# RANDOM MOMENT PROBLEMS UNDER CONSTRAINTS

HOLGER DETTE[*], DOMINIK TOMECKI[†], AND MARTIN VENKER[‡]

ABSTRACT. We investigate moment sequences of probability measures on $E \subset \mathbb{R}$ under constraints of certain moments being fixed. This corresponds to studying sections of $n$-th moment spaces, i.e. the spaces of moment sequences of order $n$. By equipping these sections with the uniform or more general probability distributions, we manage to give for large $n$ precise results on the (probabilistic) barycenters of moment space sections and the fluctuations of random moments around these barycenters. The measures associated to the barycenters belong to the Bernstein-Szegő class and show strong universal behavior. We prove Gaussian fluctuations and moderate and large deviations principles. Furthermore, we demonstrate how fixing moments by a constraint leads to breaking the connection between random moments and random matrices.

## 1. INTRODUCTION

Classical moment problems on the real line pose the question whether a given sequence of real numbers is the moment sequence of a positive Borel measure with support in a prescribed set $E \subset \mathbb{R}$ and whether such a measure, if it exists, is unique. Most notable are the Hamburger, Stieltjes and Hausdorff moment problems, which correspond to the sets $E = \mathbb{R}$, $E = \mathbb{R}_+ := [0, \infty)$ and $E$ being a compact interval, respectively. Solutions to these moment problems have been known for a long time.

In the classical moment problems one is thus interested in *all* possible moment sequences. In contrast, the random moment problem asks how a *typical* moment sequence looks like. To make this precise, let us denote by $\mathcal{P}(E)$ the set of all Borel probability measures on a Borel set $E \subset \mathbb{R}$ possessing moments of all order, and by $m_j(\mu) := \int x^j \, d\mu(x)$ the $j$-th moment of a measure $\mu \in \mathcal{P}(E)$. The set

$$\mathcal{M}_n(E) := \{(m_1(\mu), \ldots, m_n(\mu)) \mid \mu \in \mathcal{P}(E)\}$$

is called *$n$-th moment space*. It is a convex set in $\mathbb{R}^n$ with positive Lebesgue measure. Beginning with Karlin and Shapley (1953), Karlin and Studden (1966) and Kreĭn and Nudel'man (1977), geometric aspects of $\mathcal{M}_n(E)$ have been investigated in many works. A probabilistic investigation was initiated by Chang et al. (1993), who equipped $\mathcal{M}_n([0, 1])$ with the uniform distribution and studied the behavior of a fixed number of the now random moments as the dimension $n$ converges to infinity. They observed that in high dimension such a random moment sequence concentrates near the moment sequence of the arcsine distribution

$$\mu^0(dx) := \frac{1}{\pi \sqrt{x(1-x)}} \, dx. \tag{1.1}$$

Thus the moment sequence of the arcsine distribution may be seen as a probabilistic barycenter of the moment space $\mathcal{M}_n([0, 1])$. More precisely, let $(m_1^{(n)}, \ldots, m_n^{(n)})$ be drawn from the uniform distribution on $\mathcal{M}_n([0, 1])$ and $l \in \mathbb{N}$ be fixed. Then, as $n \to \infty$,

$$(m_1^{(n)}, \ldots, m_l^{(n)}) \to (m_1(\mu^0), \ldots, m_l(\mu^0)) \tag{1.2}$$

in probability. Moreover, Chang et al. (1993) proved that

$$\sqrt{n}((m_1^{(n)}, \ldots, m_l^{(n)}) - (m_1(\mu^0), \ldots, m_l(\mu^0))) \tag{1.3}$$

---

converge in distribution to a multivariate normal distribution as $n \to \infty$. Later, Gamboa and Lozada-Chang (2004) showed fluctuations on the scales of moderate and large deviations principles also for $E = [0,1]$. Dette and Nagel (2012) studied special distributions on the unbounded moment spaces $\mathcal{M}_n(\mathbb{R}_+)$ and $\mathcal{M}_{2n+1}(\mathbb{R})$ and also proved central limit theorems. They found moment sequences of a certain Marchenko-Pastur distribution and Wigner's semicircle distribution, respectively, replacing the one of the arcsine measure.

