that

\[
\mathbb{E} \left( \frac{1}{N} \sum_{i=1}^{N} 1_{\{\lambda_i \leq x\}} \right) \to \int_{-\infty}^{x} \sigma(t)dt \quad (3.6)
\]

for \(N \to \infty\), where \(\sigma(t)\) is the density of the semicircular law,

\[
\sigma(t) = \frac{1}{2\pi} \sqrt{4 - t^2} 1_{[-2,2]}.
\]

But (3.6) does not only hold if the matrix entries are i.i.d. normally distributed, it suffices that the matrix is a Wigner matrix (i.e. all the matrix entries are i.i.d.). From (3.6) it can further be inferred that for any continuous and bounded function \(f\) on \(\Sigma\) \((f \in C_b(\Sigma))\),

\[
P(|\langle L_N, f \rangle - \langle \sigma, f \rangle| > \epsilon) \to 0, \quad \text{for } N \to \infty,
\]

which is a weak law of large numbers for \((L_N)_N\). Hereby, \(\langle \mu, f \rangle\) stands for \(\int f(t)\mu(dt)\).

A much stronger statement and a not obvious improvement is the strong law of large numbers,

\[
\lim_{N \to \infty} P(L_N \xrightarrow{\text{weak}} \sigma) = 1.
\]

For the biorthogonal case, such a result for the empirical measure of the eigenvalues can also shown to hold in some situations.

Corollary 3.2 Whenever the joint density of the eigenvalues is as in (3.3) and the rate function \(I\) has a unique minimizer \(\mu^*\), we obtain under (a1) and (a2) a strong law of large numbers,

\[
P(L_n \circ X_n^{-1} \xrightarrow{\text{weak}} \mu^*) = 1.
\]
3 LARGE DEVIATION PRINCIPLE FOR BIOORTHOGONAL ENSEMBLES

Proof:
Employing the upper bound of the LDP, the strong law of large numbers follows via an application of Borel-Cantelli’s lemma, see [18, Thm. II B.3]. □

Remark:
The rate function $I$ given in (3.4) achieves its minimum value, as it is a good rate function. Let $\mu^*$ be a minimizer of $I$. Then $\mu^*$ being a probability measure on $\Sigma$ can be characterized by:

$$-\kappa \log w(x) + \frac{\kappa^2}{2} \int \log |x^\theta - y^\theta|^{-1} \mu^*(dy) + \frac{\kappa^2}{2} \int \log |x - y|^{-1} \mu^*(dy)$$

$$= \inf_{\mu \in M_1(\Sigma)} -\kappa \int \log w(x) \mu^*(dx) + \frac{\kappa^2}{2} \int \int \left\{ \log |x^\theta - y^\theta|^{-1} + |x - y| \right\} \mu^*(dx) \mu^*(dy) = -c$$

$\mu^*$- a.s., and for all $x$ except possibly on a set with null logarithmic capacity, on which it may hold that

$$-\kappa \log w(x) + \frac{\kappa^2}{2} \int \log |x^\theta - y^\theta|^{-1} \mu^*(dy) + \frac{\kappa^2}{2} \int \log |x - y|^{-1} \mu^*(dy) \geq -c,$$

where the logarithmic capacity of a measurable set $A \subset \mathbb{R}$ is defined as

$$\gamma(A) := \exp \left\{ -\inf_{\nu \in M_1(\Sigma)} \int \int \log |x - y|^{-1} \nu(dx) \nu(dy) \right\}.$$ 

This characterization can be proved similar as in the proof of Lemma 13.2[e] in [23, p.116].

3.4 Preliminary Lemmas

We are going to collect some helpful lemmas in this section and in addition to section 3.2, we define $F : \Sigma \times \Sigma \to \mathbb{R}$ by

$$F(x, y) := -\frac{\kappa^2}{2} \left( \log |x - y| + \log |x^\theta - y^\theta| \right) - \frac{\kappa}{2} \left( \log w(x) + \log w(y) \right),$$

where we set

$$F(x, y) = \infty \quad \text{if} \quad x^\theta = y^\theta \quad \text{or if} \quad \{x, y\} \cap \mathcal{N}(w) \neq \emptyset.$$ 

Note that due to our definition of $\Sigma$, $x^\theta = y^\theta$ corresponds to $x = y$.

Let $F^M(x, y)$ denote the truncated version of $F(x, y)$, for $M > 0$

$$F^M(x, y) := F(x, y) \wedge M.$$
Furthermore, we define the functions $F_n : \Sigma \times \Sigma \to \mathbb{R}$,

$$F_n(x, y) := -\frac{1}{2} \left( \frac{p(n)}{n} \right)^2 (\log |x - y| + \log |x^\theta - y^\theta|) - \frac{p(n)}{2n} (\log w_n(x) + \log w_n(y)),$$

again with $F_n(x, y) = \infty$ if $x^\theta = y^\theta$ or if $\{x, y\} \cap \mathcal{N}(w_n) \neq \emptyset$ and their truncated versions

$$F_n^M(x, y) := F_n(x, y) \wedge M, \quad M > 0.$$

Before stating and proving the announced lemmas, we start with some simple calculus and write a simple sum in a complicated way: First, we observe that in an array with $1 \leq i < j \leq p(n)$, each number $k$ appears $p(n) - 1$ times,

\[
\begin{align*}
(1,2) & \quad (1,3) & \quad \ldots & \quad (1,p(n)) \\
(2,3) & \quad (2,4) & \quad \ldots & \quad (2,p(n)) \\
& \quad \ldots \quad & \quad \ldots & \quad (k,k+1) & \quad \ldots & \quad (k,p(n)) \\
& \quad \ldots \quad & \quad \ldots & \quad (p(n)-1,p(n)),
\end{align*}
\]

where $k$ appears $p(n) - k$ times in row $k$ and one time in each of the $k - 1$ lines above. This yields,

\[
\begin{align*}
n \sum_{i=1}^{p(n)} \log w_n(x_i) &= n \sum_{i=1}^{p(n)-1} \log w_n(x_i) - n \frac{p(n)}{p(n)} \left( \sum_{k=1}^{p(n)-1} k \log w_n(x_k) - \sum_{k=1}^{p(n)} k \log w_n(x_k) \right) \\
&= \frac{n}{p(n)} \left( \sum_{k=1}^{p(n)-1} (p(n) - k) \log w_n(x_k) + \sum_{k=1}^{p(n)} (k - 1) \log w_n(x_k) \right) \\
&\quad + \frac{n}{p(n)} \sum_{k=1}^{p(n)} \log w_n(x_k) \\
&= \frac{n}{p(n)} \sum_{1 \leq i < j \leq p(n)} (\log w_n(x_i) + \log w_n(x_j)) + \frac{n}{p(n)} \sum_{i=1}^{p(n)} \log w_n(x_i).
\end{align*}
\]

But the latter expression enables us to deduce from (3.3) and the definition of $F_n$ the following identity, where we abbreviate $1_{\Sigma^p(n)}(x_1, \ldots, x_p(n))$ by $1_{\Sigma^p(n)}(x)$,

\[
\begin{align*}
q_n(x_1, \ldots, x_p(n)) &= \frac{1}{Z_n} \exp \left\{ \sum_{i < j} \log \left( |x_i - x_j| + |x_i^\theta - x_j^\theta| \right) + n \sum_{i=1}^{p(n)} \log w_n(x_i) \right\} 1_{\Sigma^p(n)}(x) \\
&= \frac{1}{Z_n} \exp \left\{ -\frac{2n^2}{p(n)^2} \sum_{i < j} F_n(x_i, x_j) + \frac{n}{p(n)} \sum_{i=1}^{p(n)} \log w_n(x_i) \right\} 1_{\Sigma^p(n)}(x). \quad (3.7)
\end{align*}
\]
One key ingredient for the proof of the upper bound will be the following lemma, which also provides that the rate function is well defined.

**Lemma 3.3**

(i) For any $M > 0$, $F_n^M(x,y)$ converges to $F^M(x,y)$ uniformly as $n \to \infty$.

(ii) $F$ is bounded from below.

**Proof:**

It is obvious that for any $x, y \in \mathbb{R}$,

$$\log |x - y| \leq \log(|x| + 1) + \log(|y| + 1),$$

which implies

$$F_n(x, y) = -\frac{1}{2} \left( \frac{p(n)}{n} \right)^2 \left( \log |x - y| + \log |x^\theta - y^\theta| \right) - \frac{p(n)}{2n} \left( \log w_n(x) + \log w_n(y) \right)$$

$$\geq -\frac{p(n)^2}{2n^2} \left( \log \{(|x| + 1)(|y| + 1)\} + \log \{(|x^\theta| + 1)(|y^\theta| + 1)\} \right)$$

$$-\frac{p(n)}{2n} \log (w_n(x)w_n(y))$$

$$= -\frac{p(n)}{2n} \left[ \log \left( \left\lfloor \frac{|x| + 1}{|x^\theta| + 1} \right\rfloor \frac{p(n)}{n} w_n(x) \right) \right. + \log \left( \left\lfloor \frac{|y| + 1}{|y^\theta| + 1} \right\rfloor \frac{p(n)}{n} w_n(y) \right].$$

In order to prove the uniform convergence of $F_n^M$, we will show that $F^M = F_n^M = M$ on some specified sets and then deal with the complement of these sets. We will start by observing that

$$\log \left( \left\lfloor \frac{|x^\theta| + 1}{|x| + 1} \right\rfloor \frac{p(n)}{n} w_n(x) \right)$$

is bounded from above.

In case of $x \in [-1,1] \cap \Sigma$ it turns out that

$$\log \left( \left\lfloor \frac{|x^\theta| + 1}{|x| + 1} \right\rfloor \frac{p(n)}{n} w_n(x) \right) \leq \log \left( 4 \frac{p(n)}{n} \right) + \log(w_n(x)).$$

For $M > 0$ choose $n_1 \in \mathbb{N}$ such that for a given $\epsilon > 0$ it holds that $|\frac{p(n)}{n} - \kappa| < \epsilon$ and $|w_n(x) - w(x)| < \epsilon \forall n \geq n_1$ and $\forall x \in [-1,1]$. Due to the uniform convergence of $\log(w_n)$ on compact sets and the condition (a1.1), which asserts the existence of some $n_2 \in \mathbb{N}$ such that $\mathcal{N}(w_n) \subseteq \mathcal{N}(w) \forall n \geq n_2$, we can find for each $\nu \in \mathcal{N}(w)$ a $\delta^{(1)}_{\nu,M} > 0$, such that $\forall n \geq \max\{n_1, n_2\}$ and $\forall x \in [-1,1]$ with $|x - \nu| < \delta^{(1)}_{\nu,M}$

$$\log \left( 4 \frac{p(n)}{n} \right) + \log(w_n(x)) \leq \log(4^{4+\epsilon}) + \log(w(x)) \leq -M.$$
Whereas for |x| ≥ 1, we observe that for n ≥ n₁ as above the following holds
\[
\log \left( [(|x|^\theta + 1)(|x|+1)]^{\frac{\mu(n)}{\mu}} w_n(x) \right) \leq \log \left( 4^{\frac{\mu(n)}{\mu} |x|^{(\theta+1)\frac{\mu(n)}{\mu}} w_n(x) \right) \leq \log \left( 4^{\kappa+\epsilon|x|^{(\theta+1)(\kappa+\epsilon)}} \sup_{m \geq n} w_m(x) \right).
\]

Assumption (a₂) provides that for each M > 0 there exists n₃ ∈ N and R_M > 0 such that for |x| ≥ R_M and all n ≥ n₃ we have that
\[
\log \left( 4^{\kappa+\epsilon|x|^{(\theta+1)(\kappa+\epsilon)}} \sup_{m \geq n} w_m(x) \right) \leq -M.
\]

In case of |x| ≤ R_M reasoning as in the first part yields for each M > 0 and ν ∈ N(w) the existence of δ^2_ν,M > 0, such that for x with |x - ν| < δ^2_ν,M we also get
\[
\log \left( 4^{\frac{\mu(n)}{\mu} |x|^{(\theta+1)\frac{\mu(n)}{\mu}} w_n(x) \right) \leq -M.
\]

Putting this together, we finally obtain the following:
For each M > 0 there exists n₀ := max\{n₁, n₂, n₃\}, δ_M,ν := min\{δ^1_M,ν, δ^2_M,ν\} > 0 and R_M > 0 such that F_n(x,y) ≥ M holds for all n ≥ n₀ on
\[
A_M := \{|x| ∨ |y| > R_M\} \cup \bigcup_{\nu \in N(w)} \{|x - \nu| \wedge |y - \nu| < \delta_{\nu,M}\}. \tag{3.9}
\]

From (3.9) we see that A₅₅ is compact.
Moreover, we find a constant C_M > 0, depending on M, such that F_n ≥ M and F ≥ M ∀ x, y ∈ B_M := \{(x, y) ∈ \mathbb{R}^2 : |x - y| < C_M\}, and therefore, a constant constant 0 < \tilde{C}_M < ∞ exists with max\_{x,y∈B_M} \{-log |x - y|, -log |x^\theta - y^\theta|\} < \tilde{C}_M. This implies the existence of yet another constant 0 < C₁,M < ∞ such that
\[
\max_{x,y∈B_M \cap A_M} \{|log |x - y||, |log |x^\theta - y^\theta||\} < C₁,M.
\]

Note here that x^\theta = y^\theta only holds if x = y. Furthermore, we introduce the notation
\[
\| f \|_{∞}^D := \sup_{x ∈ D} |f(x)|,
\]
which restricts the supremum norm to a set D.
The continuity of w_n and the compactness of A₅₅ yields that
\[
\| log w_n(x) + log w_n(y) \|_{∞}^{A₅₅} < C₂,M, \text{ for a constant } 0 < C₂,M < ∞.
\]
Now for any given $\eta > 0$ we choose $N \in \mathbb{N}$ as large that
\[
\left| \frac{p(n)^2}{2n^2} - \frac{\kappa^2}{2} \right| \leq \frac{\eta}{3C_{1,M}},
\]
\[
\| \log w_n(x) + \log w_n(y) - \log w(x) - \log w(y) \|^\infty_{A_M^c \setminus B_M} \leq \frac{2\eta}{3\kappa}
\]
and
\[
\left| \frac{\kappa}{2} - \frac{p(n)}{2n} \right| \leq \frac{\eta}{3C_{2,M}}.
\]

Thus, we find on $D := A_M^c \setminus B_M$ for all $n \geq N$:
\[
\| F_n(x, y) - F(x, y) \|^D = \left\| \left( \frac{\kappa^2}{2} - \frac{p(n)^2}{2n^2} \right) \left( \log |x - y| + \log |x^\theta - y^\theta| \right) \right.
\]
\[
- \frac{p(n)}{2n} \left( \log w_n(x) + \log w_n(y) \right) + \frac{\kappa}{2} \left( \log w(x) + \log w(y) \right) \|^D
\]
\[
\leq \frac{\eta}{3C_{1,M}} \left\| \log |x - y| + \log |x^\theta - y^\theta| \|^D + \frac{\kappa}{2n} \left( \log w_n(x) + \log w_n(y) \right) \|^D
\]
\[
+ \frac{p(n)}{2n} - \frac{\kappa}{2} \left\| \log w_n(x) + \log w_n(y) \|^D
\]
\[
\leq \frac{\eta}{3C_{1,M}} C_{1,M} + \frac{\eta}{3C_{2,M}} C_{2,M} = \eta.
\]

Hence, we have established the uniform convergence of $F_n^M$ to $F^M$ on $A_M^c \setminus B_M$, whereas on $A_M \cup B_M$ we have $F_n^M = M$. That $F^M = M$ holds on $A_M \cup B_M$ can be shown along the same lines and (i) is proven.

Besides $F \geq M$ on $A_M \cup B_M$, we also know that $F$ is real-valued and continuous on the compact set $A_M^c \cap B_M^c$ and therefore $F$ is bounded from below, which yields (ii). $\Box$

### 3.4.1 Goodness of rate function

When it comes to proving the LDP for $(Q_n \circ L_n^{-1})_n$, we consider first the finite measure
\[
P_n := Z_n Q_n,
\]
where $Z_n$ is the partition function of the Lebesgue density belonging to $Q_n$, see (3.3).

The LDP will be derived for $(P_n \circ L_n^{-1})$ at first. In this section we will pay attention to the rate function in this LDP, more precisely for its level sets, which must be compact in case of being a good rate function.

Setting
\[
H(\mu) := \int F d\mu \otimes^2, \quad H^M(\mu) := \int F^M d\mu \otimes^2,
\]

one obtains well defined maps on $\mathcal{M}_1(\Sigma)$ (due to Lemma bounded). We claim the following
3.4 PRELIMINARY LEMMAS

**Lemma 3.4** \( H \) is a good rate function that governs the LDP for \((P_n \circ L_n^{-1})_n\).

**Proof:**

First, we observe that \( F^M \) is bounded and continuous: bounded from above by definition, from below due to the last lemma and continuous because we only regard continuous weight function \( w \). Therefore, taking a weakly convergent sequence \((\mu_n)_n\) it holds that as \( n \to \infty \),

\[
H^M(\mu_n) = \int F^M d\mu_n^{\otimes 2} \to \int F^M d\mu^{\otimes 2} = H^M(\mu)
\]

which provides that \( H^M \) is weakly continuous on \( M_1(\Sigma) \) for each \( M > 0 \).

By monotone convergence, we have pointwise on \( M_1(\Sigma) \)

\[
\lim_{M \to \infty} H^M(\mu) = \lim_{M \to \infty} \int F^M d\mu^{\otimes 2} = \int \lim_{M \to \infty} F^M d\mu^{\otimes 2} = \int F d\mu^{\otimes 2} = H.
\]

Thus, \( H \) is the limit of an increasing sequence of continuous functions, therefore \( H \) is lower semi-continuous. This in turn yields that the level sets \( \{ H \leq L \} \) are closed.

But the lemma claims that they are compact, and we proceed by letting \( m_F := |\inf F| \) and \( a > 0 \). Now, for any \( \mu \in M_1(\Sigma) \) one has

\[
\left( \inf_{x,y \in [-a,a]^c} (F + m_F)(x,y) \right) \mu([-a,a]^c)^2 \leq \int \int (F + m_F)(x,y) \mu(dx) \mu(dy) \leq H(\mu) + m_F,
\]

which yields

\[
\mu([-a,a]^c)^2 \leq \frac{H(\mu) + m_F}{\inf_{x,y \in [-a,a]^c} (F + m_F)(x,y)}.
\]

Thus, \( \{ H \leq L \} \subset K_L \), for \( L \in (0, \infty) \), with

\[
K_L := \bigcap_{a > 0} \left\{ \mu \in M_1(\Sigma) : \mu([-a,a]^c) \leq \left( \frac{L + m_F}{\inf_{x,y \in [-a,a]^c} (F + m_F)(x,y)} \right)^{1/2} \right\}.
\] (3.10)

Since

\[
\lim_{a \to \infty} \inf_{x,y \in [-a,a]^c} (F + m_F)(x,y) = \infty,
\]

we find for every \( \epsilon \) a compact interval \([-a, a] \subset \Sigma \), such that \( \mu([-a, a]) \geq 1 - \epsilon \) for every \( \mu \in K_L \). Hence, \( K_L \) is tight and according to Prohorov’s theorem it is therefore weakly compact. Finally, \( \{ H \leq L \} \) being a subset of a weakly compact set and being closed, it itself is compact with respect to the weak topology. This means that the rate function \( H \) is good. \( \square \)
3 LARGE DEVIATION PRINCIPLE FOR BIORTHOGONAL ENSEMBLES

3.4.2 Exponential tightness

In order to derive a full LDP for \((P_n \circ L_n^{-1})\), we will get a weak LDP, which we want to extend into a full one. The property, which will allow such an extension is the exponential tightness of \((P_n \circ L_n^{-1})\). It is shown next, because the construction (3.10) employed in the previous proof of Lemma 3.4 will also be used in proving the next lemma.

Lemma 3.5 \((P_n \circ L_n^{-1})_n\) is exponentially tight.

Proof:
Fix \(M > 0\), and define \(K_L (L > 0)\) as in (3.10), using \(F^M\) in the place of \(F\). For every \(\mu \in K_L^c\) there exists \(a = a_\mu > 0\) such that

\[
\left( \inf_{x,y \in [-a,a]^c} F^M(x,y) + m_{F^M} \right) \mu([-a,a]^c)^2 > L + m_{F^M},
\]

hence,

\[
\inf_{x,y \in [-a,a]^c} F^M(x,y) \mu([-a,a]^c)^2 > L.
\]

(3.11)

Provided that we can show for some Borel set \(A\) in \(M_1(\Sigma)\),

\[
\limsup_{n \to \infty} \frac{1}{n^2} \log P_n(L_n \in A) \leq - \inf_{\mu \in A} \int \int F^M(x,y) \mu(dx)\mu(dy), \tag{3.12}
\]

we can finish the proof of the lemma as follows:

\[
\limsup_{n \to \infty} \frac{1}{n^2} \log P_n(L_n \in K_L^c) \leq - \inf_{\mu \in K_L^c} \int F^M d\mu^\otimes2
\]

\[
= - \inf_{\mu \in K_L^c} \left( \int_{([-a_\mu,a_\mu]^c)^2} F^M d\mu^\otimes2 + \int_{\mathbb{R}^2 \setminus([-a_\mu,a_\mu]^c)^2} F^M d\mu^\otimes2 \right)
\]

\[
\leq - \inf_{\mu \in K_L^c} \int_{([-a_\mu,a_\mu]^c)^2} \inf_{x,y \in [-a_\mu,a_\mu]^c} F^M(x,y) d\mu^\otimes2 - \inf_{\mu \in K_L^c} \int_{\mathbb{R}^2 \setminus([-a_\mu,a_\mu]^c)^2} \inf F^M d\mu^\otimes2
\]

\[
\leq -L + m_{F^M}.
\]

Since \(\inf F^M > -\infty\), we have shown that there exists a sequence \((K_L)_L>0\) with

\[
\limsup_{L \to \infty} \limsup_{n \to \infty} \frac{1}{n^2} \log P_n(L_n \in K_L^c) = -\infty,
\]

which means, that \((P_n \circ L_n^{-1})_n\) is exponentially tight. Up to now, this proof rests on proving (3.12), but that will be shown in the next section (without using Lemma 3.5, of course). \(\square\)
3.5 Proof of the main theorem

Moving on, we will show that \((P_n \circ L_n^{-1})_n\) fulfills a weak LDP.

3.5.1 Obtaining a weak LDP

Proof of the upper bound

The crucial point in proving the upper bound is to overcome the singularities of the function \(F\). Therefore, we start by observing that for \(\Delta := \{(x, y) \in \Sigma^2 : x = y\}\),

\[
L_n \otimes L_n(\Delta) = \frac{1}{p(n)^2} \sum_{i,j} \delta_{\lambda_i} \otimes \delta_{\lambda_j}(\Delta) = \frac{1}{p(n)^2} \sum_i \delta_{\lambda_i} \otimes \delta_{\lambda_i}(\Delta) = \frac{1}{p(n)}
\]  

(3.13)

holds \(m_{\mathbb{R}^{p(n)}}\)-almost surely, because of the a.s. distinct eigenvalues under the product Lebesgue measure \(m_{\mathbb{R}^{p(n)}}\).

From (3.13) it follows that \(Q_n\)-a.s.

\[
\int \int_{x \neq y} F_n^M(x, y) L_n(dx)L_n(dy) = \int \int F_n^M(x, y) L_n(dx)L_n(dy) - \frac{M}{p(n)}.
\]

(3.14)

For a Borel set \(A\) in \(\mathcal{M}_1(\Sigma)\) let us define \(\tilde{A}\) as follows

\[
\tilde{A} := \{x \in \Sigma^{p(n)} : L_n(x) \in A\}.
\]

Due to the symmetry of \(F_n\) in its arguments we find

\[
2 \sum_{i<j} F_n(x_i, x_j) = \sum_{i \neq j} F_n(x_i, x_j),
\]

and since \(m_{\mathbb{R}^{p(n)}}\) is the product Lebesgue measure on \(\mathbb{R}^{p(n)}\), we have

\[
\int \exp \left\{ \frac{2n}{p(n)} \sum_{i=1}^{p(n)} \log w_n(x_i) \right\} m_{\mathbb{R}^{p(n)}}(dx) = \left( \int \exp \left\{ \frac{2n}{p(n)} \log w_n(t) \right\} m_\mathbb{R}(dt) \right)^{p(n)}.
\]
Thus, we obtain via Hölder’s inequality

\begin{align}
P_n\left(L_n \in A\right) &= \int \exp\left\{ -\frac{2n^2}{p(n)^2} \sum_{1 \leq i < j \leq p(n)} F_n(x_i, x_j) \right\} \\
&\quad \times \exp\left\{ \frac{n}{p(n)} \sum_{i=1}^{p(n)} \log w_n(x_i) \right\} m_{R(n)}(dx) \\
&\leq \left( \int \exp\left\{ 2n \log w_n(t) \right\} m_R(dt) \right)^{\frac{p(n)}{2}} \\
&\quad \times \left( \int \exp\left\{ -\frac{2n^2}{p(n)^2} \sum_{i \neq j} F_n(x_i, x_j) \right\} m_{R(n)}(dx) \right)^{1/2} \\
&= (I) \times (II).
\end{align}

First, we examine \( \lim_{n \to \infty} \frac{1}{n^2} \log (I) \) and claim

\[ \lim_{n \to \infty} \frac{1}{n^2} \log \left( \int \exp\left\{ 2n \log w_n(t) \right\} m_R(dt) \right)^{\frac{p(n)}{2}} = 0. \]

Since \( \frac{p(n)}{n} \to \kappa \) for \( n \to \infty \) it is sufficient to show the existence of

\[ \int \exp\left\{ 2n \log w_n(t) \right\} m_R(dt). \]

(3.16)

Take \( \epsilon > 0 \) as in (a2) and let \( K \in \mathbb{R} \) be that large, so that (a2) yields for \( |t| \geq K \):

\[ |t|^{(\theta+1)(\kappa+\epsilon)} \leq 1. \]

(3.17)

Hence,

\[ \int \exp\left\{ \frac{2n}{p(n)} \log w_n(t) \right\} m_R(dt) = \int w_n(t)^{\frac{2n}{p(n)}} m_R(dt) \]

\[ = \int_{[\neg K,K]} w_n(t)^{\frac{2n}{p(n)}} m_R(dt) + \int_{[\neg K,K]^c} w_n(t)^{\frac{2n}{p(n)}} m_R(dt). \]

The first integral exists due to the continuity of \( w_n \).

For the second one, we proceed as follows:

\[ \int_{[\neg K,K]^c} w_n(t)^{\frac{2n}{p(n)}} m_R(dt) \leq \int_{[\neg K,K]^c} \left( \frac{1}{|t|^{(\theta+1)(\kappa+\epsilon)}} \right)^{\frac{2n}{p(n)}} m_R(dt) \]

\[ \leq \int_{[\neg K,K]^c} \left( \frac{1}{|t|^{(\theta+1)+\epsilon}} \right)^{\frac{2n}{p(n)}} m_R(dt) \]
3.5 PROOF OF THE MAIN THEOREM

Now choose $N \in \mathbb{N}$ as big such that we have $\forall n \geq N : \left| \frac{n}{p(n)} - \frac{1}{\kappa} \right| < \frac{1}{2\kappa}$.

Thus we end up with

\[
\int_{[-K,K]^c} \left( \frac{1}{|t|^{(\theta+1)(\kappa+\epsilon)}} \right)^{\frac{2n}{p(n)}} m_{\mathbb{R}}(dt) = \int_{[-K,K]^c} \left( \frac{1}{|t|^{(\theta+1)(\kappa+\epsilon)}} \right)^{\frac{2n}{p(n)}} m_{\mathbb{R}}(dt) \\
\leq \int_{[-K,K]^c} \left( \frac{1}{|t|^{(\theta+1)(\kappa+\epsilon)\frac{2n}{p(n)}}} \right) m_{\mathbb{R}}(dt) \\
= \int_{[-K,K]^c} \left( \frac{1}{|t|^{(\theta+1)(\kappa+\epsilon)\frac{n}{\kappa}}} \right) m_{\mathbb{R}}(dt) \\
= \int_{[-K,K]^c} \left( \frac{1}{|t|^{(\theta+1)(\kappa+\epsilon)\frac{n}{\kappa}}} \right) m_{\mathbb{R}}(dt) \\
\leq \int_{[-K,K]^c} \frac{1}{|x|^{1+\eta}} m_{\mathbb{R}}(dt),
\]

with $\eta = \theta + (\theta + 1)\frac{\epsilon}{\kappa} > 0 \ \forall \ n \geq N$, which proves the existence of the integral (3.16).

Concerning (II), we find that for any $M > 0$

\[
(\text{II}) = \left\{ \int_{\mathcal{A}} \exp \left( -2n^2 \left( \mu \otimes (F_n^M) - \frac{M}{p(n)} \right) \right) m_{\mathbb{R}^p(n)}(dx) \right\}^{1/2} \\
\leq \left\{ \exp \left( -2n^2 \inf_{\mu \in \mathcal{A}} \mu \otimes (F_n^M) - \frac{M}{p(n)} \right) \right\}^{1/2} \\
= \exp \left( -n^2 \inf_{\mu \in \mathcal{A}} \mu \otimes (F_n^M) \right) \exp \left( \frac{Mn^2}{p(n)} \right).
\]

The first part of Lemma 3.3 yields,

\[
\lim_{n \to \infty} \left( \inf_{\mu \in \mathcal{A}} \mu \otimes (F_n^M) \right) = \inf_{\mu \in \mathcal{A}} \mu \otimes (F_n^M).
\]

We have thus shown that for any Borel set $A \subset \mathcal{M}_1(\Sigma)$ one has

\[
\limsup_{n \to \infty} \frac{1}{n^2} \log P_n(\mathcal{L}_n \in A) \leq - \inf_{\mu \in A} \int \int F^M(x,y) \mu(dx) \mu(dy). \quad (3.19)
\]

Furthermore, since (3.19) is the same as (3.12), which was needed to complete the proof of Lemma 3.5, we now know that the lemma (i.e. the exponential tightness of $(P_n \circ L_n^{-1})_n$) holds.

Now we choose for $A$ the set $B(\mu, \delta) = \{ \nu \in \mathcal{M}_1(\Sigma) : d(\nu, \mu) \leq 2\delta \}$ with $\delta > 0$ and

\[
d(\mu, \nu) = \sup \left| \int fd\mu - \int fd\nu \right|
\]

53
3 LARGE DEVIATION PRINCIPLE FOR BIORTHOGONAL ENSEMBLES

where the supremum is taken over all Lipschitz functions \( f \) for which the sum of the Lipschitz constant \( l_f \) and of the uniform bound \( \| f \|_\infty \) is less than or equal to 1. That distance is compatible with the weak topology, see [11, p.356].

Since \( \mu \mapsto H^M(\mu) \) is weakly continuous, from (3.19) we obtain for any \( \mu \in \mathcal{M}_1(\Sigma) \),

\[
\inf_{\delta \to 0} \limsup_{n \to \infty} \frac{1}{n^2} \log P_n \left( L_n \in B(\mu, \delta) \right) \overset{(3.19)}{\leq} - \inf_{\delta \to 0} \inf_{\nu \in B(\mu, \delta)} \int \int F^M(x, y) \nu(dx)\nu(dy) \\
\leq - \int \int F^M(x, y) \mu(dx)\mu(dy) = -H^M(\mu).
\]

Finally, letting \( M \) go to infinity, we obtain the following upper bound,

\[
\inf_{\delta \to 0} \limsup_{n \to \infty} \frac{1}{n^2} \log P_n \left( L_n \in B(\mu, \delta) \right) \leq - H(\mu).
\]

(3.20)

**Proof of the lower bound**

Turning to the lower bound, we take \( B(\mu, \delta) \) as above and for every \( \mu \in \mathcal{M}_1(\Sigma) \) we will show a lower bound, which resembles (3.20), namely

\[
\inf_{\delta > 0} \liminf_{n \to \infty} \frac{1}{n^2} \log P_n \left( L_n \in B(\mu, \delta) \right) \geq \frac{\kappa^2}{2} \int \int (|x^\theta - y^\theta| + \log |x - y|) \mu(dx)\mu(dy) \\
+ \kappa \int \log w(x)\mu(dx) = -H(\mu).
\]

(3.21)

Next, we state two assumptions for the measure \( \mu \) and explain, why they can be made.

They read as follows:

(i) \( \mu \) has no atoms,
(ii) \( \mathcal{S} := \text{supp}(\mu) \) is a compact subset of \( \Sigma \) with \( \mathcal{S} \cap (\mathcal{N}(w) \cup \{0\}) = \emptyset \).

Assumption (i) can be made, since if it does not hold, the right side of (3.21) would equal \(-\infty\) and the inequality would be trivial. Provided (ii) does not hold, we set

\[ A_k := [-k, k] \cap \Sigma \cap \left( \bigcup_{x \in \mathbb{N} \cup \{0\}} \left( x - \frac{1}{k}, x + \frac{1}{k} \right) \right)^c \text{ and } \mu_k := \frac{1}{\mu(A_k)}\mu|_{A_k}. \]

Because of

\[ \int F d\mu_k^{\otimes 2} = \lim_{k \to \infty} \int F d\mu_k^{\otimes 2}, \]

it suffices to show inequality (3.21) for \( \mu_k \). Combining this with (a1.1) allows us to assume that the support of \( \mu \) is contained in a finite union of compact sets that does not meet \( \mathcal{N}(w) \cup \{0\} \).
This justifies assumption (ii).

The main idea of the proof of the lower bound, is to localize the eigenvalues in small sets, and benefit from the speed $n^2$, which provides that the small volumes of these sets can be neglected. The technique is called coarse-graining and we look at the following division of $\mathcal{S}$. For $j = 1, \ldots, p(n)$ let $\xi_j = \xi_j^{(n)}$ the $\frac{p(n)+1-j}{p(n)}$ quantile of $\mu$ and set $\xi^{(n)} = (\xi_{p(n)}, \ldots, \xi_1)$, $\xi_{p(n)+1} := \inf \mathcal{S}$ and $\xi_0 = \xi_1 + 1$. Therefore, $\mu([-\infty, \xi_{p(n)+1}]) = 0 = \mu((\xi_1, \infty])$. Due to assumption (ii), we have $-\infty < \xi_{p(n)+1} < \xi_{p(n)} < \cdots < \xi_1 < \xi_0 < \infty$.

For $\delta > 0$, $t \in \mathbb{R}^{p(n)}$ write

\[ \pi_n(t) := \{ i = 1, \ldots, p(n) : t_i \geq 0 \}, \quad \nu_n(t) := \{ 1, \ldots, p(n) \} \setminus \pi_n(t), \]

\[ I_n(\delta) := \{ i = 1, \ldots, p(n) : |\xi_i^{(n)} - \xi_i^{(n)}| \leq \delta \}, \]

\[ J_j^{(n)}(t, \delta) := [t_j - \delta, t_j + \delta] \cap \Sigma, \quad j = 1, \ldots, p(n), \]

\[ J_j^{(n)}(t, \delta) := \prod_{j=1}^{p(n)} J_j^{(n)}(t, \delta). \]

We may assume that $\delta$ is fixed with $0 < \delta \leq \frac{1}{\theta^p n^2}$. Set

\[ \varphi_j^{(n)} := \varphi_j^{(n, \delta)} := \inf \{ w_n(x) : x \in [\xi_j - \delta, \xi_j + \delta] \cup [\xi_{j+1}, \xi_{j-1}] \}, \quad j = 1, \ldots, p(n), \]

and analogously define $\varphi_j$ when $w_n(x)$ is replaced by $w(x)$ in the above definition. Write $\psi_n$, resp. $\psi$ for the step function which equals $\varphi_j^{(n)}$ resp. $\varphi_j$ on $[\xi_{j+1}, \xi_j]$ and is zero elsewhere

\[ \psi_n = \sum_{j=1}^{p(n)} \varphi_j^{(n)} 1_{[\xi_j, \xi_{j+1}]}, \quad \psi = \sum_{j=1}^{p(n)} \varphi_j 1_{[\xi_j, \xi_{j+1}]}. \]

Moreover, we have:
Lemma 3.6 \(\exists n_0 \in \mathbb{N} \) such that
\[
\forall n > n_0 : \mathbb{I}_n(\xi^{(n)}, \delta) \subset \{ x \in \Sigma^{(n)} : L_n(x) \in B(\mu, 2\delta) \},
\]
\[
B(\mu, 2\delta) = \{ \nu \in M_1(\Sigma) : d(\nu, \mu) \leq 2\delta \} \text{ with }
\]
\[
d(\mu, \nu) = \sup_{\{ f \text{ Lipschitz: } l_f + \|f\|_\infty \leq 1 \}} \left| \int fd\mu - \int fd\nu \right|
\]
as above.

Proof:
Let \( y = (y_p(n), \ldots, y_1(n)) \in \mathbb{I}_n(\xi^{(n)}, \delta) \).
Then we apply [4, Le. 3.3], which says that there exists a \( N(\delta) \in \mathbb{N} \) such that
\[
d(L_n(\xi^{(n)}), \mu) < \delta \quad \forall \ n \geq N(\delta).
\]
Thus, we find for every \( n \geq N(\delta), \)
\[
d(L_n(y), \mu) \leq d(L_n(y), L_n(\xi^{(n)})) + d(L_n(\xi^{(n)}), \mu) \leq d(L_n(y), L_n(\xi^{(n)})) + \delta.
\]
Therefore, we conclude by
\[
d(L_n(y), L_n(\xi^{(n)})) = \sup_{\left\{ f \text{ Lipschitz: } l_f + \|f\|_\infty \leq 1 \right\}} \left| \int fdL_n(y) - \int fdL_n(\xi^{(n)}) \right|
\]
\[
= \sup_{\left\{ f \text{ Lipschitz: } l_f + \|f\|_\infty \leq 1 \right\}} \left| \frac{1}{p(n)} \sum_{i=1}^{p(n)} (f(y_i) - f(\xi_i)) \right|
\]
\[
\leq \sup_{\left\{ f \text{ Lipschitz: } l_f + \|f\|_\infty \leq 1 \right\}} \frac{1}{p(n)} \sum_{i=1}^{p(n)} |f(y_i) - f(\xi_i)| \leq |y_i - \xi_i| \leq \delta,
\]
since \( y_i \in [\xi_i - \delta, \xi_i + \delta] \) because we chose \( y \in \mathbb{I}_n(\xi^{(n)}, \delta) \).

Now for \( n \geq n_0 \), we find
\[
P_n(L_n \in B(\mu, 2\delta)) \overset{Le.3.6}{\geq} Z_nQ_n(\mathbb{I}_n(\xi, \delta)) = \int_{\mathbb{I}_n(\xi, \delta)} \prod_{i<j} |x_i^\theta - x_j^\theta| |x_i - x_j| \prod_{i=1}^{p(n)} w_n(x_i)^n m_p(n) (dx)
\]
\[
\geq \prod_{i=1}^{p(n)} \left( \varphi_i^{(n)} \right)^n \int_{\mathbb{I}_n(\xi, \delta)} \prod_{i<j} |x_i^\theta - x_j^\theta| |x_i - x_j| \ m_p(n) (dx)
\]
\[
\geq \prod_{i=1}^{p(n)} \left( \varphi_i^{(n)} \right)^n \int_{\mathbb{I}_n(\xi, \delta)} \prod_{i<j} |x_i^\theta - x_j^\theta| |x_i - x_j| \ m_p(n) (dx), \quad (3.23)
\]

56
where the last inequality follows from \( \mathbb{I}_n(\xi, \delta) \supseteq \mathbb{J}_n(\xi, \delta) \). Now we introduce the notation
\[
\mathbb{R}^{d+} := \{ x \in \mathbb{R}^d \mid x_1 > x_2 > \ldots > x_d \}
\]
and focus on the last integral,
\[
\int_{\mathbb{J}_n(\xi, \delta)} \prod_{i<j} |x_i^\theta - x_j^\theta| |x_i - x_j| m_p(n)(dx)
\]
\[
= \int_{[-\delta, \delta]^n} \prod_{i<j} \left| (x_i + \xi_i) - (x_j + \xi_j) \right| (x_i + \xi_i)^\theta - (x_j + \xi_j)^\theta m_p(n)(dx)
\]
\[
\geq \int_{[-\delta, \delta]^n} \prod_{i<j} \left| (x_i + \xi_i) - (x_j + \xi_j) \right| (x_i + \xi_i)^\theta - (x_j + \xi_j)^\theta m_p(n)(dx)
\]
\[
= \int_{[-\delta, \delta]^n} \prod_{i<j} \left| x_i - x_j + \xi_i - \xi_j \right| (x_i + \xi_i)^\theta - (x_j + \xi_j)^\theta m_p(n)(dx)
\]
(3.24)

On \( \mathbb{R}^p(n)^+ \), the following inequalities hold,
\[
\prod_{i<j-1} |x_i - x_j + \xi_i - \xi_j| \geq \prod_{i<j-1} |\xi_i - \xi_j| \quad (3.25)
\]
and as it can easily be seen (e.g. from \( ab < (a + b)^2 \), for \( a, b > 0 \))
\[
\prod_{i=1}^{p(n)-1} |x_i - x_{i+1} + \xi_i - \xi_{i+1}| \geq \prod_{i=1}^{p(n)-1} \left| x_i - x_{i+1} \right|^{\frac{1}{2}} |\xi_i - \xi_{i+1}|^{\frac{1}{2}}. \quad (3.26)
\]

Next, we will show that
\[
\left| (x_i + \xi_i)^\theta - (x_j + \xi_j)^\theta \right| \geq \left| x_i^\theta + \xi_i^\theta - x_j^\theta - \xi_j^\theta \right|, \quad (3.27)
\]
where the modulus can be omitted since both terms are greater than 0 on \( \mathbb{R}^{p(n)^+} \) due to our choice of the \( \xi_i \). If \( \xi_i > \xi_j > 0 \) and \( x_i > x_j > 0 \), we find that
\[
(x_i + \xi_i)^\theta - (x_j + \xi_j)^\theta = x_i^\theta + \xi_i^\theta - x_j^\theta - \xi_j^\theta + \sum_{k=1}^{\theta-1} \binom{\theta}{k} \left( x_i^k \xi_i^{\theta-k} - x_j^k \xi_j^{\theta-k} \right) > x_i^\theta + \xi_i^\theta - x_j^\theta - \xi_j^\theta. \quad (3.28)
\]
In case of \( 0 > \xi_i > \xi_j \) and \( 0 > x_i > x_j \), we first recall that this only can happen while \( \theta \) is odd. Nevertheless, (3.28) also holds: Within the sum, either \( \theta - k \) is odd and \( k \) is even or vice versa. Hence, if \( k \) is even \( 0 > \xi_i^{\theta-k} x_i^k > \xi_j^{\theta-k} x_j^k \), while the other case follows in an
3 LARGE DEVIATION PRINCIPLE FOR BIOORTHOGONAL ENSEMBLES

analogous way.
The last case to consider is \( \xi_i > 0 > \xi_j \) and \( x_i > 0 > x_j \). This case also only comes up for \( \theta \) being odd, and we find \( x_i^k \xi_i^{\theta-k} > 0 \) and \(-x_j^k \xi_j^{\theta-k} > 0 \) since either \( k \) is even or \( \theta - k \).

Thus, we have proven (3.27) and this can on one hand further be used to obtain either

\[
x_i^\theta + \xi_i^\theta - x_j^\theta - \xi_j^\theta > (x_i^\theta - x_j^\theta)^2 (\xi_i^\theta - \xi_j^\theta)^2
\]

(3.29)
or on the other hand yield

\[
x_i^\theta + \xi_i^\theta - x_j^\theta - \xi_j^\theta > \xi_i^\theta - \xi_j^\theta,
\]

(3.30)
since \( x_i^\theta > x_j^\theta \) on \( \mathbb{R}^{p(n)^+} \) for \( \theta \in \mathbb{N} \).