However, to speak of a typical moment sequence it would be desirable to have a certain universality in the sense that the limiting moment sequences should not strongly depend on the probability distribution which has been put on the moment space $\mathcal{M}_n(E)$. The question of universality is even more prominent in the case of unbounded $E$ as then $\mathcal{M}_n(E)$ is unbounded itself and thus can not carry the "natural" uniform distribution. Therefore recently Dette et al. (2018) gave a unifying view on the random moment problem by identifying classes of distributions on $\mathcal{M}_n(E)$ in all three cases $E = [a,b], \mathbb{R}_+, \mathbb{R}$ that admit universal behavior: On $\mathbb{R}_+$ and $\mathbb{R}$, the moment sequences in these classes always converge in the large $n$ limit to the moment sequences of the Marchenko-Pastur distributions and those of the semicircle distributions, respectively. For $E = [a,b]$, the arcsine measure was found in Dette et al. (2018) to be rather a special member of the universal family of Kesten-McKay (or free binomial) distributions than being universal itself.

The occurrence of the three families of Kesten-McKay, Marchenko-Pastur and semicircle distributions is somewhat curious and suggests a connection to random matrix theory, where these distributions appear as limits of empirical spectral distributions for the so-called Jacobi, Laguerre and Wigner ensembles, respectively. They are characterized as equilibrium measures to certain external fields on $\mathbb{R}$. We will illuminate the connection of random moments to equilibrium measures, orthogonal polynomials and random matrix theory in the course of this paper.

From a geometric point of view, moment spaces are interesting convex sets that admit special parametrizations and therefore allow for a detailed analysis. For instance, the moment space $\mathcal{M}_n([0,1])$ is a convex body contained in $[0,1]^n$. It is very far from other convex bodies like balls, hypercubes or cross-polytopes regarding the strength of the dependence between coordinates. For example, to obtain Gaussian fluctuations of random points in the three mentioned classical convex bodies, one needs to involve a growing number of coordinates due to the rather mild dependence structure. In striking contrast to that, (1.3) with $l = 1$ shows that even the first coordinate $m_1^{(n)}$ shows for $n \to \infty$ Gaussian fluctuations under the uniform distribution on $\mathcal{M}_n([0,1])$, indicating a very strong dependence between all moments. The analysis of moment spaces generally uses special independent coordinates that unravel the dependence structure of moments.

In the present work we investigate the behavior of random moment sequences when certain moments are fixed by a constraint. This corresponds to the question of a typical moment sequence when some moments are a priori known. From a more geometric perspective, we investigate slices of the moment space $\mathcal{M}_n(E)$, thereby providing a better insight in its shape. In particular, constraining moment sequences enables to study different regions in moment spaces, see e.g. Figure 1 below, where we show the moment space $\mathcal{M}_2([0,1])$ and constrained moment spaces obtained by fixing the third moment $m_3$. Of special interest are existence and structure of (probabilistic) barycenters of these constrained moment spaces as well as volume and more refined probabilistic questions like fluctuation laws of the random constrained moment sequences.

Let us now make things precise. Under a constraint $\mathcal{C}$ we understand a finite collection of integer indices $1 \le i_1 < \ldots < i_k, k \in \mathbb{N}$ and corresponding values $c_{i_1}, \ldots, c_{i_k}$ which we denote as $\mathcal{C} := \{m_{i_1} = c_{i_1}, \ldots, m_{i_k} = c_{i_k}\}$. For instance, the constraint $\mathcal{C} = \{m_1 = c_1\}$ means fixing the expectation, whereas $\mathcal{C} = \{m_1 = c_1, m_2 = c_2\}$ also fixes the variance. We allow for $k = 0$, corresponding to unconstrained moment spaces. We now want to examine the moments of probability measures whose $i_j$-th moment is given by $c_{i_j}, j = 1, \ldots, k$.