Putting together (3.25), (3.26), (3.29) and (3.30), we can bound (3.24) from below by

\[
\prod_{i<j} \left[ (\xi_i - \xi_j) (\xi_i^\theta - \xi_j^\theta) \right] \prod_{i=1}^{p(n)-1} \left[ (\xi_i - \xi_{i+1})^2 (\xi_i^\theta - \xi_{i+1}^\theta)^2 \right] \times
\]

\[
\int_{\left\{ -\delta,0 \right\} \times [0,\delta] \times \mathbb{R}^{p(n)}} \prod_{i=1}^{p(n)-1} \left[ (x_i - x_{i+1})^2 (x_i^\theta - x_{i+1}^\theta)^2 \right] m_{p(n)}(dx).
\]

(3.31)

Applying the mean value theorem for \( \theta \geq 2 \) (while for \( \theta=1 \) the following inequality will be an equality) now yields

\[
x_i^\theta - x_{i+1}^\theta = (x_i - x_{i+1}) \theta \zeta^\theta - 1,
\]

with \( \zeta \in [x_{i+1}, x_i] \).

Since \( |x_i| \leq \delta \forall i \) and \( \delta < \frac{1}{\theta \pi^2} \), we easily see that \( 0 < \theta \zeta^\theta - 1 < 1 \).

Therefore, \( 0 < x_i^\theta - x_{i+1}^\theta \leq x_i - x_{i+1} \) and we obtain

\[
\int_{\left\{ -\delta,0 \right\} \times [0,\delta] \times \mathbb{R}^{p(n)}} \prod_{i=1}^{p(n)-1} \left[ (x_i - x_{i+1})^2 (x_i^\theta - x_{i+1}^\theta)^2 \right] m_{p(n)}(dx)
\]

\[
\geq \prod_{i=1}^{p(n)-1} (x_i^\theta - x_{i+1}^\theta) m_{p(n)}(dx)
\]

\[
= \frac{1}{\theta p(n)} \int_{\left\{ -\delta,0 \right\} \times [0,\delta] \times \mathbb{R}^{p(n)}} \prod_{i=1}^{p(n)-1} (x_i - x_{i+1}) \prod_{i=1}^{p(n)} |x_i|^\frac{1-\theta}{2\pi^2} m_{p(n)}(dx)
\]

(3.32)

\[
\geq \frac{1}{\theta p(n)} \int_{\left\{ -\delta,0 \right\} \times [0,\delta] \times \mathbb{R}^{p(n)}} \prod_{i=1}^{p(n)-1} (x_i - x_{i+1}) m_{p(n)}(dx)
\]

(3.33)
The equality (3.32) is a simple application of the transformation formula, while the inequality (3.33) holds, because \( \theta \) is an integer and \( \delta < 1 \). If \( |x_i| \in [0, \delta^\theta] \), \( |x_i|^{\frac{1 - \theta}{\theta}} \geq 1 \) follows from
\[
\delta^\theta \geq |x_i| \iff |x_i|^{\frac{1 - \theta}{\theta}} \geq \delta^{1 - \theta} \geq 1
\]
since \( \frac{1 - \theta}{\theta} \leq 0 \).

For the last integral, we substitute \( u_{p(n)} = x_{p(n)}, u_{i-1} = x_{i-1} - x_i, i = p(n), \ldots, 2 \) and obtain
\[
\int_{[-\delta^\theta, 0]^p(n) \times [0, \delta^\theta]^p(n)} \prod_{i=1}^{p(n)-1} (x_i - x_{i+1}) m_{p(n)}(dx) = \int_{U_n} \prod_{i=2}^{p(n)} u_i m_{p(n)}(dx),
\]
where
\[
U_n = \left\{ u = (u_{p(n)}, \ldots, u_1) \mid u_p(n) \in [-\delta^\theta, \delta^\theta], u_i \in [0, \delta^\theta], i = 1, \ldots, p(n) - 1 : \sum_{i=1}^{p(n)} u_i \leq \delta^\theta \right\}.
\]

Since \( \cap_{1 \leq i \leq p(n)} \left\{ u_i \in [0, \frac{\delta^\theta}{p(n)}]\right\} \subset U_n \), we find the following lower bound of the last integral
\[
\int_{[0, \frac{\delta^\theta}{p(n)}]^p(n)} \prod_{i=2}^{p(n)} u_i m_{p(n)}(dx) = \left( \frac{1}{2} \right)^{p(n)-1} \left( \frac{\delta^\theta}{p(n)} \right)^{2(p(n)-1)+1}. \tag{3.34}
\]

From (3.33), (3.31) and (3.34) we get,
\[
P_n(L_n \in B(\mu, 2\delta)) \geq \prod_{i=1}^{p(n)} \left( \varphi_i^{(n)} \right) \prod_{i<j-1} \left[ (\xi_i - \xi_j) (\xi_i^\theta - \xi_j^\theta) \right] \times \prod_{i=1}^{p(n)-1} \left[ (\xi_i - \xi_{i+1})^{\frac{1}{2}} (\xi_i^\theta - \xi_{i+1}^\theta)^{\frac{3}{2}} \right] \left( \frac{1}{\theta p(n)} \right)^{\frac{1}{2}} \left( \frac{\delta^\theta}{p(n)} \right)^{2(p(n)-1)+1}.
\]

Hence,
\[
\frac{1}{n^2} \log P_n(L_n \in B(\mu, 2\delta)) \geq \left( \frac{p(n)}{n^2} \right)^2 \frac{1}{p(n)^2} \sum_{i<j-1} \log \left[ (\xi_i - \xi_j) (\xi_i^\theta - \xi_j^\theta) \right] \tag{3.35}
+ \left( \frac{p(n)}{n^2} \right)^2 \frac{1}{2p(n)^2} \sum_{i=1}^{p(n)-1} \log \left[ (\xi_i - \xi_{i+1}) (\xi_i^\theta - \xi_{i+1}^\theta) \right] \tag{3.36}
+ \frac{1}{n} \sum_{i=1}^{p(n)} \log \varphi_i^{(n)} \tag{3.37}
+ \frac{1}{n^2} \log \left( 2 \left( \frac{1}{2\delta} \right)^{p(n)} \right) + \frac{2p(n) - 1}{n^2} (\log \delta^\theta - \log p(n)). \tag{3.38}
\]
3 LARGE DEVIATION PRINCIPLE FOR BIORTHOGONAL ENSEMBLES

It is easily seen that (3.38) converges to 0 as $n$ tends to infinity. Now, for (3.35) and (3.36) observe that for $\theta = 1$, resp. $\theta \in \mathbb{N}$,

$$
\int_{x < y} \log(y^\theta - x^\theta) \mu(dx) \mu(dy) - \int_{(x,y) \in [\xi_{p(n)} + 1, \xi_{p(n)}]} \log |y^\theta - x^\theta| \mu(dx) \mu(dy)
$$

$$
= \sum_{i < j} \int_{(x,y) \in [\xi_{j+1}, \xi_j] \times [\xi_i, \xi_i + 1]} \log(y^\theta - x^\theta) \mu(dx) \mu(dy)
$$

$$
+ \frac{1}{2} \sum_{i=1}^{p(n)-1} \int_{(x,y) \in [\xi_i, \xi_i + 1]^2} \log |y^\theta - x^\theta| \mu(dx) \mu(dy)
$$

$$
\leq \frac{1}{p(n)^2} \sum_{i < j} \log(\xi_i^\theta - \xi_j^\theta + 1) + \frac{1}{2p(n)^2} \sum_{i=1}^{p(n)-1} \log(\xi_i^\theta - \xi_i^\theta + 1).
$$

Because $\xi_{p(n)+1}$ was defined to be the infimum of the support $S$ and we look at measures without atoms, we observe that $\xi_{p(n)} \to \xi_{p(n)+1} = \inf S$ for $n \to \infty$. Thus, we find

$$
\lim_{n \to \infty} \int_{(x,y) \in [\xi_{p(n)} + 1, \xi_{p(n)}]^2} \log |y^\theta - x^\theta| \mu(dx) \mu(dy) = \int_{\emptyset} \log |y^\theta - x^\theta| \mu(dx) \mu(dy) = 0.
$$

We have shown that the limit of (3.35) and (3.36) can be bounded from below in the following way,

$$
\int \int \log(y^\theta - x^\theta) \mu(dx) \mu(dy) = \frac{1}{2} \int_{x < y} \log(y^\theta - x^\theta) \mu(dx) \mu(dy)
$$

$$
\leq \frac{1}{2} \lim_{n \to \infty} \left( \frac{1}{p(n)^2} \sum_{i < j} \log(\xi_i^\theta - \xi_j^\theta + 1) + \frac{1}{2p(n)^2} \sum_{i=1}^{p(n)-1} \log(\xi_i^\theta - \xi_i^\theta + 1) \right). \tag{3.39}
$$

As to (3.37), observe that

$$
\frac{1}{n} \sum_{j=1}^{p(n)} \log \phi_j^{(n)} = \frac{p(n)}{n} \int \log \psi_n d\mu,
$$

where $\psi_n$ was defined in (3.22). Denote by $l$ a Lipschitz constant of $\log w$ on $S$ according to (a1.3). For $\eta > 0$ write

$$
l^{n} := \max \{ |\log w(x) - \log w(y)| : |x - y| \leq \eta \}.
$$

Note that $l^{n} \leq l \eta$, again because of the Lipschitz continuity (a1.3) and since $N(w) \cap S = \emptyset$. Next, define

$$
M(n, \delta) := \bigcup_{j \in I_n(\delta)} [\xi_{j+1}^{(n)}, \xi_j^{(n)}]
$$
3.5 PROOF OF THE MAIN THEOREM

and

\[ C := \max \{ \log w(x) : x \in S \} - \min \{ \log w(x) : x \in S \}, \]

where assumption (ii) provides that \( C \) is finite.

Let \( \epsilon > 0 \). Since for all \( n \) and \( j \geq 1 \) one has \( \mu \left( [\xi_{j+1}^{(n)}, \xi_j^{(n)}] \right) = 1/p(n) \), and since \( \delta \) is fixed, one has \( \mu(M(n, \delta)) \leq \epsilon \) for large \( n \). Now let \( n \) be large enough such that one also has \( \| \log w_n - \log w \|_{\infty} \leq \epsilon \), which is possible because of (a1.2). Then

\[
\int |\log \psi_n - \log w| d\mu \leq \int |\log \psi_n - \log w \pm \log \psi| d\mu
\]

\[
\leq \sum_{j \in I_n(\delta)} \int [\xi_{j+1}, \xi_j] \{ |\log \psi - \log w| + \epsilon \} d\mu
\]

\[
+ \sum_{j \in I_n(\delta)} \int [\xi_{j+1}, \xi_j] \{ |\log \psi - \log w| + \epsilon \} d\mu + \epsilon
\]

\[
\leq \frac{p(n)}{p(n)} (\epsilon^2 + \epsilon) + (C + \epsilon)\epsilon
\]

\[
\leq 2l\delta + (C + \epsilon + 1)\epsilon
\]

This yields

\[
\limsup_{n \to \infty} \frac{1}{n} \sum_{j=1}^{p(n)} \log \varphi_j^{(n)} - \kappa \int \log w \ d\mu = O(\delta). \tag{3.40}
\]

Putting together (3.35)-(3.40) yields the lower bound (3.21), whenever the empirical measure lies in a ball around some measure \( \mu \) in \( M_1(\Sigma) \).

3.5.2 Extending the weak LDP into a full one

Finally, for any \( \mu \in M_1(\Sigma) \) we have obtained

\[
\inf_{\delta > 0} \limsup_{n \to \infty} \frac{1}{n^2} \log P_n(L_n \in B(\mu, \delta)) \leq -H(\mu) \tag{3.41}
\]

and

\[
\inf_{\delta > 0} \liminf_{n \to \infty} \frac{1}{n^2} \log P_n(L_n \in B(\mu, \delta)) \geq -H(\mu). \tag{3.42}
\]

As \( \Sigma \) is a closed subset of the Polish space \( \mathbb{R} \), it is Polish as well, see [2, Chap. 26], which in turn guarantees that \( M_1(\Sigma) \) is a Polish space, cf. [42, Thm. 6.2].

Since the \( B(\mu, \delta) \) are a base for the weak topology on \( M_1(\Sigma) \), we have to verify the condition...
of Theorem 1.4 resp. [11, Thm. 4.1.11], which is

\[ I(\mu) := \sup_{\delta > 0} \liminf_{n \to \infty} \frac{1}{n^2} \log P_n(L_n \in B(\mu, \delta)) \]

This is equivalent to show that

\[ \inf_{\delta > 0} \liminf_{n \to \infty} \frac{1}{n^2} \log P_n(L_n \in B(\mu, \delta)) = \inf_{\delta > 0} \limsup_{n \to \infty} \frac{1}{n^2} \log P_n(L_n \in B(\mu, \delta)). \]  

(3.43)

Comparing (3.43) to (3.41) and (3.42), we can see that Theorem 1.4 is applicable, which therefore provides a weak LDP for \((P_n \circ L_n^{-1})_n\) on \(M_1(\Sigma)\) with rate function \(H\) and speed \(n^2\).

But our original aim was to derive a LDP for \((Q_n \circ L_n^{-1})_n\), which means that we have to control the partition function \(Z_n\), since \(P_n = Z_n Q_n\).

Setting \(A = G = M_1(\Sigma)\) in the lower and upper bound at the LDP for \((P_n \circ L_n^{-1})_n\), we obtain

\[ \lim_{n \to \infty} \frac{1}{n^2} \log Z_n = - \inf_{\mu \in M_1(\Sigma)} \int F d\mu \otimes^2. \]

The right hand side has been shown to be \(< +\infty\) by Lemma 3.3 (ii) and establishes (3.5).

Now,

\[ \frac{1}{n^2} \log Q_n(L_n \in A) = \frac{1}{n^2} (\log P_n(L_n \in A) - \log Z_n) \]

for any Borel set \(A\) in \(M_1(\Sigma)\). Hence Theorem 3.1 is proven.

### 3.6 Examples

We conclude this section by the application of Theorem 3.1 to three classes of matrix ensembles, that are mentioned in [6] and where the joint distribution of the eigenvalues is of form (3.3). As can clearly be seen from (3.3), these ensembles are specified by their domain \(\Sigma\) and their corresponding weight functions \(w_n\).

#### Biorthogonal Jacobi ensembles

This ensemble has two real parameters \(\alpha, \beta > -1\) and its weight function is as for (orthogonal) Jacobi polynomials, where we adapted the notation to suit (3.3). The ensemble is characterized by

\[ \Sigma = (0, 1), \quad w_n(x) = x^\alpha (1 - x)\beta, \quad \alpha, \beta > -1. \]  

(3.44)
Note that in [6], the factor \((1 - x)\beta\) does not appear. While technical reasons led Borodin to that restriction of setting \(\beta\) equal to zero, we can allow for general Jacobi weights of form (3.44). For the orthogonal case \((\theta = 1)\), it is explained in [21, Chap. 2.8] that this orthogonal Jacobi ensemble appears in the canonical correlation analysis, where the correlations coefficients turn out to be the square root of the eigenvalues of a special kind of matrix. It is proved there, that the eigenvalues of these type of matrices follow a joint distribution function of type (3.1) with Jacobi weight functions (3.44).

For the following treatment in the biorthogonal setting \((\theta \in \mathbb{N})\) we will distinguish two cases. First, we look at constant parameters \(\alpha\) and \(\beta\), but finish with the case of \(n\)-dependent parameters \(\alpha(n)\) and \(\beta(n)\). In both cases, \(p(n)\) is such that \(\frac{p(n)}{n} \to \kappa\) for \(n \to \infty\).

**Constant parameters**

When taking real parameters \(\alpha, \beta > -1\) (where we allow for \(\alpha = 0\) or \(\beta = 0\)) we find \(\lim_{n \to \infty} w_n = 1\). Thus, \(\log(w) = 0\) and the conditions (a1) are trivially satisfied. The sequence \((Q_n \circ L_n^{-1})_n\) obeys a LDP with speed \(n^2\) and rate function

\[
I(\mu) = \frac{\kappa^2}{2} \int \int \left\{ \log \left| x^\theta - y^\theta \right|^{-1} + \log \left| x - y \right|^{-1} \right\} \mu(dx) \mu(dy) + c_{Jac},
\]

where \(\mu \in \mathcal{M}_1(\Sigma)\) and as above

\[
c_{Jac} := \lim_{n \to \infty} \frac{1}{n^2} \log Z_n^{Jac} < \infty.
\]

**\(n\)-dependent parameters**

Next, we assume that the parameters \(\alpha(n)\) and \(\beta(n)\) are such that

\[
\lim_{n \to \infty} \frac{\alpha(n)}{n} = \alpha > -1 \quad \text{and} \quad \lim_{n \to \infty} \frac{\beta(n)}{n} = \beta > -1.
\]

Hence, \(\lim_{n \to \infty} w_n = w(x) = x^\alpha (1 - x)^\beta\). This implies \(\lim_{n \to \infty} \log w_n \to \log w\). On compact subsets of \((0, 1)\) the convergence is uniform, since the continuous function \(\log w\) achieves a minimum and maximum on these sets and also has bounded derivatives there. Therefore, \(\log w\) is also Lipschitz continuous on compact subsets and (a1) is fullfilled. Theorem 3.1 yields that the sequence \((Q_n \circ L_n^{-1})_n\) obeys a LDP with speed \(n^2\) and rate function

\[
I(\mu) = \frac{\kappa^2}{2} \int \int \left\{ \log \left| x^\theta - y^\theta \right|^{-1} + \log \left| x - y \right|^{-1} \right\} \mu(dx) \mu(dy)
- \kappa \int \alpha \log x \ \mu(dx) - \kappa \int \beta \log (1 - x) \mu(dx) + \tilde{c}_{Jac},
\]
where $\mu \in \mathcal{M}_1(\Sigma)$ and \[ \tilde{c}_{Jac} := \lim_{n \to \infty} \frac{1}{n^2} \log \tilde{Z}_{Jac}^n < \infty. \]

For $\theta = 1$, we have recovered the result from [31, Prop. 2.1].

### Biorthogonal Laguerre ensembles/Disordered bosons

The second class of biorthogonal Laguerre ensembles is given by $\Sigma = (0, \infty)$, $w_n(x) = x^\alpha e^{-\frac{x}{\tau}}$, for a real parameter $\alpha > -1$.

This setup describes one interesting quasi-particle system in physics, which is the system of disordered bosons and has been introduced in [37], cf. chapter 2. The probability distribution for the positive eigenvalues of the matrix $iX$ (where $X$ is as in (2.16)) turns out to be

$$d\mu_{n,\alpha}(x_1, \ldots, x_n) = \frac{1}{Z_{n,\alpha}} \prod_{k=1}^n x_k^\alpha e^{-\tau x_k} \prod_{1 \leq i < j \leq n} |x_i - x_j| |x_i^2 - x_j^2| 1_{\Sigma^n}(x_1, \ldots, x_n) dx_k.$$  

Thus, it corresponds to $\theta = 2$, $p(n) = n$ and the weight function is slightly modified,

$$w_n(x) = x^\alpha e^{-\frac{x}{\tau}}.$$  

We remember that $\tau^{-1}$ was the variance of the independent and normally distributed random variables, that were used to construct the stability matrix $h$, cf. (2.17) and (2.18). Taking the variance $\tau^{-1}$ equal to $n^{-1}$ we will again separately treat the cases of a constant and $n$-dependent parameter and for the sake of generality, we will also keep $p(n)$ with $\lim_{n \to \infty} \frac{p(n)}{n} = \kappa$.

### Constant parameter

Choosing a real constant $\alpha > -1$ provides $w_n(x) \to w(x) = e^{-x}$ for $n \to \infty$, with $\mathcal{N}(w_n) = \mathcal{N}(w) = \emptyset$ on $(0, \infty)$. In addition, $\log w(x) = -x$ is Lipschitz on compact sets and $\log w_n = \frac{\alpha}{n} \log x - x$ converges uniformly to $-x$ on compact sets, so (a1) holds in this context. Because of

$$\lim_{x \to +\infty} x^{(\theta+1)(1+\epsilon)+\frac{\kappa}{2}} e^{-x} = 0,$$

for any $n \in \mathbb{N}$, (a2) is established as well. As a corollary of our main Theorem 3.1 we get a large deviation principle for $(Q_n \circ L_n^{-1})_n$.

**Corollary 3.7** $(Q_n \circ L_n^{-1})_n$ satisfies a LDP on $\mathcal{M}_1(\Sigma)$ with respect to the weak topology with speed $n^2$ and good rate function

$$I(\mu) = -\frac{\kappa^2}{2} \int \int \log \left(|x^2 - y^2| |x - y|\right) \mu(dx) \mu(dy) + \kappa \int x \mu(dx) + c_{Log},$$

where $\mu \in \mathcal{M}_1(\Sigma)$ and

$$\tilde{c}_{Jac} := \lim_{n \to \infty} \frac{1}{n^2} \log \tilde{Z}_{Jac}^n < \infty.$$
where $\mu \in \mathcal{M}_1(\Sigma)$ and 
\[ c_{\text{Lag}} := \lim_{n \to \infty} \frac{1}{n^2} \log Z_{n}^{\text{Lag}} < \infty. \]

**n-dependent parameter**

Taking a sequence $(\alpha(n))_{n \in \mathbb{N}}$ with $\lim_{n \to \infty} \frac{\alpha(n)}{n} = \alpha > -1$, $w_n(x)$ tends to $w(x) = x^\alpha e^{-x}$. Again, we have $\mathcal{N}(w_n) = \mathcal{N}(w) = \emptyset$ and on compact subsets of $(0, \infty)$, $\log w_n$ converges uniformly towards $\log w$, while the bounded derivative of $\log w(x) = \alpha \log x - x$ on such sets is responsible for $\log w$ being Lipschitz. We also find 
\[ \lim_{x \to +\infty} x^{(\theta+1)(1+\epsilon)+\frac{\alpha(n)}{n}} e^{-x} = 0, \]
for any $n \in \mathbb{N}$, which gives condition (a2). On that basis, we get the following corollary of Theorem 3.1.

**Corollary 3.8** For biorthogonal Laguerre ensembles, $(Q_n \circ L_n^{-1})_n$ satisfies a LDP on $\mathcal{M}_1(\Sigma)$ with respect to the weak topology with speed $n^2$ and good rate function 
\[ I(\mu) = -\frac{\kappa}{2} \int \int \log \left( |x^2 - y^2| |x - y| \right) \mu(dx) \mu(dy) \]
\[ -\kappa \int \alpha \log x \mu(dx) + \kappa \int x \mu(dx) + \tilde{c}_{\text{Lag}}, \]
where $\mu \in \mathcal{M}_1(\Sigma)$ and 
\[ \tilde{c}_{\text{Lag}} := \lim_{n \to \infty} \frac{1}{n^2} \log Z_{n}^{\text{Lag}} < \infty. \]

Note that we have stuck to the choice $\theta = 2$ stemming from the disordered boson system.