**Definition 1.1** (Admissible Constraint). For a constraint $\mathcal{C}$ we denote by

$$\mathcal{P}^{\mathcal{C}}(E) := \{\mu \in \mathcal{P}(E) \mid m_{i_j}(\mu) = c_{i_j} \; \forall \, 1 \leq j \leq k\}$$

the space of probability measures on $E$ fulfilling constraint $\mathcal{C}$ and by

$$\mathcal{M}_n^{\mathcal{C}}(E) := \{(m_1(\mu), \ldots, m_n(\mu)) \mid \mu \in \mathcal{P}^{\mathcal{C}}(E)\}$$

the constrained $n$-th moment space. A constraint $\mathcal{C}$ is called *admissible* for $E$, if the intersection $\mathcal{M}_{i_k}^{\mathcal{C}}(E) \cap \text{Int} \, \mathcal{M}_{i_k}(E)$ is nonempty, where here and lateron, Int denotes the interior.

For ease of notation, we will assume throughout the article that $\mathcal{C}$ is an admissible constraint for $E$ with indices $i_1, \ldots, i_k$ and corresponding values $c_{i_1}, \ldots, c_{i_k}$. Note that the notion of admissibility depends on the set $E$.

Let us illustrate how a constraint restricts the moment space and allows to study different regions of the space. Figure 1 shows $\mathcal{M}_2^{\mathcal{C}}([0,1])$ for $\mathcal{C} = \{m_3 = c_3\}$ (encircled in black) inside of $\mathcal{M}_2([0,1])$ (encircled in grey) for two values of $c_3$. In the first plot $c_3 = 0.3125$ which is the third moment of the arcsine measure. The second plot is for $c_3 = 0.1$.



(A) $m_3 = 0.3125$ (arcsine law)  (B) $m_3 = 0.1$
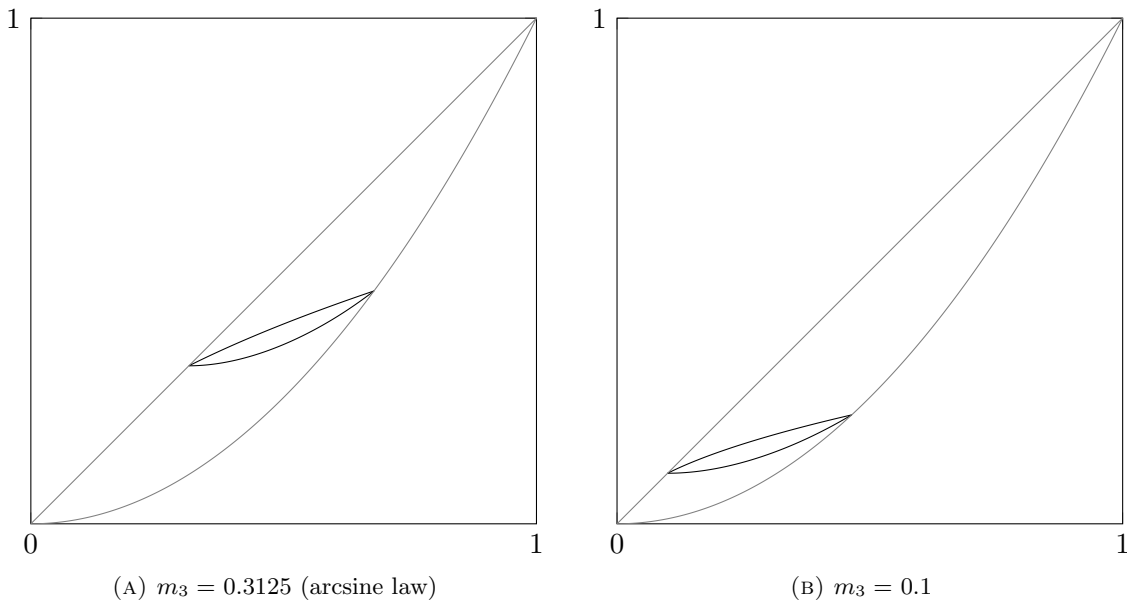
FIGURE 1. Visualizations of $\mathcal{M}_2^{\mathcal{C}}([0,1])$ under constraints on $m_3$.