**Remark:**

(1) In [37], the authors proved
\[ \mathbb{E} \left[ \frac{1}{N} \sum_{i=1}^{N} 1_{\{\lambda_i \leq x\}} \right] \to \int_{-\infty}^{x} \rho(t) dt \quad \text{for } N \to \infty, \]
with 
\[ \rho(t) = \frac{1}{2\pi} \left( \frac{t}{b} \right)^{-\frac{1}{2}} \left( 1 + \sqrt{1 - \left( \frac{t}{b} \right)^2} \right)^{\frac{1}{2}} - \left( 1 - \sqrt{1 - \left( \frac{t}{b} \right)^2} \right)^{\frac{1}{2}}, \]
(3.45)
for $0 < t \leq b := 3\sqrt{3}$.

We improved this result, see discussion after Corollary 3.2, obtaining a strong law of large numbers for the empirical distribution of the eigenvalues,
\[ \lim_{N \to \infty} P(L_N \overset{\text{weak}}{\longrightarrow} \rho) = 1. \]
(2) There exists a unique minimizer $\mu^*$ for $I$, $I(\mu^*) = 0$. Its density was computed by two independent methods by means of a variational calculation or by using biorthogonal polynomials in [37, Sect. IV and Sect.V.D.].

Now, we present a third way. Choose $\kappa = 1$ and let $\alpha$ be the $l$ in [37]. We have to find a probability measure $\mu^*$, which fulfills

$$- \log x + \int \log |x^2 - y^2|^{-1} \mu^*(dy) + \int \log |x - y|^{-1} \mu^*(dy) = -\tilde{c}_{\text{Log}}. \tag{3.46}$$

Provided that the corresponding rate function $I$ of Corollary 3.8 is convex (what the authors claim in [37]), we can find that $\mu^*$ is unique. In order to show that it has the density $\rho$ as in (3.45), one needs only to show that $\mu^* = \rho(x)dx$ satisfies (3.46). For this calculation we need to exactly know $\tilde{c}_{\text{Log}} = \lim_{N \to \infty} \log Z_{\text{Log}}^N$, which is of a ‘Selberg integral’-type. But this approach requires the knowledge of the minimizer $\mu^*$ (or resp. of $\rho$) and does not present a way of deducing $\rho$.

**Biorthogonal Hermite ensembles**

The last class of ensembles, to which our LDP will be applied are biorthogonal Hermite ensembles, which are of a much more general form than known from Hermite polynomials. These ensembles are defined by

$$\Sigma = (-\infty, \infty), \quad w_n(x) = |x|^\alpha e^{-\frac{x^2}{n}}, \tag{3.47}$$

and the real parameter has to satisfy $\alpha > -1$ here as well. The special case of $\alpha = 0$ yields the common Hermite weights. As in the last example, we choose the variance to be $\frac{1}{n}$, which gives weight functions $w_n(x) = |x|^\frac{\alpha}{n} e^{-x^2}$.

**Constant parameter**

Taking $\alpha > -1$, $w_n(x) \to w(x) = e^{-x^2}$ for $n \to \infty$, and it can be shown as above that conditions (a1) and (a2) are met. The rate function for the LDP for the empirical measure of eigenvalues appearing in biorthogonal ensembles with Hermite weight functions $e^{-x^2}$ is as follows,

$$I(\mu) = -\frac{\kappa^2}{2} \int \int \log (|x^\theta - y^\theta| |x - y|) \mu(dx) \mu(dy) + \kappa \int x^2 \mu(dx) + c_{\text{Her}},$$

where $\mu \in \mathcal{M}_1(\Sigma)$ and

$$c_{\text{Her}} := \lim_{n \to \infty} \frac{1}{n^2} \log Z_{n}^{\text{Her}} < \infty.$$
3.6 EXAMPLES

\textbf{n-dependent parameter}

When the parameter depends on \( n \) in such a way that \( \lim_{n \to \infty} \frac{a(n)}{n} = \alpha > -1 \), it is obvious that \( w_n(x) \to w(x) = |x|^\alpha e^{-x^2} \) for \( n \to \infty \) and again, conditions (a1) and (a2) are met.

We finish this section as well as the whole chapter by a corollary of Theorem 3.1, which gives a new behavior concerning Hermite ensembles, since an application of \( n \)-dependent weight functions in this setting is as yet unknown.

\textbf{Corollary 3.9} For biorthogonal Hermite ensembles, \( (Q_n \circ L_n^{-1})_n \) satisfies a LDP on \( \mathcal{M}_1(\Sigma) \) with respect to the weak topology with speed \( n^2 \) and good rate function

\[
I(\mu) = -\frac{\kappa^2}{2} \int \int \log \left( |x^\theta - y^\theta| |x - y| \right) \mu(dx) \mu(dy) + \kappa \int x^2 \mu(dx) - \kappa \int \alpha \log x \mu(dx) + \tilde{c}_{\text{Her}},
\]

where \( \mu \in \mathcal{M}_1(\Sigma) \) and

\[
\tilde{c}_{\text{Her}} := \lim_{n \to \infty} \frac{1}{n^2} \log \tilde{Z}_{n}^{\text{Her}} < \infty.
\]
3 LARGE DEVIATION PRINCIPLE FOR BIORTHOGONAL ENSEMBLES
4 Fluctuations in a Multi-Matrix model

As we have introduced in detail in chapter two, the multi-matrix model is a model, that generalizes the classical GUE \( (2.5) \) in two distinct ways. First, it is a model in which we look at \( m \) matrices at once, where \( m \in \mathbb{N} \) is a fixed integer number. This obviously has been the reason to name it multi-matrix model. Moreover, this model not only allows for several matrices but it is also a perturbation of the classical GUE. The perturbation is given by a special type of potential, which will be specified later on.

Our first result, that has been obtained in that multi-matrix model, is a moderate deviation principle (MDP) for normalized traces of a non-commutative monomial. Moreover, we could obtain an alternative proof of a special version of a CLT for normalized traces of a non-commutative monomial, which was first derived in [25] in greater generality. Both results are connected to map enumeration. In the rate function of the MDP as well as in the variance of the CLT does a function appear, which can be understood as a generating function for special types of maps. Thus, this chapter continues by spelling out this connection in detail in an extensive excursus, in which we thoroughly introduce all the corresponding quantities, so that the combinatorics behind the connection of such matrix integrals and map enumeration becomes apparent. The rate function of our MDP will also be analyzed with respect to their combinatorial interpretation.

Then we will give the proofs of our results and generalize our results from monomials to polynomials.

The potential \( V \) will always be a polynomial with coefficients lying in a vicinity of the origin. But there is another restriction on the choice of coefficients. In the multi-matrix model, they have to provide that \( V \) satisfies some convexity condition (which will be Definition 4.3 below). This also covers the one-matrix model, but we also consider a second class of admissible coefficients, that is sufficient to obtain similar results in the one-matrix case. We will compare these two choices and it turns out that this comparison for the one-matrix model rewards us with the extension of our results to some potentials, that do not obey the notion of convexity.
Finally, this chapter ends by giving examples of multi-matrix models as they appear in physics, but that have also attracted the interest of mathematicians. These models are the Ising model on random graphs, a matrix model coupled in a chain and the so called $q$-Potts model.

### 4.1 Notations

Let us recall the general setup as mentioned in the second chapter. In the multi-matrix model, we consider a set of $m$-tuples $H = (H_1, \ldots, H_m)$ of $N \times N$ hermitian matrices, where the (complex) matrix entry in line $k$ and row $l$ of the matrix $H_i$ is referred to by $H_i(kl)$. Since the matrices are hermitian, we only need to specify the distribution of the entries on and above the diagonal, and they are such that $\Re(H_i(kl)), k < l$, $\Im(H_i(kl)), k < l$, $2^{-\frac{1}{2}} H_i(kk)$ is a family of independent real Gaussian variables of variance $(2N)^{-1}$.

The space of $m$ hermitian matrices of dimension $N \times N$ whose entries follow that Gaussian distribution will be denoted by $\mathcal{H}_N(\mathbb{C})^m$.

#### 4.1.1 The measure under investigation

If we denote the Lebesgue measure on $\mathcal{H}_N(\mathbb{C})^m$ by $d^N H$, the law of the classical $m$-dimensional Gaussian unitary ensemble (GUE) is as follows:

$$\mu^N(d H) = \frac{1}{Z^N} \exp \left\{ -\frac{N}{2} \text{tr} \left( \sum_{i=1}^m H_i^2 \right) \right\} d^N H,$$

where $Z^N$ is the normalizing constant.

In this chapter, we will examine the following perturbation of the $m$-dimensional GUE, which is also a measure on $\mathcal{H}_N(\mathbb{C})^m$:

$$\mu^N_V(d H) = \frac{1}{Z^N_V} \exp \left\{ -N \text{tr} (V(H_1, \ldots, H_m)) \right\} \mu^N(d H),$$

where $\mu^N(d H)$ is the measure (4.1), $Z^N_V$ is the normalizing constant for the perturbed measure $\mu^N_V$ and $V(H_1, \ldots, H_m)$ is the so-called potential function. We will now specify the potential $V$ of our model more precisely.

#### 4.1.2 The potential $V$

In the course of the whole chapter, we consider polynomial potential functions $V(H_1, \ldots, H_m)$ of form

$$V(H_1, \ldots, H_m) := V(t(H_1, \ldots, H_m)) := \sum_{i=1}^n t_i q_i(H_1, \ldots, H_m)$$
with \( n \in \mathbb{N}, \ t = (t_1, \ldots, t_n) \in \mathbb{C}^n \) and fixed non-commutative monomials \( q_j, j = 1, \ldots, n \), of type
\[
q_j(H_1, \ldots, H_m) = H_{j_1} \cdots H_{j_{r_j}} \quad \text{for} \quad j_k \in \{1, \ldots, m\}, \ k = 1, \ldots, r_j, \ \text{with} \ r_j \geq 1.
\]

From (4.2) we see that the perturbed measure reduces to the one of the \( m \)-dimensional GUE, provided that we set our potential function equal to zero.

We will abbreviate \( V(H_1, \ldots, H_m) \) resp. \( V_t(H_1, \ldots, H_m) \) by \( V \) resp. \( V_t \), whenever it does not cause any confusion. We know the general form of \( V_t \) from (4.3) and in this paragraph, we will introduce two conditions that the potential must fulfill.

**Self-adjoint potentials**

**Definition 4.1** We say that the potential \( V_t \) is self-adjoint, if \( V_t = V_t^\dagger \) holds with respect to the involution \( \dagger \) that is given for all \( z \in \mathbb{C} \) and all monomials \( q_l(H_1, \ldots, H_m) = H_{l_1} \cdots H_{l_p} \) by
\[
(zq_l)^\dagger = (zH_{l_1} \cdots H_{l_p})^\dagger = zH_{l_p} \cdots H_{l_1}.
\]

(4.4)

Thus, if the potential \( V \) is self-adjoint, all the appearing monomials \( q_l \) are also self-adjoint.

Note that a self-adjoint potential \( V_t \) always has a real trace, which can be seen as follows. First, we now that \( \text{tr}(q_l) = \text{tr}(q_l^\dagger) \), because \( V \) is self-adjoint. Next, we observe that for hermitian matrices \( X_j \) and self-adjoint monomials \( q_l \), we obtain
\[
\text{tr}(q_l) = \text{tr}(H_{l_1} \cdots H_{l_p}) = \text{tr}(H_{l_p} \cdots H_{l_1}) = \text{tr}(q_l^\dagger),
\]

(4.5)

which yields \( \text{tr}(q_l) = \text{tr}(q_l^\dagger) = \text{tr}(q_l^\dagger) = \text{tr}(q_l) \in \mathbb{R} \). Hence, for a self-adjoint potential \( V_t \), it holds that \( \text{tr}(V_t) \in \mathbb{R} \) as can be seen from:
\[
2V_t = V_t + V_t^\dagger = \sum_{l=1}^n t_l q_l + \sum_{l=1}^n \overline{t_l} q_l^\dagger = \sum_{l=1}^n \Re(t_l)(q_l + q_l^\dagger) + \sum_{l=1}^n i \Im(t_l)(q_l - q_l^\dagger)
\]

\[\iff \quad V_t = \sum_{l=1}^n \Re(t_l)\frac{q_l + q_l^\dagger}{2} - \sum_{l=1}^n i \Im(t_l)\frac{q_l - q_l^\dagger}{2i},\]

(4.6)

because we have \( \text{tr}(q_l) = \text{tr}(q_l^\dagger) \in \mathbb{R} \).

Throughout the whole chapter, we will assume that \( V_t \) is self-adjoint.
4 FLUCTUATIONS IN A MULTI-MATRIX MODEL

\textbf{c-convexity}

Next, we introduce the notion of convexity and c-convexity, since all our results require that the appearing potential function \( V \) is c-convex.

\textbf{Definition 4.2} We will say that \( V \) is convex, if for any \( N \in \mathbb{N} \)
\[
\varphi_V^N : \mathcal{H}_N(\mathbb{C})^m = (\mathbb{R}^{N^2})^m \rightarrow \mathbb{R} \quad \begin{array}{c}
(H_1, \ldots, H_m) \\
\text{tr} (V(H_1, \ldots, H_m))
\end{array}
\]
is a convex function of the entries of the hermitian matrices \( H_1, \ldots, H_m \).

If we add a Gaussian potential \( \frac{1}{2} \sum_{i=1}^{m} H_i^2 \) to \( V \) (and that exactly happens in (4.2) through the density of the measure (4.1)) the hypothesis on \( V \) can be relaxed in the following way:

\textbf{Definition 4.3} If there exists \( c > 0 \), such that for all \( N \in \mathbb{N} \), the function
\[
\varphi_V^N : \mathcal{H}_N(\mathbb{C})^m = \mathbb{R}^{N^2m} \rightarrow \mathbb{R} \quad \begin{array}{c}
(H_1, \ldots, H_m) \\
\text{tr} (V(H_1, \ldots, H_m) + \frac{1-c}{2} \sum_{i=1}^{m} H_i^2)
\end{array}
\]
is real and convex as a function of the entries of the \( N^2m \) matrices, we say that \( V \) is c-convex.

Note that the requirement that \( \text{tr}(V) \) is real is included in the definition of c-convexity, while it is automatically met in case of \( V \) being self-adjoint as seen above.

As the mapping \( \varphi_V^N \) is defined on \( \mathbb{R}^{N^2m} \), an alternative definition of \( V \) being c-convex, is that the Hessian of the mapping \( \varphi_V^N \) is positive semi-definite and the condition of c-convexity implies that \( Z_V^N \) is automatically finite, see [25]. An example of a c-convex \( V \) is
\[
V = \sum_{i=1}^{n} P_i \left( \sum_{k=1}^{m} \alpha_k^i H_k \right) + \sum_{k,l} \beta_{k,l} H_k H_l
\]
with convex real polynomials \( P_i \) in one unknown and for all \( l, \sum_k |\beta_{k,l}| \leq (1 - c) \). This is due to Klein’s Lemma, see [23], which we now state.

\textbf{Lemma 4.4 (Klein’s lemma)} Let \( f : \mathbb{R} \rightarrow \mathbb{R} \) be a convex function. Then, if \( H = (H(ij))_{1 \leq i,j \leq N} \) is a \( N \times N \) hermitian matrix, the mapping
\[
\psi_V^N : \mathbb{C}^{N(N+1)/2} \rightarrow \mathbb{R} \quad \begin{array}{c}
(H(ij))_{1 \leq i \leq j \leq N} \\
\sum_{i=1}^{N} f(\lambda_i(H))
\end{array}
\]
is convex, where \( (\lambda_i(H))_{i=1,\ldots,N} \) denotes the real eigenvalues of \( H \). Furthermore, if \( f \) is twice differentiable with \( f''(x) > c \) for all \( x \), the mapping \( \psi_V^N \) is also twice differentiable and with its Hessian is bounded from below by \( c I_{N \times N} \).
As \( V \) is self-adjoint, the monomials are also self-adjoint and consist of products of hermitian matrices, and such a product is also hermitian, which can be seen as follows,

\[
q_l = H_{i_1} \cdots H_{i_k} = \overline{H_{i_1}^t \cdots H_{i_k}^t} = (H_{i_1} \cdots H_{i_k})^t = ((H_{i_1} \cdots H_{i_k})^t)^t
\]

\[
q_l^* \equiv \overline{(H_{i_1} \cdots H_{i_k})^t} = (H_{i_1} \cdots H_{i_k})^* = q_l^*.
\]

Thus, the trace of each monomial is a function of non-negative eigenvalues, and Klein’s lemma will be a helpful tool to establish \( c \)-convexity. Hence, a special class of examples are \( V \) of the form

\[
V_t = \sum_{i=1}^{n} t_i \left( \sum_{k=1}^{m} \alpha_k^i H_k \right)^{2p_i}
\]

with real non-negative \( t_i \)'s, integers \( p_i \)'s and real \( \alpha \)'s. Note that this class is very similar to the classes of potentials \( V \) that were considered in [20] in the one-matrix case \( m = 1 \). In this one-dimensional case a different assumption for \( V_t \) was made and in section 4.7 we will compare these two assumptions. It will turn out that there are polynomial potential functions, which are not \( c \)-convex but do fall in the class of functions considered in [20], and explicit examples will be given.

Taking now the potential \( V_t \) as in (4.3), we next define for any \( \eta > 0 \) and \( c > 0 \)

\[
B_{\eta,c} = \left\{ t \in \mathbb{C}^n \mid |t| = \max_{1 \leq i \leq n} |t_i| \leq \eta, V_t \text{ is } c-\text{convex} \right\}.
\]

From now on, let \( E \) denote the expectation with respect to the probability measure \( \mu_{V_t}^N \) defined in (4.2), unless it is otherwise stated.

Now, we can state our results.

\section*{4.2 Results}

For a self-adjoint potential \( V \), the non-commutative monomial \( q_l \) appearing in \( V \) will also be self-adjoint and we define

\[
\varphi_l : \mathcal{H}_N(\mathbb{C})^m \rightarrow \mathbb{R}^m \rightarrow \text{tr}(q_l(H_1, \ldots, H_m)) - E[\text{tr}(q_l(H_1, \ldots, H_m))]
\]

where the expectation is with respect to \( \mu_{V_t}^N \). As \( \varphi_l \) is continuous, it is measurable.
Theorem 4.5 (A Moderate Deviation Principle) Let \( t = (t_1, \ldots, t_n) \in \mathbb{R}^n, c > 0 \) and \( V_t = \sum_{i=1}^{n} t_i q_i(H_1, \ldots, H_m) \) be self-adjoint with \( t_i \neq 0 \). Then there exists \( \eta > 0 \), such that for \( n \in \mathbb{N} \), with \( t \in B_{\eta,c} \cap \mathbb{R} \) the sequence of distributions \( \left( \mu_{V_t}^n \circ \left( \frac{1}{N^{\gamma}} \bar{\varphi}_l \right) \right)^{-1} \) obeys a MDP on \( \mathbb{R} \) with speed \( N^{2\gamma} \) and rate function

\[
I(x) = \frac{x^2}{2} \left( \frac{\partial^2}{\partial t_l^2} F^0(t) \right)^{-1}
\]

for any \( 0 < \gamma < 1 \). The function \( F^0(\cdot) \) is given by

\[
F^0(t) = \sum_{k \in \mathbb{N}^n} \frac{(-t)^k}{k!} C^k_0,
\]

where \( k! = \prod_{i=1}^{n} k_i! \), \( (-t)^k = \prod_{i=1}^{n} (-t_i)^{k_i} \) and \( C^k_0 \) being the number of maps on a surface of genus 0 with \( k_i \) vertices of type \( q_i \).

As mentioned earlier on, only CLT-type results have been obtained in the one-matrix as well as in the multi-matrix model, see [20] for \( m = 1 \) and [25] for \( m \geq 1 \). Besides the MDP in [10], Theorem 4.5 is one of the first MDPs in connection with random matrices. In [10], the MDP for the distribution function of the empirical measure is derived with respect to the topology of distribution in law, our MDP for the scaled trace of a non-commutative monomial holds on \( \mathbb{R} \). Moreover, the MDP covers the whole range of possible scalings \( \frac{1}{N^{\gamma}} \), for \( 0 < \gamma < 1 \).

Before we will give some details on the function \( F^0 \) and the connection between matrix models and map enumeration, we will state a special case of the CLT obtained by Guionnet and Maurel-Segala in [25], for which we will give an alternative proof.

Theorem 4.6 (A Central Limit Theorem) Let \( t = (t_1, \ldots, t_n) \in \mathbb{R}^n \) and \( c > 0 \). Then there exists \( \eta > 0 \), such that for a self-adjoint \( V_t = \sum_{i=1}^{n} t_i q_i(H_1, \ldots, H_m) \), \( t_i \neq 0 \), \( n \in \mathbb{N} \), with \( t \in B_{\eta,c} \cap \mathbb{R} \) and any fixed \( l \in \{1, \ldots, n\} \) the distribution of the random variable

\[
\bar{\varphi}_l := \text{tr}(q_l(H_1, \ldots, H_m)) - \mathbb{E}[\text{tr}(q_l(H_1, \ldots, H_m))]
\]

with respect to \( \mu_{V_t}^n \) converges towards the normal distribution with expectation 0 and variance \( \frac{\partial^2}{\partial t_l^2} F^0(t) \).

As the second derivative of the function \( F^0 \) appears in Theorem 4.5 and 4.6, we want to stress the combinatorial interpretation of this function (and its derivatives) in terms of generating functions for some kind of graphs in the following excursus.
4.3 Excursus: Map enumeration and matrix integrals

In the sixties, Tutte started to deal with the problem of enumerating maps when he was studying combinatorial problems, such as the four color problem. Some ten years later, t’Hooft draw the first connection between some physical quantities in Quantum Chromodynamics (QCD) and Feynman diagrams. Building on that observation, enumeration of maps could be related to problems in Quantum Field Theory and matrix integrals could be used to count maps (e.g. see [7], [5]). This was only later on put on a solid mathematical foundation in [20], [25] and [39].

In this section, we try to explain how matrix integrals and map enumeration are related.

4.3.1 Basic notions for map enumeration

Let us first start by giving some definitions from graph theory, which are relevant for the enumeration of maps.

Definition 4.7 A graph $G=(V,E)$ consists of a set of vertices $V$, and a set of edges $E$, where $E \subset V \times V$.

The graph is said to be directed, if the edges are pointing from one vertex to another and otherwise, the graph is undirected. Provided that the vertices and edges are labelled, we speak of a labelled graph. The number of incoming (resp. outgoing) edges at a vertex is called the degree of the vertex.

A surface is a compact oriented two-dimensional manifold without boundary. Each surface has a genus $g$, which is the number of handles (or holes) it possesses. Up to homeomorphism, there is only one surface with a given genus, see [1, Sect. 5.4], and therefore, we will characterize a surface only by its genus from now on.

![Surfaces of genus g= 0, 1 and 2](image)

Figure 4.1: Surfaces of genus $g= 0, 1$ and $2$
Definition 4.8 A map is a connected graph drawn on surface, such that the edges do not intersect and that cutting the surface along the edges provides a disjoint union of sets (which are also called faces), where each face is homeomorphic to an open disk.

Note that the condition to obtain faces homeomorphic to open disks is equivalent to demand that the genus of the surface should be minimal in order to avoid intersecting edges.

Due to the definition of a map, we see that not every graph leads to a map, once it is drawn on a surface, see Figure 4.2 and 4.3.

Finally, we conclude our sequence of definitions with a reminder of the Euler characteristic $\chi$ for a surface of genus $g$

$$\chi = 2 - 2g = \#\text{vertices} + \#\text{faces} - \#\text{edges}.$$  

4.3.2 One-matrix case

Now we restrict ourself to the one-matrix case and recall some facts about Gaussian integrals, which will help us to see later on, how Wick’s formula ties in with a graphical interpretation. This section relies on the great survey paper by Zvonkin, [53].
Gaussian integrals and Wick’s formula for hermitian matrices

On the real line, we know that the normal (or Gaussian) distribution has the Gaussian measure \( \mu \) as density
\[
d\mu(x) := \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}} dx,
\]
where \( x, m, \sigma \in \mathbb{R}, \sigma > 0 \) and \( dx \) is the Lebesgue measure on \( \mathbb{R} \).

For any measure on a space \( X \) and for any function \( f : X \rightarrow \mathbb{R} \) or \( f : X \rightarrow \mathbb{C} \), we will write \( \mathbb{E}_\mu[f] := \mathbb{E}[f] \) for the expectation of \( f \) with respect to \( \mu \):
\[
\mathbb{E}[f] := \int_X f(x) d\mu(x),
\]
where the space \( X \) and the measure \( \mu \) should be clear from the context, but note that the expectation \( \mathbb{E} \) in this excursus is not taken with respect to the measure \( \mu_N^V \).

Let us recall some basic facts for the Gaussian measure:
\[
\begin{align*}
\mathbb{E}[1] &= 1 \\
\mathbb{E}[x] &= m \\
\mathbb{E}[(x-m)^2] &= \sigma^2 \\
\varphi(t) := \mathbb{E}[e^{itx}] &= e^{imt-\frac{\sigma^2 t^2}{2}},
\end{align*}
\]
which basically says that \( \mu \) is a probability measure on \( \mathbb{R} \), that the mean of the distribution is \( m \) and the variance \( \sigma^2 \). Since the characteristic function \( \varphi(t) \) uniquely determines a probability distribution, we could also have defined the Gaussian measure via (4.9). For the standard normal distribution \( m \) is set to 0 and \( \sigma^2 \) equals 1.

Since we want to operate on \( \mathbb{R}^d \) later on, we need to define a Gaussian measure on \( \mathbb{R}^d \). For \( x, y \in \mathbb{R}^d \) let \( \langle x, y \rangle \) denote the canonical scalar product \( \langle x, y \rangle = \sum_{i=1}^d x_i y_i \).

**Definition 4.9** A measure \( \mu \) on \( \mathbb{R}^d \) is called Gaussian, if its characteristic function is as follows,
\[
\varphi(t) = \int_{\mathbb{R}^d} e^{i\langle t, x \rangle} d\mu(x) = \exp \left\{ i\langle m, t \rangle - \frac{1}{2} \langle Ct, t \rangle \right\},
\]
where \( m, t \in \mathbb{R}^d \) and \( C \) is a symmetric \( (d \times d) \)-matrix.

The vector \( m \) is called mean vector of the measure \( \mu \) and will be set to 0 from now on. Thus, \( \mathbb{E}[x_i] = 0 \ \forall \ i = 1, \ldots, d \) and the covariance matrix \( C \) has entries \( c_{ij} = \mathbb{E}[x_i x_j], i, j = 1, \ldots, d \).

When \( C \) is non-degenerate, the measure has density
\[
d\mu(x) = \text{Const} \times \exp \left\{ -\frac{1}{2} \langle Bx, x \rangle \right\} dx,
\]
where
\[
B := \begin{pmatrix} 1 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & 1 \end{pmatrix}.
\]
where $B = C^{-1}$. In order to normalize that measure, we have to take $\text{Const} = (2\pi)^{-\frac{d}{2}}(\det B)^{\frac{1}{2}}$. From (4.11) we see that the measure $\mu$ has an even density function and therefore we find $\mathbb{E}[f] = 0$ for every monomial $f(x) = x^{2k+1}$, where $k \in \mathbb{N}$.

The calculation of the expectation of monomials with an even degree or of any other polynomial can be simplified by the following theorem, which will also be the link to map enumeration as we will show later on.

**Theorem 4.10 (Wick’s formula)** Let $f_1, \ldots, f_{2k}$ be a set of linear functions of $x_1, \ldots, x_d$, and the expectation taken w.r.t. to the Gaussian measure (4.11), then

$$
\mathbb{E}[f_1 f_2 \cdots f_{2k}] = \sum_{\text{Wick couplings}} \mathbb{E}[f_{p_1} f_{q_1}] \mathbb{E}[f_{p_2} f_{q_2}] \cdots \mathbb{E}[f_{p_k} f_{q_k}],
$$

where a Wick coupling is a permutation $p_1 q_1 p_2 q_2 \cdots p_k q_k$ of the set of indices $1, 2, \ldots, 2k$, such that $p_1 < p_2 < \cdots < p_k$ and $p_1 < q_1, p_2 < q_2, \ldots, p_k < q_k$.

Note that the number of Wick couplings of a set $1, 2, \ldots, 2k$ is equal to

$$(2k - 1)!! := 1 \times 3 \times 5 \times \cdots \times (2k - 1).$$

As an example let us consider the measure (4.11) with $\mathbb{R}^2$ and $C = 1_{2 \times 2}$, where we want to calculate the expectation $\mathbb{E}[x_1^2 x_2^2]$. Thus $f_1(x_1, x_2) = x_1 = f_2(x_1, x_2)$ and $f_3(x_1, x_2) = x_2 = f_4(x_1, x_2)$. The 3 Wick couplings of 1234 are 1234, 1324 and 1423 and therefore

$$
\mathbb{E}[x_1^2 x_2^2] = \mathbb{E}[f_1 f_2 f_3 f_4] = \mathbb{E}[f_1 f_2] \mathbb{E}[f_3 f_4] + \mathbb{E}[f_1 f_3] \mathbb{E}[f_2 f_4] + \mathbb{E}[f_1 f_4] \mathbb{E}[f_2 f_3]
$$

$= \mathbb{E}[x_1^2] \mathbb{E}[x_2^2] + \mathbb{E}[x_1 x_2] \mathbb{E}[x_1 x_2] + \mathbb{E}[x_1 x_2] \mathbb{E}[x_1 x_2] = 1 + 0 + 0 = 1,$

since $\mathbb{E}[x_i x_j] = c_{ij}$ for $i, j = 1, 2$.

Because we want to operate on the space of hermitian matrices $\mathcal{H}_N(\mathbb{C})$ (remember, that we are still in the one-matrix case, $m = 1$), we will now show, that this is only a special case for the Gaussian measure on $\mathbb{R}^d$. Given a $N \times N$ hermitian matrix $H = (h_{ij})$, let us use the following notation, $\text{Re}(h_{ij}) = x_{ij}$ and $\text{Im}(h_{ij}) = y_{ij}$. Due to the fact that $h_{ij} = \overline{h_{ji}}$, every hermitian matrix is specified by exactly $N^2$ parameters, which are $h_{ij} = x_{ii} \in \mathbb{R}, i = 1, \ldots, N$ and $x_{ij}, y_{ij}, 1 \leq i < j \leq N$. Therefore, $\mathcal{H}_N(\mathbb{C})$ is isomorphic to $\mathbb{R}^{N^2}$ and the corresponding Lebesgue measure is

$$
d^N(H) := \prod_{i=1}^{N} dx_{ii} \prod_{i<j} dx_{ij} dy_{ij}.
$$
With some abuse of notation, \( d^N(H) \) is the same measure as \( d^N(H) \) in (4.1) for \( m = 1 \). In order to specify the measure, we need a quadratic form on \( \mathcal{H}_N(\mathbb{C}) \), which will be \( \text{tr}(H^2) \). An easy calculation shows, that

\[
\text{tr}(H^2) = \sum_{i=1}^{N} x_{ii} + 2 \sum_{i<j} \left( x_{ij}^2 + y_{ij}^2 \right).
\]

Now we identify \( z \in \mathbb{R}^{N^2} \) with

\[
(x_{11}, x_{22}, \ldots, x_{NN}, x_{12}, \ldots, x_{1N}, x_{23}, \ldots, x_{2N}, \ldots, x_{N-1N}, y_{12}, \ldots, y_{1N}, y_{23}, \ldots, y_{2N}, \ldots, y_{N-1N})^t
\]

and find that the corresponding \( N^2 \times N^2 \) matrix \( B \) for this quadratic form to yield \( \text{tr}(H^2) = \langle Bz, z \rangle \) is diagonal, with the first \( N \) diagonal terms being 1 and the other \( N^2 - N \) diagonal terms being 2:

\[
B = \begin{pmatrix}
1 & & & & \\
& \ddots & & & \\
& & 1 & & \\
& & & \ddots & \\
& & & & 2
\end{pmatrix}.
\]

Thus, the partition function \( Z^N \) of the measure of the GUE (4.1) can be given explicitly and the measure itself is of form

\[
\mu^N(dH) = \frac{1}{Z^N} \exp \left\{ -\frac{N}{2} \text{tr}(H^2) \right\} d^N H = \left( \frac{1}{\pi^{N^2/2N}} \right)^\frac{1}{2} \exp \left\{ -\frac{N}{2} \text{tr}(H^2) \right\} d^N H,
\]

whereas \( B \) was replaced by \( \tilde{B} = NB \) for the sake of consistency with (4.1). As the new covariance matrix we obtain \( C = \tilde{B}^{-1} \) and this yields \( \mathbb{E}[x_{ii}] = \frac{1}{N} \) and \( \mathbb{E}[x_{ij}^2] = \frac{1}{2N} = \mathbb{E}[y_{ij}^2] \), \( i < j \). Regarding the matrix entries \( h_{ij} \) as linear functions of the variables \( x_{ij}, y_{ij} \), we observe that

\[
\mathbb{E}[h_{ij}h_{kj}] = \frac{1}{N} \delta_{il} \delta_{jk},
\]

(4.13) since \( \mathbb{E}[h_{ii}^2] = \mathbb{E}[x_{ii}^2] = \frac{1}{N} \) and \( \mathbb{E}[h_{ij}h_{ji}] = \mathbb{E}[x_{ij}^2 + y_{ij}^2] = \frac{1}{2N} + \frac{1}{2N} = \frac{1}{N} \), while in every other case only off-diagonal terms of \( C \) are to be considered, which are 0.

This enables us now to calculate the integral of some polynomial, say \( \text{tr}(H^{2k}) \) via Wick’s formula. We will look at a special case, choosing \( k = 5 \) and firstly, we see that

\[
\text{tr}(H^{10}) = \sum h_{1i1j}h_{1i2j}h_{1i3j}h_{1i4j}h_{1i5j}h_{1i6j}h_{1i7j}h_{1i8j}h_{1i9j}h_{1i10j}h_{1i11j}.
\]
where the sum is taken over all $N^{10}$ combination of indices $i_1, \ldots, i_{10}$. In order to compute the integral $\mathbb{E}[\text{tr}(H^{2k})]$, we apply Wick’s formula and by writing $f_i := h_{i,i+1}$, we will take a closer look at one Wick coupling (out of the 945 ones), which contribute to the sum, namely the following: $f_1 f_2 f_3 f_4 f_5 f_6 f_8 f_7 f_9$, which corresponds to

$$\mathbb{E}[h_{i_1 i_2} h_{i_4 i_5}] \mathbb{E}[h_{i_2 i_3} h_{i_3 i_4}] \mathbb{E}[h_{i_5 i_6} h_{i_{10} i_1}] \mathbb{E}[h_{i_6 i_7} h_{i_8 i_9}] \mathbb{E}[h_{i_7 i_8} h_{i_9 i_{10}}].$$

(4.14)

Due to (4.13), we know what the restrictions for the indices are, provided that the contribution of that Wick coupling to the sum should be nonzero:

$$\mathbb{E}[h_{i_1 i_2} h_{i_4 i_5}] = \frac{1}{N} \text{ if } i_1 = i_5, \ i_2 = i_4,$$

$$\mathbb{E}[h_{i_2 i_3} h_{i_3 i_4}] = \frac{1}{N} \text{ if } i_2 = i_4,$$

$$\mathbb{E}[h_{i_5 i_6} h_{i_{10} i_1}] = \frac{1}{N} \text{ if } i_5 = i_1, \ i_6 = i_{10},$$

$$\mathbb{E}[h_{i_6 i_7} h_{i_8 i_9}] = \frac{1}{N} \text{ if } i_6 = i_9, \ i_7 = i_8,$$

$$\mathbb{E}[h_{i_7 i_8} h_{i_9 i_{10}}] = \frac{1}{N} \text{ if } i_7 = i_{10}, \ i_8 = i_9,$$

(4.15)

leading to

$$i_1 = i_5, \ i_2 = i_4, \ i_3, \ i_6 = i_7 = i_8 = i_9 = i_{10}.$$

This means that we can choose 4 indices out of $N$ possible ones (always $i_3$ and for example, $i_1, i_4$ and $i_9$), leading to $N^4$ possibilities, which contribute to the sum. Thus, we need to get for each coupling, the number of ‘free parameters’ $f$, which contributes $\frac{1}{N^7} N^f = N^{f-5}$ to the integral, while the factor $\frac{1}{N^5}$ stems from (4.14) and (4.15).

Having this in mind, we now turn to the combinatorial interpretation.

**Wick’s formula and graphs**

The graph, we are going to consider is a 10-sided polygon, in which the vertices are labelled $i_1, \ldots, i_{10}$, has only one face and in which the edges are directed, such that the directed edges form a circle.
What we are doing next is called *gluing*.

By gluing, we mean that we merge two directed edges such that the two, which are glued together must be of opposite direction. The vertex, which marks the start from one directed edge will thus be identified with the vertex, where the other edge ends. We are now gluing the edges of our polygon pairwise in the following way:

Thus, we must identify the vertex $i_1$ with $i_5$, $i_2$ with $i_4$ (coming from the first gluing), and also $i_6$ with $i_{10}$, $i_6$ with $i_9$, $i_7$ with $i_8$, $i_7$ with $i_{10}$ and $i_8$ with $i_9$. This leads to the following
equalities $i_1 = i_5$, $i_2 = i_4$, $i_3 = i_7 = i_8 = i_9 = i_{10}$. These are exactly the equalities, obtained from our Wick coupling above and indeed, there is a one-to-one correspondence between Wick couplings and the gluing of the edges of a polygon pairwise. To see this, label the edge from $i_l$ to $i_{l+1}$ by $l$. Now glue the first edge to any other edge. Then proceed by choosing the next edge that has not been glued and glue it to one of the remaining edges. By continuing this, we find a Wick coupling, in which the first edge of each pairing marks the $p_i$’s, while the second one defines the $q_i$’s. Since the first edge can be glued to any of the $2k-1$ edges, the next edge therefore only has $2k-3$ possible edges to be glued with, until the last two edges must be glued together, we easily see, that there are exactly $(2k-1)!!$ possibilities of pairwise gluings. The number of ‘free parameters’ $f$ which contribute to the sum are now the number of vertices of the map, that was produced by gluing the edges of the polygon. In order to calculate this number, we use the Euler characteristic, knowing that our resulting map has 5 edges and that the polygon as well as the map has only one face. Therefore,

$$\chi = 2 - 2g = \#\text{vertices} - \#\text{edges} + \#\text{faces} = \#\text{vertices} - 5 + 1,$$

which yields

$$\#\text{vertices} = 5 + 1 - 2g.$$

More generally, we can say about a $2k$-sided polygon, that it has $k$ edges after gluing and as well one face, hence,

$$\#\text{vertices} = k + 1 - 2g.$$

Up to know, we have omitted to speak about the genus. As is seen from the Euler characteristic, the genus of the resulting map after gluing affects the contribution to the total sum, while we have to remember, that each contribution is $N^{f-k}$. The integrals of form $\mathbb{E}[\text{tr}(H^{2k})]$ can now be written in terms of a genus expansion. Defining

$$\epsilon_g(k) := \#(\text{gluings of a } 2k\text{-sided polygon resulting in a map of genus } g)$$

$$= \#(\text{labeled one-face map of genus } g \text{ with } k \text{ edges})$$

and observing, that the maximal genus is equal to $\lfloor \frac{k}{2} \rfloor$, we have established

$$\mathbb{E}[\text{tr}(H^{2k})] = \sum_{g=0}^{\lfloor \frac{k}{2} \rfloor} \epsilon_g(k)N^{k+1-2g-k} = \sum_{g=0}^{\lfloor \frac{k}{2} \rfloor} \epsilon_g(k)N^{1-2g}. \quad (4.16)$$

82
The result stated in (4.16) was first obtained by Harer and Zagier in their famous paper [28]. However, they counted different objects than one-face maps or gluings of polygons, but their combinatorial way of proceeding is just an alternative approach to map enumeration and leads to exactly the same counting problem, as we explain next.

**Alternative graphical interpretation**

When considering an alternative graphical formulation for Wick’s theorem, we have to think of an edge as two arrows pointing in opposite directions.

![Figure 4.6: An edge consisting of two arrows](image)

A vertex, where edges meet, is now represented by a star, such that the connected arrows are of the same direction.

![Figure 4.7: A vertex depicted by a star](image)

Gluing the edges, such that each of the corresponding arrows are joined with respect to their orientation, produces a so called *fat-graph.*
Figure 4.8: All possible fat-graphs with one star of degree 4

Consider now a star with 2\(k\) edges, labeled \(h_{i_1i_2}, h_{i_2i_3}, \ldots, h_{i_{2k}i_1}\), which induces that the arrows of edge \(h_{i_im}\) are labeled \(i_l\) and \(i_m\). By gluing two edges, say \(h_{i_1i_2}\) and \(h_{i_2i_m}\) we find that in order to obtain the orientation, we need \(i_1 = i_m\) and \(i_2 = i_l\). Thus, as above we recover all the Wick couplings. There is one vertex, \(k\) edges and only the number of faces is unknown but via Euler’s characteristic equal to \(1 - 2g + k\). This leads to the observation that \(\epsilon_g(k)\) in (4.18) can be replaced by

\[
\tilde{\epsilon}_g(k) := \# \{\text{fat-graphs of genus } g \text{ with one star of degree } 2k\}. 
\]

**Remark:** 1.) As was pointed out in [53] and [34] \(\epsilon_0(k) = \tilde{\epsilon}_0(k) = C_k = \frac{1}{k+1} \binom{2k}{k}\), where \(C_k\) are the well-known Catalan numbers.

Furthermore, since \(\epsilon_0(k)\) is the leading term in \(N\), the following holds:

\[
\lim_{N \to \infty} \frac{\mathbb{E} \left[ \text{tr}(H^k) \right]}{N} = C_k. \tag{4.17}
\]

In 1958, proving (4.17) was a crucial step for Wigner when deriving the semicircle law, see [50].

2.) An advantage of the second graphical interpretation of the Wick formula is the following generalization it allows, and which can also be found e.g. in [34]: For integrals of type \(\mathbb{E} \left[ \prod_{i=1}^{k} (\text{tr}(H^{p_i}))^{l_i} \right]\) it holds that

\[
\mathbb{E} \left[ \prod_{i=1}^{k} (\text{tr}(H^{p_i}))^{l_i} \right] = \sum_F N^{\# \text{faces} - \# \text{edges}},
\]
where the sum is taken over all fat-graphs $F$ (not necessarily connected!) consisting of $l_i$ vertices of degree $p_i$, $i = 1, \ldots, k$. By Euler’s characteristic, this can be reformulated in terms of an genus expansion,

$$
\mathbb{E} \left[ \prod_{i=1}^{k} \text{tr}(H^{k_i})^{l_i} \right] = \sum_F N^\# \text{faces} - \# \text{edges} \\
= \sum_g \# \{ \text{fat-graphs of genus } g \text{ with } l_i \text{ vertices of degree } k_i \} N^{2g - \sum_{i=1}^{k} l_i} 
$$

(4.18)

**Symmetric matrices**

When considering symmetric matrices $S = (s_{ij})_{1 \leq i,j \leq N}$, we will examine the relevant maps, that need to be counted for Wick’s theorem. It turns out that they are the graphs of the hermitian case up to some modifications. The measure on the space of symmetric $N \times N$ matrices, $\mathcal{S}^N(\mathbb{R})$, is

$$
\mu_s^N(dS) = \frac{1}{Z_s^N} \exp \left\{ -\frac{N}{4} \text{tr}(S^2) \right\} d^N S, 
$$

(4.19)

where $d^N S$ is the Lebesgue measure on $\mathcal{S}^N(\mathbb{R})$.

Furthermore, $\mu_s^N(dS)$ can be understood as a measure on $\mathbb{R}^{N(N+1)/2}$ being of form (4.11), where the inverse of the covariance matrix is a diagonal matrix with $N$ diagonal terms equal to $\frac{N}{2}$ and the other $\frac{N(N-1)}{2}$ diagonal terms being $N$,

$$
B = \begin{pmatrix}
\frac{N}{2} & & & \\
& \frac{N}{2} & & \\
& & \ldots & \\
& & & N
\end{pmatrix} 
\in \mathbb{R}^{\frac{N(N+1)}{2} \times \frac{N(N+1)}{2}}.
$$

Thus we find

$$
\mathbb{E} [s_{ij} s_{kl}] = \frac{1}{N} (\delta_{il} \delta_{jk} + \delta_{ik} \delta_{jl}).
$$

In contrast to (4.13), the additional term $\delta_{ik} \delta_{jl}$ appears. That exactly means, that we do not necessarily have to glue the directed edges in opposite direction in the one-face case. This means that the direction of the edges can be neglected and we can glue edges either way.
Concerning fat graphs, we also observe that the orientation of the arrows is meaningless, which means, that we can also allow for twisted edges. These unoriented fat-graphs are also referred to as *Moebius graphs*.

Figure 4.9: Some unoriented fat-graphs of genus \( g = 1 \) and with one star with degree 4

And all that was said in the hermitian setting now transfers to the symmetric case, when we move from directed graphs to undirected graphs.

For an algorithm, to obtain the number of the appearing graphs (either directed or undirected), see [43].

**A 'historical' integral**

Concluding the considerations concerning the one-matrix case, we will have a closer look at one integral of the ‘first hour’,

\[
Z_{V_t}^N = \mathbb{E} \left[ \exp \left\{ -N \text{tr}(tH^4) \right\} \right], \ t \in \mathbb{R},
\]

where \( V_t = tH^4 \) using the notation of (4.2) and (4.3). We will briefly give the heuristic here, as first given by [7], [5] and later spelled out in [53].

Since we only know how to integrate polynomials with the help of Wick’s formula, the obvious thing to do is to expand the exponential function,

\[
\exp \left\{ -N \text{tr}(tH^4) \right\} = \sum_{n=1}^{\infty} \frac{(-t)^n}{n!} N^n (\text{tr}(H^4))^n.
\]

When interchanging the sum and the expectation, we can only proceed formally, meaning that the following identity is to be understood in the sense that any derivative with respect
4.3 Excursus: Map Enumeration and Matrix Integrals

to $t$ of both sides match at $t = 0,$

$$Z_{V_t} = \sum_{n=1}^{\infty} \frac{(-t)^n}{n!} N^n E \left[(\text{tr}(H^4))^n\right].$$

When employing the Wick formula for

$$E[(\text{tr}(H^4))^n] = E[\sum h_{i_1j_1}h_{j_1k_1}h_{k_1i_1}h_{i_2j_2}h_{j_2k_2}, \ldots, h_{i_ni_n}],$$

we have $n$ vertices of degree 4, where the edges consist of two arrows and are labeled in the following manner.

![Diagram](image)

**Figure 4.10:** Configuration for $(\text{tr}(H^4))^n$

Gluing edges regarding the orientation of the arrows now produces a (Feynman) diagram, which is

- a graph with $n$ vertices of degree 4,
- with $4n/2 = 2n$ edges, with a prescribed cyclic order around each vertex,
- and where the gluing of edges yields chains of equalities, e.g. $i_1 = k_3 = \ldots = i_1, j_1 = j_3 = \ldots, j_1$ (if $h_{i_1j_1}$ is coupled to $h_{k_3j_3}$), which correspond to faces.

Hence, each diagram contributes $N^f \times N^{-2n}$ to the sum, where $f := \# \text{faces}.$ Due to

$$\chi = 2 - 2g = n - 2n + f \iff f = n + 2 - 2g$$

this factor becomes $N^{n+2-2g-2n} = N^{2-2g-n},$ which in turn gives

$$Z_{V_t} = \sum_{n=1}^{\infty} \frac{(-t)^n}{n!} \sum_g \# \{\text{diagram of genus } g \text{ with } n \text{ vertices of degree 4}\} N^{2-2g}. \quad (4.21)$$

We remark here, that diagrams do not need to be connected.

Taking the logarithm of an exponential generating functions as in (4.21) is known by combinatorists to mark the transition from unconnected diagrams to connected diagrams, see e.g.
Remember, that connected diagrams are also called maps, therefore

\[
\log Z_{V_t} = \sum_{n=1}^{\infty} \frac{(-t)^n}{n!} \sum_g \# \{ \text{maps of genus } g \text{ with } n \text{ vertices of degree } 4 \} \frac{N}{2}^{-2g}.
\]

Multiplying each side by \( \frac{N}{2}^{-2} \) and changing the order of summation, one obtains heuristically the genus expansion

\[
\frac{1}{N^2} \log Z_{V_t} = \sum_g \left( \frac{1}{N^2} \right)^g E_g(t),
\]

where

\[
E_g(t) = \sum_{n=1}^{\infty} \frac{(-t)^n}{n!} \{ \text{maps of genus } g \text{ with } n \text{ vertices of degree } 4 \}.
\]

In the genus expansion, \( \frac{1}{N^2} \) plays the role of the formal parameter.

In [20], this expansion was put on solid mathematical grounds, in the sense that summing the right hand side of (4.22) up to genus \( g = k \), establishes this equality up to an error of order \( N^{-2k-2} \), where potentials of type \( V_t = \sum_{i=1}^{\nu} t_i H_i \), with \( t \in \mathbb{R}^\nu \) being in the neighborhood of 0, were considered. Nevertheless, it is also shown, that \( E_g(t) \) is analytic around 0. The extension to the multi matrix case, was done in a mathematically rigorous way in [25] and [39].

### 4.3.3 Multi-matrix case

Now we turn to \( m \) independent hermitian matrices, \( H_1, \ldots, H_m \).

**Stars, maps and gluing**

The presence of more than one matrix corresponds to the appearance of \( m \) different colors in the graphical setting. The colors are denoted by \( l \), where \( l \in \{1, \ldots, m\} \). The central object, that is considered at the beginning of the chapter and in our results, is now a non-commutative monomial in \( m \) indeterminants,

\[
q_l(H_1, \ldots, H_m) = H_{l_1} H_{l_2} \cdots H_{l_{k_l}}, \quad \text{with } k_l \in \mathbb{N}, l_j \in \{1, \ldots, m\}, 1 \leq j \leq k_l,
\]

where the non-commutativity stems from the matrix multiplication. For monomials like this, we define vertices, that are of type \( q_l \) as follows.

**Definition 4.11** We say that a vertex is of type \( q_l \), if it has \( k_l \) colored half-edges, one marked half edge and an orientation, such that the marked half-edge is of color \( l_1 \), the next one with respect to the orientation is of color \( l_2 \) and so forth, until the last half-edge is colored with \( l_{k_l} \).
Thus, we obtain a bijection between monomials and stars. Moreover, the graphical interpretation of the involution \(^\dagger\) as defined in (4.4) is quite simple. Just shift the marker of the first half-edge towards the next neighboring half-edge against the orientation of the vertex and afterwards reverse the orientation of the vertex. In order to construct maps out of these stars (or simply one of them), we need to glue them according to the following rule:

- Half-edges, which are glued together, must be of the same color,
- The resulting edges are not allowed to intersect,
- Half-edges are glued with respect to their orientation, if the half-edges are thought to be composed of two opposite arrows, as was the case when considering fat-graphs in the hermitian case above.

Following the orientation of one edge, we describe a circle, which surrounds one face.

**Remark:**
Due to this cyclic orientation at the end of the half-edges at any vertex, there exists a unique way of embedding the graph into the surface, a result which might go back to Hamilton, see comments in [53].
As before we count maps up to homeomorphism of the oriented surface.
Unperturbed case

When considering the unperturbed case of the $m$-dimensional GUE, Voiculescu was able to extend Wigner’s result (4.17) to multi matrices in [48]. For a monomial $q_l$ define $C_0(q_l)$ by

$$C_0(q_l) := \#\{\text{planar maps with one star of type } q_l\}.$$  

Under certain moment conditions he showed for any $l_j \in \{1, \ldots, m\}, 1 \leq j \leq k,

$$\lim_{N \to \infty} \frac{1}{N} \text{tr}(H_{l_1}H_{l_2} \cdots H_{l_k}) \mu^N(d\,H) = C_0(H_{l_1}H_{l_2} \cdots H_{l_k}),$$  \hspace{1cm}(4.23)

where $\mu^N(d\,H)$ as in (4.1) and furthermore,

$$\lim_{N \to \infty} \frac{1}{N} \text{tr}(H_{l_1}H_{l_2} \cdots H_{l_k}) = C_0(H_{l_1}H_{l_2} \cdots H_{l_k}) \text{ almost surely.}$$

Obviously, these results can be adjusted to hold for any polynomial $p$ in $m$ non-commutative indeterminants.

When quantities like $\text{tr}(q_l)$ are examined, the position of the marked edge in $q_l$ is not of utmost importance for its value, since it is invariant under cyclic permutation with respect to the given orientation, but for counting purposes, these two stars are distinct and all maps resulting from these two are counted.

Perturbed case

The next step was to exchange the measure $\mu^N(d\,H)$ by the perturbed measure $\mu^N_{V_t}(d\,H)$ as in (4.2), where the potential $V_t$ is allowed to be of form $V_t = \sum_{i=1}^n t_i q_i(H_1, \ldots, H_m)$. The first
results were obtained in [24] and [25], where similar to (4.18), (4.22) and (4.23) the leading asymptotic for either the expectation of a monomial or the logarithm of the partition function was deduced. The authors obtained an expansion in terms of the genus up to a genus of 2, while in [39] the genus expansion was finally given for any arbitrary genus along with the order of the appropriate error term, when the potential $V_t$ is $c$-convex and the parameter $t$ small enough.

The exact formulation of (4.23) with the perturbed measure now reads:

**Theorem 4.12 (Theorem 1.2 in [39])** Let $V_t = \sum_{i=1}^{n} t_i q_i (H_1, \ldots, H_m)$ and $c > 0$. For all $g \in \mathbb{N}$, there exists $\eta_g > 0$, such that for all $t \in B_{\eta_g, c}$ and for all monomials $P$

$$E_{\nu_{V_t}} \left[ \frac{1}{N} \text{tr}(P) \right] = \mathcal{C}_0^g(P) + \frac{1}{N^2} \mathcal{C}_1^g(P) + \cdots + \frac{1}{N^{2g}} \mathcal{C}_g^g(P) + o \left( \frac{1}{N^{2g}} \right),$$

where $C_i^g(P)$ is the generating function for maps of genus $g$ with some fixed vertices,

$$C_i^g(P) = \sum_{k \in \mathbb{N}} \frac{(-t)^k}{k!} \mathcal{C}_i^k(P)$$

with $k! = \prod_i k_i!$, $(-t)^k = \prod_i (-t_i)^k$ and $\mathcal{C}_i^k(P)$ the number of maps of genus $g$ with $k_i$ vertices of type $q_i$ and one vertex of type $P$.

The climax and last point of this excursus will be the genus expansion of the logarithm of the partition function in the perturbed $m$-dimensional GUE (cf. (4.22) in the unperturbed situation).

**Theorem 4.13 (Theorem 1.1 in [39])** Let $V_t = \sum_{i=1}^{n} t_i q_i (H_1, \ldots, H_m)$ and $c > 0$. For all $g \in \mathbb{N}$, there exists $\eta_g > 0$, such that for all $t \in B_{\eta_g, c}$, $Z_{V_t}^N$ has the following expansion

$$F_{V_t}^N := \frac{1}{N^2} \log Z_{V_t}^N = F_0^g(t) + \frac{1}{N^2} F_1^g(t) + \cdots + \frac{1}{N^{2g}} F_g^g(t) + o \left( \frac{1}{N^{2g}} \right),$$

(4.24)

where $F_g^g$ is the generating function for maps of genus $g$ associated with $V_t$,

$$F_g^g(t) = \sum_{k \in \mathbb{N}} \frac{(-t)^k}{k!} \mathcal{C}_g^k$$

and $k! = \prod_i k_i!$, $(-t)^k = \prod_i (-t_i)^k$ and $\mathcal{C}_g^k$ is the number of maps on a surface of genus $g$ with $k_i$ vertices of type $q_i$.

Now it is obvious, why such an expansion is called genus expansion and why the leading term $F_0^g(t)$ was called the planar approximation. For our purposes, we will only apply that theorem for $g = 1$.

This theorem will play an important part in the proof of our results, but before we come to that, we collect some of the properties of the rate function.
4 FLUCTUATIONS IN A MULTI-MATRIX MODEL

4.4 Rate function

Since the rate function is of form

\[ I(x) = \frac{x^2}{2} \left( \frac{\partial^2}{\partial t^2} F^0(t) \right)^{-1}, \]

we see that the derivation of the function \( F^0(t) \) appears and in this section we will give one result by Maurel-Segala [39], how this derivation can be computed and afterwards, we give more details on the graphical interpretation.

The basic statement of the theorem is, that we are able to derive the expansion (4.24) on a term by term basis. In fact, this means that the asymptotics of much more observables are available.

For \( j = (j_1, \ldots, j_n) \in \mathbb{N}^n \), we introduce the operator of derivation

\[ D_j = \frac{\partial^\sum_{i,j_i} \partial t_1 \cdots \partial t_n}{\partial t_1 \cdots \partial t_n}. \]

**Theorem 4.14 (Theorem 1.3 in [39])** Let \( V_t = \sum_{i=1}^n t_i q_i(H_1, \ldots, H_m) \) and \( c > 0 \). For all \( j = (j_1, \ldots, j_n) \in \mathbb{N}^n \), for all \( g \in \mathbb{N} \), there exists \( \eta_g > 0 \), such that for all \( t \in B_{\eta_g,c} \),

\[ D_j F_{V_t}^N = D_j F^0(t) + \cdots + \frac{1}{N^{2g}} D_j F^g(t) + o \left( \frac{1}{N^{2g}} \right), \quad (4.25) \]

where \( D_j F^g(t) \) is the generating function for maps of genus \( g \) associated with \( V_t \) with some fixed vertices:

\[ D_j F^g(t) = \sum_{k \in \mathbb{N}^n} \frac{(-t)^k}{k!} C_g^{k+j}, \]

where \( C_g^k \) is again the number of maps on a surface of genus \( g \) with \( k_i \) vertices of type \( q_i \).

This theorem will also be involved in the proofs, that are to follow.

In order to see that the derivative adds some fixed points to the maps, that we count, consider as a simple example \( D_j \) with \( j = (2, 0, \ldots, 0) \in \mathbb{N}^n \), where we refer to [39, Thm. 7.4 and Le. 4.4] for more details,

\[ D_j F^g(t) = \frac{\partial^2}{\partial t_1^2} F^g(t) = \frac{\partial^2}{\partial t_1^2} \sum_{k \in \mathbb{N}^n} \frac{(-t)^k}{k!} C_g^k = \sum_{k \in \mathbb{N}^n} \frac{\partial^2}{\partial t_1^2} \frac{(-t)^k}{k!} C_g^k \]

\[ = \sum_{k \in \mathbb{N}^n} \frac{(-1)^2 (-t_1)^{k_1-2}}{(k_1 - 2)!} \prod_{i \leq 2} \frac{(-t_i)^{k_i}}{k_i!} C_g^k = \sum_{k \in \mathbb{N}^n} \frac{(-t)^k}{k!} C_g^{k+2e_1} \]

\[ = \sum_{k \in \mathbb{N}^n} \frac{(-t)^k}{k!} C_g^{k+j}, \quad (4.26) \]
where $e_1 = (1, 0, \ldots, 0) \in \mathbb{N}^n$.

That we could exchange the derivative and the sum is due to the analyticity of $F^g(t)$ for each $g$ when $t$ is in the vicinity of 0, see reference above.

Thus, we find for example for $V_t = t_1 H_1^2 H_2^2 H_4^2 + t_2 H_2 H_1 H_2$ and $j = (2, 0)$, that $D_j F^0$ counts all planar maps with $k_1 + 2$ vertices of type $q_1 = H_1^2 H_2^2 H_4^2$ and $k_2$ vertices of type $q_2 = H_2 H_1 H_2$.

Moreover, we have seen that the rate function $I$ from Theorem 4.5 is the same as

$$I(x) = \frac{x^2}{2} \left( \sum_{k \in \mathbb{N}^n} \frac{(-t)^k}{k!} \sigma^{k+2e_1}_0 \right)^{-1}.$$  

4.5 Proofs

Now we turn to the proof of our results.

Out of these two theorems cited above, we deduce an asymptotic expansion for the moment generating function of

$$\varphi_l := \varphi_l(H) := \text{tr}(q_l(H_1, \ldots, H_m))$$

and

$$\varphi_l := \varphi_l(H) := \varphi_l(H) - \mathbb{E}[\varphi_l(H)].$$

Remember, that $\mathbb{E}$ denotes the expectation with respect to the probability measure $\mu_{V_t}^N$ and let $e_l$ denote the $l$-th unit vector $(0, \ldots, 0, 1, 0, \ldots, 0) \in \mathbb{N}^n$.

**Lemma 4.15** Let $t \in \mathbb{R}^n$ and $c > 0$. Then there exists $\eta > 0$, such that for

$$V_t = \sum_{i=1}^n t_i q_i(H_1, \ldots, H_m), \quad n \in \mathbb{N}, \text{ with } t \text{ in } B_{n,c} \cap \mathbb{R}, \text{ and any fixed } l \in \{1, \ldots, n\} \text{ we have for any } s \in \mathbb{R} \text{ that}$$

$$\mathbb{E}\left(\exp(s \varphi_l)\right) = \exp \left\{ N^2 \left[ F^0 \left( t - \frac{s}{N} e_l \right) - F^0(t) + \frac{s}{N} \frac{\partial}{\partial t_l} F^0(t) \right] + F^1 \left( t - \frac{s}{N} e_l \right) - F^1(t) + \frac{s}{N} \frac{\partial}{\partial t_l} F^1(t) + o(1) \right\}. $$

In the non-centered case, we find

$$\mathbb{E}\left(\exp(s \varphi_l)\right) = \exp \left\{ N^2 \left[ F^0 \left( t - \frac{s}{N} e_l \right) - F^0(t) \right] + F^1 \left( t - \frac{s}{N} e_l \right) - F^1(t) + o(1) \right\}.$$ 

**Proof:**

Consider for any $s \in \mathbb{R}$ and $l \in \{1, \ldots, n\}$ the potential $V_{t-\frac{s}{N} e_l}$. We easily see that

$$V_{t-\frac{s}{N} e_l} = \sum_{i=1}^n t_i q_i(H_1, \ldots, H_m) + \left( t_l - \frac{s}{N} \right) q_l(H_1, \ldots, H_m) = V_t - \frac{s}{N} q_l(H_1, \ldots, H_m).$$
We abbreviate \( \tilde{V} := \tilde{V}(t, s, l) := V_{t - \frac{s}{N}e_l} \) and obtain
\[
Z^N_{\tilde{V}} = \int_{\mathcal{H}_N(C)^m} \exp \left( -N \left( \text{tr}(\tilde{V}(H)) \right) \right) \mu^N(dH) \\
= \int_{\mathcal{H}_N(C)^m} \exp\left( -N(\text{tr}(V_i)) \right) \exp(s \text{ tr}(q_l(H))) \mu^N(dH) \\
= \mathbb{E}[\exp(s \text{ tr}(q_l(H)))] Z^N_{V_i}.
\]
Hence we get
\[
\mathbb{E}[\exp(s \text{ tr}(q_l(H)))] = \frac{Z^N_{\tilde{V}}}{Z^N_{V_i}}. \tag{4.27}
\]

We want to apply Theorem 4.13 for both terms on the right hand side of the last equality. Fix \( c > 0 \). Then by Theorem 4.13 there exists a \( \eta := \eta(c) > 0 \) such that the expansion (4.24) holds true for \( Z^N_{V_i} \) for all \( t \in B_{\eta, c} \).

Now for fixed \( s \in \mathbb{R} \) we choose \( N \) sufficiently large such that \( |s/N| < \varepsilon \) for a \( \varepsilon > 0 \). We abbreviate \( \tilde{t} := t - \frac{s}{N}e_l \). For any \( t \in B_{\eta, c} \) the polynomial \( V_i \) is \( c \)-convex, thus for \( N \) sufficiently large we can find a \( c' > 0 \) such that \( V_i \) is \( c' \)-convex. For this \( c' > 0 \) we can find a \( \eta' := \eta'(c') > 0 \) such that (4.24) holds true for any \( t \in B_{\eta', c'} \), where \( V_i \) is \( c' \)-convex. Now we choose \( \varepsilon < \eta' \) and \( \eta'' := \min(\eta' - \varepsilon, \eta) \).

Since \( \eta'' \leq \eta \), the condition \( |t| \leq \eta'' \) induces \( |t| \leq \eta \) and therefore, the relevant set to consider is \( B_{\eta'', c} \). Summarizing, for fixed \( s \in \mathbb{R} \) and \( N \) sufficiently large, for any \( t \in B_{\eta', c} \) we can apply Theorem 4.13 for both terms in (4.27).

Remark that we apply Theorem 4.13 taking \( g = 1 \).

Now we obtain the following asymptotic expansion in the non-centered case:
\[
\frac{Z^N_{\tilde{V}}}{Z^N_{V_i}} = \exp \left\{ N^2 F^0 \left( t - \frac{s}{N}e_l \right) + F^1 \left( t - \frac{s}{N}e_l \right) + o(1) \right\} \\
\times \exp \left\{ -N^2 F^0(t) - F^1(t) + o(1) \right\} \\
= \exp \left( N^2 \left[ F^0 \left( t - \frac{s}{N}e_l \right) - F^0(t) \right] + F^1 \left( t - \frac{s}{N}e_l \right) - F^1(t) + o(1) \right).}
\]

In the centered case, we also have to consider the contribution of \( \mathbb{E}[\exp(-s \mathbb{E}[q_l(H)])] \) to the asymptotic expansion.
Therefore, we calculate,
\[
\frac{\partial}{\partial t} \log Z_N^V = \frac{1}{Z_N^V} \frac{\partial}{\partial t} \int_{\mathcal{H}_N(C)^m} \exp \left( -N \left( \text{tr}(V(H)) \right) \right) \mu^N(dH) \tag{4.28}
\]
\[
= \frac{1}{Z_N^V} \int_{\mathcal{H}_N(C)^m} \frac{\partial}{\partial t} \exp \left( -N \left( \text{tr}(V(H)) \right) \right) \mu^N(dH) \tag{4.29}
\]
\[
= \frac{1}{Z_N^V} \int_{\mathcal{H}_N(C)^m} -N \text{tr}(q_l(H_1, \ldots, H_m)) \exp \left( -N \left( \text{tr}(V(H)) \right) \right) \mu^N(dH)
\]
\[
= \mathbb{E} \left[ -N \text{tr}(q_l) \right] = -N \mathbb{E} \left[ \text{tr}(q_l) \right].
\]

In order to interchange the differentiation and integration in (4.28), we applied [19, Thm. 5.7], which only demands that the integral appearing in (4.29) exists. Theorem 4.12 establishes that existence and hence, we find
\[
s \mathbb{E} \left[ \text{tr}(q_l) \right] = -\frac{s}{N} \frac{\partial}{\partial t} \log Z_N^V
\]
\[
\iff \exp \left( -s \mathbb{E} \left[ \text{tr}(q_l) \right] \right) = \exp \left( \frac{s}{N} \frac{\partial}{\partial t} \log Z_N^V \right). \tag{4.30}
\]

Finally, (4.27), (4.30) and Theorem 4.14 yield
\[
\mathbb{E} \left( \exp \left( s \overline{\phi}_l \right) \right) = \frac{Z_N^V}{Z_N^V} \times \exp \left( \frac{s}{N} \frac{\partial}{\partial t} \log Z_N^V \right)
\]
\[
= \exp \left\{ N^2 \left[ F^0 \left( t - \frac{s}{N} e_l \right) - F^0(t) \right] + F^1 \left( t - \frac{s}{N} e_l \right) - F^1(t) + o(1) \right\}
\]
\[
\times \exp \left\{ \frac{s}{N} \left( N^2 \frac{\partial}{\partial t} F^0(t) + \frac{\partial}{\partial t} F^1(t) + o(1) \right) \right\},
\]
and therefore we obtain
\[
\mathbb{E} \left( \exp \left( s \overline{\phi}_l \right) \right) = \exp \left\{ N^2 \left[ F^0 \left( t - \frac{s}{N} e_l \right) - F^0(t) + \frac{s}{N} \frac{\partial}{\partial t} F^0(t) \right]
\]
\[
+ F^1 \left( t - \frac{s}{N} e_l \right) - F^1(t) + \frac{s}{N} \frac{\partial}{\partial t} F^1(t) + o(1) \right\}.
\]

This lemma is now the crucial ingredient for the proofs of our main results.

**Proof of the central limit theorem, Theorem 4.6:**

By Taylors theorem, we have
\[
F^i \left( t - \frac{s}{N} e_l \right) - F^i(t) + \frac{s}{N} \frac{\partial}{\partial t} F^i(t) = \frac{s^2}{2N^2} \frac{\partial^2}{\partial t^2} F^i(e_N^i), \quad i = 0, 1, \tag{4.31}
\]
where
\[ \xi^l_N \in \left( t \wedge \left( t - \frac{s}{N} e_l \right), t \vee \left( t - \frac{s}{N} e_l \right) \right). \]

Combining this with Lemma 4.15, we find the limit of the moment generating function of \( \varphi^l \) to be that of a centered random variable having a normal distribution with variance \( \frac{\partial^2}{\partial t^2} F^0(t) \):

\[
\lim_{N \to \infty} \mathbb{E} \left[ \exp(s \varphi^l) \right] = \lim_{N \to \infty} \exp \left( \frac{s^2}{2} \frac{\partial^2}{\partial t^2} F^0(\xi^l_N) + \frac{s^2}{2N^2} \frac{\partial^2}{\partial t^2} F^1(\xi^l_N) + o(1) \right) \\
= \exp \left( \frac{s^2}{2} \frac{\partial^2}{\partial t^2} F^0(t) \right),
\]

since \( \xi^l_N \to t \), for \( N \to \infty \), and \( F^0, F^1 \) being differentiable functions of any order around the origin by Theorem 4.14 (they are even analytic functions, see [39, Le. 4.4]).

By Levy’s continuity theorem the CLT is established.

**Proof of the moderate deviation principle, Theorem 4.5:**

In order to apply the Gärtner-Ellis theorem (cf. Theorem 1.9 or Theorem 2.3.6 in [11]), we have to calculate the limit of the properly scaled cumulant generating function of the random variable \( \frac{1}{N^\gamma} \varphi^l \), where \( 0 < \gamma < 1 \) and let \( s \in \mathbb{R} \),

\[
\lim_{N \to \infty} \frac{1}{N^{2\gamma}} \log \mathbb{E} \left[ \exp \left( s N^{2\gamma} \frac{1}{N^\gamma} \varphi^l \right) \right] = \lim_{N \to \infty} \frac{1}{N^{2\gamma}} \log \mathbb{E} \left[ \exp \left( s N^\gamma \varphi^l \right) \right].
\]

First, we want to apply Lemma 4.15, but instead of a fixed number \( s \in \mathbb{R} \), we have now the term \( sN^\gamma \) in the exponent. Going carefully through the proof of the Lemma shows, that the the result still holds, when \( s \) is replaced by \( sN^\gamma \), since the most important property is that \( \frac{N^\gamma}{N} \) tends to 0 because of \( 0 < \gamma < 1 \).

Hence,

\[
\lim_{N \to \infty} \frac{1}{N^{2\gamma}} \log \mathbb{E} \left[ \exp \left( s N^\gamma \varphi^l \right) \right] \\
= \lim_{N \to \infty} \frac{1}{N^{2\gamma}} \log \left( \exp \left( N^2 \left[ F^0 \left( t - \frac{sN^\gamma}{N} e_l \right) - F^0(t) + \frac{sN^\gamma}{N} \frac{\partial}{\partial t} F^0(t) \right] \right) + F^1 \left( t - \frac{sN^\gamma}{N} e_l \right) - F^1(t) + \frac{sN^\gamma}{N} \frac{\partial}{\partial t} F^1(t) + o(1) \right) \\
\]
4.6 GENERALIZATION

The next step is to employ Taylor’s theorem (4.31),

\[
\lim_{N \to \infty} \frac{1}{N^{2\gamma}} \left\{ N^2 \left[ F^0 \left( t - \frac{sN^\gamma}{N} \epsilon_l \right) - F^0(t) + \frac{sN^\gamma}{N} \frac{\partial}{\partial t} F^0(t) \right] \right.
\]

\[
+ F^1 \left( t - \frac{sN^\gamma}{N} \epsilon_l \right) - F^1(t) + \frac{sN^\gamma}{N} \frac{\partial}{\partial t} F^1(t) + o(1) \bigg\} = \lim_{N \to \infty} \frac{1}{N^{2\gamma}} \left[ N^2 \left( \frac{sN^\gamma}{N} \right)^2 \frac{\partial^2}{\partial t^2} F^0(\xi^0_N) + \frac{1}{2} \left( \frac{sN^\gamma}{N} \right)^2 \frac{\partial^2}{\partial t^2} F^1(\xi^1_N) + o(1) \right].
\]

Finally, we use the fact of \( F^0 \) being differentiable of any order around the origin along with \( \xi^0_N \in \left( t \wedge (t - \frac{sN^\gamma}{N} \epsilon_l), t \vee (t - \frac{sN^\gamma}{N} \epsilon_l) \right) \), which yields \( \xi^0_N \to t \), for \( N \to \infty \), since \( \gamma \in (0, 1) \).

Thus we get,

\[
\lim_{N \to \infty} \frac{1}{N^{2\gamma}} \log E \left[ \exp \left( s N^\gamma \varphi_l \right) \right] = \frac{s^2}{2} \frac{\partial^2}{\partial t^2} F^0(t). \tag{4.32}
\]

In order for the Gärtner-Ellis theorem to yield the full MDP, we have to examine, whether the function \( \frac{s^2}{2} \frac{\partial^2}{\partial t^2} F^0(t) \) fullfills the three conditions of Theorem 1.9. The conditions are satisfied, since

\[
\mathcal{D} := \left\{ s \in \mathbb{R} \left| \frac{s^2}{2} \frac{\partial^2}{\partial t^2} F^0(t) < \infty \right. \right\} \neq \emptyset,
\]

\( \frac{s^2}{2} \frac{\partial^2}{\partial t^2} F^0(t) \) is differentiable in \( \mathcal{D}^o \), thus continuous and finally, for every sequence \( (s_n) \) with \( \lim_{n \to \infty} s_n = \infty \), we find that \( s_n \frac{\partial^2}{\partial t^2} F^0(t) \to \infty \) for \( n \to \infty \).

Now we can apply the Gärtner-Ellis theorem and finish the proof by calculating the Legendre-Fenchel transform of \( \frac{s^2}{2} \frac{\partial^2}{\partial t^2} F^0(t) \), which is

\[
I(x) = \sup_{s \in \mathbb{R}} \left\{ sx - \frac{s^2}{2} \frac{\partial^2}{\partial t^2} F^0(t) \right\} = \frac{x^2}{2} \left( \frac{\partial}{\partial t^2} F^0(t) \right)^{-1}.
\]

\( \square \)

4.6 Generalization

Theorems 4.6 and 4.5 can be generalized to hold for traces of polynomials,

\[
P := \sum_{i=1}^{n} \alpha_i \text{tr}(q_i),
\]

97
where $\alpha_i \in \mathbb{R}, i = 1, \ldots, n$ and $q_i$ non-commutative monomial appearing in the potential function $V_i$. We will briefly state the crucial steps of the calculations and only look at
\[ \psi = \alpha_1 \text{tr}(q_i) + \alpha_2 \text{tr}(q_i), \]
where $l_1, l_2 \in \{1, \ldots, n\}$, $l_1 \neq l_2$, $\alpha_1, \alpha_2 \in \mathbb{R}$, and
\[ \overline{\psi} = \alpha_1 \text{tr}(q_i) + \alpha_2 \text{tr}(q_i) - \mathbb{E}[\alpha_1 \text{tr}(q_i) + \alpha_2 \text{tr}(q_i)] \]
respectively. In order to be able to apply Theorem 4.13 and a Taylor expansion in two variables, we argue as in the proof of Lemma 4.15 that for $c > 0$ we can find a $\eta = \eta(c) > 0$ such that $t - \kappa \frac{s}{N}(\alpha_1 e_i + \alpha_2 e_i) \in B_{\eta,c}$ for every $\kappa \in [0, 1]$ and $N$ sufficiently large. With the new potential
\[ V_{i- \frac{s}{N}(\alpha_1 e_i + \alpha_2 e_i)} = V_i - \frac{s}{N}(\alpha_1 q_i + \alpha_2 q_i), \]
we find that the free energy to be
\[ Z^{N}_{V_{i- \frac{s}{N}(\alpha_1 e_i + \alpha_2 e_i)}} = \mathbb{E}[e^{s(\alpha_1 \text{tr}q_i + \alpha_2 \text{tr}q_i)}] Z^{N}_{V_i} \]
and therefore we get
\[ \mathbb{E}[e^{s(\alpha_1 \text{tr}q_i + \alpha_2 \text{tr}q_i)}] = Z^{N}_{V_{i- \frac{s}{N}(\alpha_1 e_i + \alpha_2 e_i)}} (Z^{N}_{V_i})^{-1}. \quad (4.33) \]
Centering $\psi$, we need to calculate
\[ \mathbb{E}[\alpha_1 \text{tr}q_i + \alpha_2 \text{tr}q_i]. \]
Since we have
\[ \frac{\partial}{\partial u} V_{i- \frac{s}{N}(\alpha_1 e_i + \alpha_2 e_i)} = -\frac{\alpha_1}{N} \frac{\partial}{\partial t_1} V_{i- \frac{s}{N}(\alpha_1 e_i + \alpha_2 e_i)} - \frac{\alpha_2}{N} \frac{\partial}{\partial t_2} V_{i- \frac{s}{N}(\alpha_1 e_i + \alpha_2 e_i)}, \]
we find
\[ \mathbb{E}[\alpha_1 \text{tr}q_i + \alpha_2 \text{tr}q_i] = \frac{\partial}{\partial u} \log \mathbb{E}[e^{u(\alpha_1 \text{tr}q_i + \alpha_2 \text{tr}q_i)}] \bigg|_{u=0} \]
\[ = \frac{\partial}{\partial u} \log Z^{N}_{V_{i- \frac{s}{N}(\alpha_1 e_i + \alpha_2 e_i)}} \bigg|_{u=0} \]
\[ = -\frac{1}{N} \left( \alpha_1 \frac{\partial}{\partial t_1} \log Z^{N}_{V_i} + \alpha_2 \frac{\partial}{\partial t_2} \log Z^{N}_{V_i} \right). \quad (4.34) \]
The moment generating function now turns out to be (via Theorem 4.13)
\[ \mathbb{E}[e^{\overline{\psi}}] \overset{(4.33)}{=} \frac{Z^{N}_{V_{i- \frac{s}{N}(\alpha_1 e_i + \alpha_2 e_i)}}}{Z^{N}_{V_i}} \exp \left\{ \frac{s}{N} \sum_{i=1}^{2} \alpha_i \frac{\partial}{\partial t_i} \log Z^{N}_{V_i} \right\} \]
\[ \overset{(4.34)}{=} \exp \left\{ N^2 \left( F^0 \left( t - \frac{s}{N} (\alpha_1 e_i + \alpha_2 e_i) \right) - F^0(t) \right) + \frac{s}{N} \sum_{i=1}^{2} \alpha_i \frac{\partial}{\partial t_i} F^0(t) \right) \]
\[ + F^1 \left( t - \frac{s}{N} (\alpha_1 e_i + \alpha_2 e_i) \right) - F^1(t) + \frac{s}{N} \sum_{i=1}^{2} \alpha_i \frac{\partial}{\partial t_i} F^1(t) + o(1) \right\}. \]
Hence we obtain

\[ \mathbb{E} \left[ e^{s\bar{\psi}} \right] = \exp \left\{ \frac{s^2}{2} \sum_{i,j=1}^{2} \alpha_i \alpha_j \frac{\partial^2}{\partial t_i \partial t_j} F^0(\xi_N) + O \left( \frac{1}{N^2} \right) \right\}, \]  

(4.35)

and therefore

\[ \lim_{N \to \infty} \mathbb{E} \left[ e^{s\bar{\psi}} \right] = \exp \left\{ \frac{s^2}{2} \sum_{i,j=1}^{2} \alpha_i \alpha_j \frac{\partial^2}{\partial t_i \partial t_j} F^0(t) \right\}. \]

Here, we used the Taylor expansion in two variables and

\[ \xi_N = t - \kappa \frac{s}{N} (\alpha_1 e_{t_1} + \alpha_2 e_{t_2}), \] for a \( \kappa \in (0, 1) \).

Having thus obtained the CLT for \( \bar{\psi} \) under \( \mu_N^{\mathbb{V}} \), we can use the expansion (4.35) to obtain the MDP for \( \frac{1}{N^{2\gamma}} \bar{\psi} \) under \( \mu_N^{\mathbb{V}} \), where \( \gamma \in (0, 1) \), with speed \( N^{2\gamma} \) and rate function

\[ I(x) = \frac{x^2}{2} \left( \sum_{i,j=1}^{2} \alpha_i \alpha_j \frac{\partial^2}{\partial t_i \partial t_j} F^0(t) \right)^{-1}. \]

via the Gärtner-Ellis approach:

\[ \lim_{N \to \infty} \frac{1}{N^{2\gamma}} \log \mathbb{E} \left[ \exp \left\{ s N^{\gamma} \bar{\psi} \right\} \right] = \lim_{N \to \infty} \frac{1}{N^{2\gamma}} \left( \frac{s N^{\gamma}}{N} \right)^2 \left[ \sum_{i,j=1}^{2} \alpha_i \alpha_j \frac{\partial^2}{\partial t_i \partial t_j} F^0(\xi_N) + o(1) \right] \]

\[ = \frac{s^2}{2} \sum_{i,j=1}^{2} \alpha_i \alpha_j \frac{\partial^2}{\partial t_i \partial t_j} F^0(t), \]

since \( \xi_N = t - \kappa \frac{N^{s\gamma}}{N} (\alpha_1 e_{t_1} + \alpha_2 e_{t_2}), \) with \( \kappa \in (0, 1) \) and \( F^0 \) being differentiable of any order for \( t \) sufficiently small.

As above, it holds that \( \frac{s^2}{2} \sum_{i,j=1}^{2} \alpha_i \alpha_j \frac{\partial^2}{\partial t_i \partial t_j} F^0(t) \) is finite for \( s \in \mathbb{R} \), differentiable in \( s \) and goes to \( \infty \) for \( s \to \infty \). Thus all the conditions of Theorem 1.9 are met and the MDP on the real line for \( \bar{\psi} \) is established.

### 4.7 Comparisons in the one-matrix model

In this section, we will restrict our attention to the one-matrix model. In that case, the expansion of \( \log Z_N^{\mathbb{V}_t} \) as in (4.24) for \( m = 1 \) coincides with the expansion of \( \log Z_N^{\mathbb{V}_t} \) obtained by Ercolani and McLaughlin in [20]. While in [20] it was allowed for potentials of type \( \sum_{i=1}^{\nu} t_i H^i \), these are exactly the potentials \( \mathbb{V}_t = \sum_{i=1}^{n} t_i q_i(H) \), since the monomials \( q_i(H) \) are of form \( H^k \), for some \( k \in \mathbb{N}_0 \), and we can take \( \nu = n \) and \( q_i(H) = H^i \). But the domains of the
corresponding parameter $t$ appearing in the potential differ and our comparison will focus on these different domains.  
So far, our results relied on Theorem 4.13 and therefore, we could only allow for $c$-convex potential functions $V_t$ and allow our parameter $t$ to be out of 

$$B_{\eta,c} \cap \mathbb{R} = \{ t \in \mathbb{R}^n \mid |t| \leq \eta, V_t \text{ is } c\text{-convex} \}.$$  

Now, we will compare $t$ and the conditions on $V_t$, which are of the above type, with those coming from the expansion from Ercolani and McLaughlin. Their expansion holds for $t \in \mathbb{T}(T, \gamma)$, for a $T > 0$ and a $\gamma > 0$, where 

$$\mathbb{T}(T, \gamma) = \left\{ t \in \mathbb{R}^n \mid |t| \leq T, t_n > \gamma \sum_{j=1}^{n-1} |t_j| \right\}.$$  

Thus, we still focus on an neighborhood of 0, but the $c$-convexity is replaced by the condition $t_n > \gamma \sum_{j=1}^{n-1} |t_j|$. Our analytic comparison cannot give an all-embracing treatment of all possible choices of parameters, but we claim the following:

**Claim 1:** If $n$ is even and $T > 0$ and $\gamma > 1$ are such that $T < \frac{1}{n(n-1)^2(1+\gamma)}$, then it holds that 

$$\mathbb{T}(T, \gamma) \subset B_{T,c(T,\gamma)}$$  

for some $c(T, \gamma) > 0$.

**Claim 2:** If $n$ is uneven, for $V_t$ we can find $T > 0$ and $\gamma > 0$, such that $t \in \mathbb{T}(T, \gamma)$, whereas $\forall \eta > 0$ and $\forall c > 0$ we find $t \notin B_{\eta,c}$.

Before proving our claims, let us remark, what this means with respect to the asymptotic expansion of $\log Z_N^{V_t}$. In the situation of claim 1, this expansion does not hold a priori as we did not choose $T$ and $\gamma$ with respect to that condition but rather arbitrary. But we can deduce that for our choice of $T$ and $\gamma$ as above, the expansion of $\log Z_N^{V_t}$ holds for any $t \in \mathbb{T}(T, \gamma)$, because the potential can be shown to be $c$-convex. Because of this observation, it is sometimes mentioned that the $c$-convexity encompasses the condition of Ercolani and McLaughlin for $T$ sufficiently small and $\gamma$ sufficiently large, see e.g. [23].

While the first claim only holds for even $n$, claim 2 states that for uneven $n$, only the expansion of Ercolani and McLaughlin is applicable, since $V_t$ cannot be $c$-convex in this situation.

**Proof of Claim 1:**

Let $T > 0$ and $\gamma > 1$ be as in claim 1 and the potential $V_t$ is of form $V_t(H) = \sum_{i=1}^{n} t_i H^i$, where $H$ is a hermitian matrix. First, we observe that $\text{tr}(V_t + \frac{1}{2} H^2)$ is always real. Since $\text{tr}(H^2)$ is always real for a hermitian matrix $H$, we also find that $\text{tr}(V_t)$ is real, since $V_t$ is self-adjoint.
for \( t_i \in \mathbb{R} \), and we have moreover, that the trace of each monomial \( H^k \) for \( k = 1, \ldots, n \) is real,

\[
\text{tr } (H^k) = \text{tr } \left( (\mathcal{P}^t)^k \right) = \text{tr } \left( (\mathcal{P}^k)^t \right) = \text{tr } \left( \mathcal{P}^k \right) = \text{tr } (H^k).
\]

We will deduce the existence of a \( c(T, \gamma) \) such that \( V_i \) is \( c \)-convex by an application of Klein’s lemma, Lemma 4.4, and therefore we show that \( f(x) := \sum_{i=1}^{n} t_i x^i + \frac{1}{2} x^2 \) is a convex function. Thus, we need to establish the positivity of the second derivative,

\[
g(x) := f''(x) = \sum_{i=2}^{n} t_i (i-1)i x^{i-2} + 1 - c \geq 0 \text{ for any } x \in \mathbb{R}.
\]

Observe, that \( g(0) = 2t_2 + 1 - c > 0 \): Because of \( \gamma > 1 \), we see that for \( n \geq 2 \), we find \( T < \frac{1}{2} \), which therefore provides that \( 2t_2 + 1 > 0 \). Hence, we can find a \( c_1(T, \gamma) > 0 \) such that \( g(0) = 2t_2 + 1 - c_1(T, \gamma) > 0 \).

The next case to consider is that of \( \gamma \geq 1 \). As \( t \in \mathbb{T}(T, \gamma) \), it is obvious from (4.36) that

\[
g(x) = \sum_{i=1}^{n} t_i (i-1)i x^{i-2} + 1 - c > \sum_{i=1}^{n-1} (t_i (i-1)i x^{i-2} + \gamma |t_i|(n-1)n x^{n-2}) + 1 - c.
\]

Because \( n \) is even and \( \gamma > 1 \), it holds for every \( i \leq n \) that

\[
\gamma |t_i|(n-1)n x^{n-2} = |\gamma |t_i|(n-1)n x^{n-2}| > |t_i (i-1) i x^{i-2}|
\]

which gives \( g(x) > 0 \) for \( x \) with \( |x| \geq 1 \).

When \( |x| < 1 \), we start with the observation that

\[
0 \leq \left| \sum_{i=1}^{n-1} (t_i (i-1)i x^{i-2} + \gamma |t_i|(n-1)n x^{n-2}) \right| < T(n-n)^2 n + \gamma T n(n-1)^2
\]

\[
= n(n-1)^2 T(1+\gamma) < 1,
\]

which yields

\[
g(x) > \sum_{i=1}^{n-1} (t_i (i-1)i x^{i-2} + \gamma |t_i|(n-1)n x^{n-2}) + 1 > 0.
\]

Thus, we can find a \( c_2(T, \gamma) > 0 \) such that

\[
g(x) > \sum_{i=1}^{n-1} (t_i (i-1)i x^{i-2} + \gamma |t_i|(n-1)n x^{n-2}) + 1 - c_2(T, \gamma) > 0.
\]

Hence, we choose \( c(T, \gamma) = \min \{ c_1(T, \gamma), c_2(T, \gamma) \} \) and applying Klein’s lemma yields

\[
\mathbb{T}(T, \gamma) \subset B_{T, c(T, \gamma)}.
\]

\[\square\]
Proof of Claim 2:
In case of uneven $n$, take e.g. the potential $V_t = 2t^2H^2 + t^3H^3$. Obviously, this potential is not $c$-convex (take e.g. the case $N = 1$, in which the matrix $H$ reduces to one real-valued unknown $h$ and we immediately see that $\text{tr} \left( V_t(h) + \frac{1-c}{2} h^2 \right) = \frac{4t^2+1-c}{2} h^2 + t^3 h^3$ is not a convex function in $h$ for any $c > 0$, regardless of our choice of $\eta > 0$). Along the same lines, we see that any potential $V_t$ with $t_n \neq 0$ for $n$ uneven cannot be $c$-convex. □

Moreover, there is one other possible extension of our results by using the expansion from [20] rather than Theorem 4.13, which we are going to explain next. When using the expansion (4.24), a CLT and a MDP for the centered and properly scaled random variable $\text{tr}(q_l) = \text{tr}(X^l)$ could only be obtained if $t_l \neq 0$. Considering the case of $V_t = t_n H^n, n \geq 2$, either even or uneven, we see that the expansion can be applied to the numerator and denominator in (4.27) even for $t_l$ with $l < n$: Provided that $t \in T^0(T, \gamma)$, where $T^0(T, \gamma)$ denotes the interior of the set $T(T, \gamma)$, it still holds for large $N$ that $\max \{ t_n, |s_N| \} < T$ and $t_n > |s_N|$, for $s \in \mathbb{N}$. And once this expansion was established, the proofs can be left unchanged to yield the CLT and MDP for the distribution of the centered and properly scaled random variable $\text{tr}(X^l) = \sum_{i=1}^{N} \lambda_i^l$ with $l < n$, although $t_l = 0$.

We have seen that in the one-matrix case, $m = 1$, the condition of $c$-convexity includes the condition of Ercolani and McLaughlin when having a positive coefficient of an even leading order, whereas only the latter condition is applicable in some special cases.

### 4.8 Examples

We conclude this chapter by three examples of multi-matrix models, which can be found in [22], where for each model the leading term of the genus expansion (i.e. the first order term) for the free energy (which is the logarithm of the partition function) was given as a solution of a variational problem. These models mostly arouse in physics and have been of interest due to their application to Quantum Field Theory and string theory as well.

#### Ising model on random graphs

The first example is the random Ising model on random graphs, in which two matrices and an interaction term appear, $m = 2$. The Gibbs measure of that model is given by

$$
\mu^N_{ls}(dH_1, dH_2) = \frac{1}{Z^N_{ls}} \exp \left\{ -N \left( \text{tr}(V_{t_1}^1(H_1)) + \text{tr}(V_{t_2}^2(H_2)) - t_3 \text{tr}(H_1H_2) \right) \right\} \mu^N(dH_1) \mu^N(dH_2).
$$
Here, $V_i^j(H_i)$ are convex self-adjoint polynomials depending on the parameter $t_i, i = 1, 2,$ and $t_3 \in \mathbb{R}$ with $|t_3| < 1,$ which is easily seen to guarantee that the function $V_1^1(H_1) + V_2^2(H_2) - t_3H_1H_2$ is $c$-convex, see also [24].

This model has been studied with regard to the first order asymptotic of the logarithmic partition function in [22] by using large deviations techniques and the first order could be given by a variational formula. It was shown, that for $t_3 = 1,$

$$\lim_{N \to \infty} \frac{1}{N^2} \log Z_N^{Is} = \inf_{\mu, \nu, \in M_1(\mathbb{R})} \left\{ \int V_1^1(x) \mu(dx) + \int V_2^2(x) \nu(dx) - \Sigma(\mu) - \Sigma(\nu) - I^{(2)}(\mu, \nu) \right\} - 2 \inf_{\nu, \in M_1(\mathbb{R})} \left\{ \frac{1}{2} \int x^2 \nu(dx) - \Sigma(\nu) \right\},$$

where

$$\Sigma(\mu) = \int \int \log |x - y| \mu(dx)\mu(dy),$$

is the so called non-commutative entropy and the object $I^{(2)}(\mu, \nu)$ is related to so called spherical integrals and itself specified by a contraction principle from a process level large deviation result, where we refer to [22] for more details.

Choosing the parameters $t_i, i = 1, 2, 3,$ small enough, the free energy can also be expanded into a generation function for maps. For example, potential functions $V_j^j$ of type

$$V_j^j = \sum_{i=1}^{n_j} t_{j,i}H_j^i, \quad j = 1, 2,$$

(4.37)

define the appearing $n_j$ vertices which are colored with color $j$ and are of degree $i, i = 1, \ldots, n_j$ for $j = 1, 2.$ The interaction $H_1H_2$ now only establishes edges between vertices, that are of different color. Note that we now color vertices instead of edges, but that does not change the combinatorial interpretation. And the parameter $t$ for that potential can now be stated more precisely, $t = (t_{1,1}, \ldots, t_{1,n_1}, t_{2,1}, \ldots, t_{2,n_2}, t_3)$.

In order to see that this model generalizes the Ising model on $\mathbb{Z}^2,$ take $V_1^1(H) = V_2^2(H) = tH^4$ and for small enough $t,$ we obtain

$$\frac{1}{N^2} \log Z_N^{Is} = \frac{1}{1 - c^2} \sum_{g=0}^k \frac{1}{N^{2g}} \sum_{k,l} \frac{1}{k!} \left( \frac{-t}{(1 - c^2)^2} \right)^k \frac{(-c)^l}{l!} C(k, l, g) + o\left(N^{-2k}\right),$$

where $C(k, l, g)$ is the number of maps of genus $g$ with $k$ vertices of degree 4 being assigned one of two possible colors with exactly $l$ edges between vertices of different colors. In the canonical Ising model on $\mathbb{Z}^2,$ $C(k, l, g)$ is replaced by the number of configurations on a subset of $\mathbb{Z}^2$ rather than on random graphs, where the genus has to be related to the boundary condition
Let us denote the $N$ real eigenvalues of the two $N \times N$ hermitian matrices $H_j^i$ by $\lambda^j_{i,j}$, $j = 1, 2, i = 1, \ldots, N$. Taking the potential functions $V_j^i$, $j = 1, 2$ of form (4.37), we obtain for small enough $t$ for $l \leq n$ with $t_{j,j} \neq 0$ in $V_j^i$ a CLT or a MDP for $\text{tr}(H_j^i) = \sum_{i=1}^{N} \lambda^i_{j,i}$ under $\mu^N_{ls}$.

As seen in section 4.6, the CLT and MDP can in case of small enough $t_{j,i}$ and $t_{3}$ also be extended to hold for traces of polynomials $P$ of type

$$P(H_1, H_2) = \sum_k \alpha_k H_1^k + \sum_i \beta_i H_2^i + \delta H_1 H_2,$$

for $\alpha_k, \beta_i, \delta \in \mathbb{R}$, where we only have to take care, that any monomial appearing in $P$ also has a non-vanishing parameter in its original potential function, i.e. $t_{1,k}, t_{2,i}, t_{3} \neq 0$.

**Chain model**

The chain model can be seen as an extension of the Ising model of the previous paragraph and was considered in [38]. This model finally is an example, which allows for general $m \in \mathbb{N}$.

In that model, the potential function $V$ looks like

$$V(H_1, \ldots, H_m) = \sum_{i=1}^{m} V_{i}^i(H_i) - \sum_{i=2}^{m} t_{2,i}(H_{i-1}H_i),$$

where the $V_i^i$ are convex self-adjoint functions depending on the parameter $t_{1,i}$, $i = 1, \ldots, m$, e.g. of type (4.37). If we take $t_{2,i}$, $i = 1, \ldots, m$ small enough, we can find a $c > 0$, such that $V$ is $c$-convex. Note here, that in the references above, the $t_{2,i}$ are set to 1. While for our results to be applicable, we have to consider only small $t_{1,i}$ and small enough $t_{2,i}$ for $i = 1, \ldots, m$.

Then we can as well obtain CLTs and MDPs for the trace of any monomial or for the trace of any polynomial which consists of monomials appearing somewhere in $V_i$ under the measure

$$\mu^N_{Ch}(dH_1, \ldots, dH_m) = \frac{1}{Z^N_{Ch}} \exp \left\{-N \left( \sum_{i=1}^{m} \text{tr}(V_{i}^i(H_i)) - \sum_{i=2}^{m} t_{2,i} \text{tr}(H_{i-1}H_i) \right) \right\} \prod_{i=1}^{m} \mu^N(dH_i).$$

The name of the model is caused by the structure of dependencies, it allows. Each matrix interacts with its predecessor, thus constituting a chain.
Q-Potts model on random graphs

Changing the chain model slightly, by not coupling each matrix to its predecessor but always to the first matrix $H_1$, the potential function $V$ changes to

$$V(H_1, \ldots, H_q) = \sum_{i=1}^{q} V_{i,1}^i(H_i) - \sum_{i=2}^{q} t_{2,i}(H_1 H_i),$$

where everything else stays the same. We just changed $m$ to $q$, since this model is known (in case of $t_{2,i} = 1, i = 2, \ldots, q$) under the name of q-Potts model and appears for example in [52]. Concerning the CLT and MDP, all that was said in the chain model carries over to the q-Potts model. Due to the similarity of the chain model and the q-Potts model in terms of its potential function and therefore in terms of its Gibbs measure, it will not be too illuminative to give anymore details for the q-Potts model and for this reason, we abandon it and prefer to simply refer to the example above.
4 FLUCTUATIONS IN A MULTI-MATRIX MODEL
## List of Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Explanation</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(X_i)_{i \in \mathbb{N}}$</td>
<td>sequence of random variables</td>
<td>14</td>
</tr>
<tr>
<td>$\overline{X}_n$</td>
<td>empirical mean of $X_1 \ldots, X_n$</td>
<td>14</td>
</tr>
<tr>
<td>$(\mu_n)_{n \in \mathbb{N}}$</td>
<td>sequence of distributions</td>
<td>14</td>
</tr>
<tr>
<td>$I$</td>
<td>rate function</td>
<td>14</td>
</tr>
<tr>
<td>$a_n$</td>
<td>speed</td>
<td>14</td>
</tr>
<tr>
<td>$\Gamma$</td>
<td>Polish space</td>
<td>14</td>
</tr>
<tr>
<td>$B_\Gamma$</td>
<td>Borel $\sigma$-algebra of $\Gamma$</td>
<td>14</td>
</tr>
<tr>
<td>$N_L$</td>
<td>level set</td>
<td>14</td>
</tr>
<tr>
<td>$\langle \cdot, \cdot \rangle$</td>
<td>scalar product</td>
<td>15</td>
</tr>
<tr>
<td>$\mathcal{D}_f$</td>
<td>domain, where $f$ is finite</td>
<td>15</td>
</tr>
<tr>
<td>$\mathcal{D}'_f$</td>
<td>interior of $\mathcal{D}_f$</td>
<td>15</td>
</tr>
<tr>
<td>$M(\lambda)$</td>
<td>moment generating function</td>
<td>15</td>
</tr>
<tr>
<td>$\mathcal{A}$</td>
<td>base of topology of $\Gamma$</td>
<td>17</td>
</tr>
<tr>
<td>$(\Omega, \mathcal{F}, \mathbb{P})$</td>
<td>probability space</td>
<td>18</td>
</tr>
<tr>
<td>$L_n$</td>
<td>empirical measure</td>
<td>18</td>
</tr>
<tr>
<td>$\mathcal{L}(X_1)$</td>
<td>law of random variable $X_1$</td>
<td>18</td>
</tr>
<tr>
<td>$\delta$</td>
<td>indicator function</td>
<td>18</td>
</tr>
<tr>
<td>$\mathcal{M}_1(\Gamma)$</td>
<td>space of probability measures on $\Gamma$</td>
<td>19</td>
</tr>
<tr>
<td>$\Lambda$</td>
<td>scaled limit of cumulant generating function</td>
<td>20</td>
</tr>
<tr>
<td>$\Lambda^*$</td>
<td>Legendre-Fenchel transform of $\Lambda$</td>
<td>20</td>
</tr>
<tr>
<td>$(X_1, \ldots, X_m)$</td>
<td>$m$-tuple of random matrices</td>
<td>23</td>
</tr>
<tr>
<td>$X^*$</td>
<td>adjoint matrix</td>
<td>23</td>
</tr>
<tr>
<td>$P(X)$</td>
<td>density of joint distribution of matrix entries</td>
<td>24</td>
</tr>
<tr>
<td>$dX$</td>
<td>differential of $X$</td>
<td>25</td>
</tr>
<tr>
<td>Symbol</td>
<td>Explanation</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>------</td>
</tr>
<tr>
<td>$\Delta(\lambda)$</td>
<td>Vandermonde determinant</td>
<td>25</td>
</tr>
<tr>
<td>$e^i, i = 1, 2, 3, 4$</td>
<td>Pauli matrices</td>
<td>27</td>
</tr>
<tr>
<td>$q_N$</td>
<td>density of joint distribution of eigenvalues</td>
<td>27</td>
</tr>
<tr>
<td>$Q_N$</td>
<td>joint distribution of eigenvalues</td>
<td>27</td>
</tr>
<tr>
<td>$Z_N$</td>
<td>partition function of joint eigenvalue distribution</td>
<td>27</td>
</tr>
<tr>
<td>$\beta$</td>
<td>parameter indicating real, complex or quaternion matrix entries</td>
<td>28</td>
</tr>
<tr>
<td>$H$</td>
<td>Hamiltonian</td>
<td>28</td>
</tr>
<tr>
<td>$N \wedge M$</td>
<td>minimum of real numbers $N, M$</td>
<td>28</td>
</tr>
<tr>
<td>$h$</td>
<td>stability matrix</td>
<td>32</td>
</tr>
<tr>
<td>$d_\mu(h)$</td>
<td>probability distribution of $h$</td>
<td>32</td>
</tr>
<tr>
<td>$z$</td>
<td>complex number</td>
<td>33</td>
</tr>
<tr>
<td>$\Re(z)$</td>
<td>real part of $z$</td>
<td>33</td>
</tr>
<tr>
<td>$\Im(z)$</td>
<td>imaginary part of $z$</td>
<td>33</td>
</tr>
<tr>
<td>$\mathcal{H}_N(\mathbb{C})^m$</td>
<td>space of $m$-tuples of hermitian matrices</td>
<td>33</td>
</tr>
<tr>
<td>$d^N_H$</td>
<td>Lebesgue measure on $\mathcal{H}_N(\mathbb{C})^m$</td>
<td>33</td>
</tr>
<tr>
<td>$\mu_N^\mu(dH)$</td>
<td>law of $m$-dimensional GUE</td>
<td>33</td>
</tr>
<tr>
<td>$\mu_N^V(dH)$</td>
<td>measure of the multi-matrix model</td>
<td>33</td>
</tr>
<tr>
<td>$V$</td>
<td>polynomial potential</td>
<td>33</td>
</tr>
<tr>
<td>$q_j$</td>
<td>non-commutative polynomial in $m$ indeterminants</td>
<td>33</td>
</tr>
<tr>
<td>$Z_N^V$</td>
<td>partition function in the multi-matrix model</td>
<td>33</td>
</tr>
<tr>
<td>$\mathbb{E}[X]$</td>
<td>expectation of random variable $X$</td>
<td>34</td>
</tr>
<tr>
<td>$g$</td>
<td>genus of a surface</td>
<td>34</td>
</tr>
<tr>
<td>$F_g(t)$</td>
<td>generating function for maps</td>
<td>34</td>
</tr>
<tr>
<td>$w$, resp. $w_n$</td>
<td>weight functions</td>
<td>40</td>
</tr>
<tr>
<td>$L^2$</td>
<td>space of square Lebesgue-integrable functions</td>
<td>40</td>
</tr>
<tr>
<td>$(p_i)_{i \in \mathbb{N}}$</td>
<td>family of orthogonal polynomials</td>
<td>40</td>
</tr>
<tr>
<td>$\theta$</td>
<td>integer-valued parameter for biorthogonal ensembles</td>
<td>40</td>
</tr>
<tr>
<td>$\Sigma$</td>
<td>closed subset of $\mathbb{R}$</td>
<td>41</td>
</tr>
<tr>
<td>$p(n)$</td>
<td>number of eigenvalues</td>
<td>41</td>
</tr>
<tr>
<td>Symbol</td>
<td>Explanation</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>------</td>
</tr>
<tr>
<td>$\mathcal{N}(f)$</td>
<td>set of zeros of $f$</td>
<td>42</td>
</tr>
<tr>
<td>$C_b(\Sigma)$</td>
<td>space of continuous and bounded functions on $\Sigma$</td>
<td>43</td>
</tr>
<tr>
<td>weak convergence</td>
<td></td>
<td>43</td>
</tr>
<tr>
<td>$\gamma(A)$</td>
<td>logarithmic capacity of set $A$</td>
<td>44</td>
</tr>
<tr>
<td>$|\cdot|_\infty$</td>
<td>supremum norm</td>
<td>47</td>
</tr>
<tr>
<td>$|\cdot|_A$</td>
<td>supremum norm restricted to a set $A$</td>
<td>47</td>
</tr>
<tr>
<td>$m_{\mathbb{R}^n}(dx)$</td>
<td>product Lebesgue measure on $\mathbb{R}^n$</td>
<td>51</td>
</tr>
<tr>
<td>$l_f$</td>
<td>Lipschitz constant of function $f$</td>
<td>54</td>
</tr>
<tr>
<td>$\xi^{(n)} = (\xi_{p(n)}, \ldots, \xi_1)$</td>
<td>vector of quantiles</td>
<td>55</td>
</tr>
<tr>
<td>$\psi_n, \psi$</td>
<td>step functions</td>
<td>55</td>
</tr>
<tr>
<td>$B(\mu, \delta)$</td>
<td>ball in $\mathcal{M}_1(\Sigma)$ around $\mu$ with radius $\delta$</td>
<td>56</td>
</tr>
<tr>
<td>$\mathbb{R}^{d,+}$</td>
<td>space of ordered vectors in $\mathbb{R}^d$</td>
<td>57</td>
</tr>
<tr>
<td>$\dagger$</td>
<td>involution acting on monomials</td>
<td>71</td>
</tr>
<tr>
<td>$c$</td>
<td>parameter for $c$-convexity</td>
<td>72</td>
</tr>
<tr>
<td>$\lambda_1, \ldots, \lambda_N$</td>
<td>eigenvalues of hermitian matrix</td>
<td>72</td>
</tr>
<tr>
<td>$F^0(t)$</td>
<td>generating function for maps in a multi-matrix model</td>
<td>74</td>
</tr>
<tr>
<td>$\varphi_l$</td>
<td>trace of monomial $q_l$</td>
<td>75</td>
</tr>
<tr>
<td>$\overline{\varphi_l}$</td>
<td>centered trace of monomial $q_l$</td>
<td>75</td>
</tr>
<tr>
<td>$G = (V, E)$</td>
<td>graph with set of vertices $V$ and set of edges $E$</td>
<td>76</td>
</tr>
<tr>
<td>$\chi$</td>
<td>Euler’s characteristic</td>
<td>76</td>
</tr>
<tr>
<td>$f$</td>
<td>number of free parameters</td>
<td>80</td>
</tr>
<tr>
<td>$C_k$</td>
<td>$k$-th Catalan number</td>
<td>84</td>
</tr>
<tr>
<td>$\mathcal{S}^N(\mathbb{R})$</td>
<td>space of $N \times N$ symmetric matrices</td>
<td>85</td>
</tr>
<tr>
<td>$C_g^k$</td>
<td>number of maps on a surface of genus $g$ with $k_i$ vertices of type $q_i$</td>
<td>91</td>
</tr>
<tr>
<td>$C_g^k(P)$</td>
<td>number of maps on a surface of genus $g$ with $k_i$ vertices of type $q_i$ and one vertex of type $P$</td>
<td>91</td>
</tr>
<tr>
<td>$\mathcal{D}_j$</td>
<td>operator of derivation</td>
<td>92</td>
</tr>
<tr>
<td>$s \vee t$</td>
<td>maximum of real numbers $s, t$</td>
<td>96</td>
</tr>
</tbody>
</table>
LIST OF SYMBOLS
List of Figures

4.1 Surfaces of genus $g=0$, 1 and 2 ............................................. 75
4.2 Example of a graph not leading to a map ................................... 76
4.3 Map with 2 vertices of degree 2 and 4 on sphere and torus ........... 76
4.4 A one-face Polygon with 10 directed sides ................................ 81
4.5 Gluing pairwise the edges of the polygon ................................. 81
4.6 An edge consisting of two arrows ............................................. 83
4.7 A vertex depicted by a star ....................................................... 83
4.8 All possible fat-graphs with one star of degree 4 ....................... 84
4.9 Some unoriented fat-graphs of genus $g=1$ and with one star with degree 4 .......... 86
4.10 Configuration for $(\text{tr}(H^1))^n$ ............................................. 87
4.11 Star of type $q_l = H_1^2 H_2^2 H_1 H_2^3$ and of type $q_l^* = H_2^3 H_1 H_2^3 H_1^2$ .......... 89
4.12 Maps of genus 0 and 1 with two stars of type $H_1 H_2 H_1 H_2$ ................. 90
Bibliography


BIBLIOGRAPHY