This article is structured as follows. In the next section, we will first consider the uniform distribution on $\mathcal{M}_n^{\mathcal{C}}(E)$ for a bounded interval $E$. We will see that a uniformly distributed random moment sequence converges for $n \to \infty$ in the sense of (1.2), where the limiting measure has a density w.r.t. the arcsine measure (1.1). However, it is in general not the equilibrium measure under the constraint $\mathcal{C}$ to the uniform external field as might be expected from the discussion above. While an equilibrium measure under a constraint $\mathcal{C}$ should be obtained by minimizing Voiculescu's free entropy (from free probability theory, see e.g. (Akemann et al., 2011, Chapter 22 and references therein)) to the arcsine measure over the set of measures compatible with the constraint, we rather find the Kullback-Leibler divergence or relative entropy from classical probability playing a major role. This parting of the ways of the random moment problems and random matrix theory will be explained in detail in Section 3.1 and has its roots in the fact that introducing a constraint breaks the asymptotic equivalence of the spectral measure encountered in the random moment problem and the empirical spectral measure encountered in random matrix theory. Nevertheless, the limiting measures we find in this paper belong to a famous class of measures as well, the so-called Bernstein-Szegő class. Section 2 also provides a computation of the volume of the constrained moment spaces, a central limit theorem as well

as moderate and large deviations principles. Section 3 takes a broader approach and defines more general classes of distributions on the constrained moment spaces, in particular on the unbounded moment spaces $\mathcal{M}_n^{\mathcal{C}}(\mathbb{R}_+)$ and $\mathcal{M}_n^{\mathcal{C}}(\mathbb{R})$. For generic densities on these spaces, we identify the universal structures of the limiting moment sequences and give results on the fluctuations around these limits on several scales. We provide a detailed and extensive analysis, even in those cases where the random moment sequences do not have a single limit but rather concentrate around a finite set of limit points. Sections 4, 5 and 6 are devoted to the proofs of the results in Section 3, whereas the results in Section 2 are proved in Section 7.

## 2. Uniformly distributed random moment sequences

In this section we study the case of random moment sequences uniformly distributed in $\mathcal{M}_n^{\mathcal{C}}(E)$ for $E$ being a compact interval. Without loss of generality we will choose $E = [0,1]$, since by the linear transformation $t \mapsto a + t(b-a)$ the results for $[0,1]$ can be transferred to any interval $[a,b]$. Note that although Section 3 covers more general distributions on the moment spaces, the results in the present section are more explicit and not easily deduced from the ones of Section 3.

Recall our convention that $\mathcal{C}$ is an admissible constraint (see Definition 1.1) of the form $m_{i_1} = c_{i_1}, \ldots, m_{i_k} = c_{i_k}$. The constrained moment space $\mathcal{M}_n^{\mathcal{C}}([0,1])$ can be identified canonically with the set of the $(n-k)$-dimensional vectors of the unconstrained moments $(m_j, 1 \leq j \leq n, j \neq i_1, \ldots, i_k)$. The set of unconstrained moments is a convex and compact subset of the $(n-k)$-dimensional unit cube and has due to admissibility of $\mathcal{C}$ non-zero Lebesgue measure. Pushing forward the $n-k$-dimensional Lebesgue measure from the unconstrained moments to $\mathcal{M}_n^{\mathcal{C}}([0,1])$, we can equip $\mathcal{M}_n^{\mathcal{C}}([0,1])$ with the uniform distribution, which allows us to investigate the behaviour of a "typical" moment sequence on the constrained moment spaces.

Throughout this section let $(m_1^{(n)}, \ldots, m_n^{(n)})$ be drawn from the uniform distribution on $\mathcal{M}_n^{\mathcal{C}}([0,1])$. Here and lateron, we will tacitly assume that all random variables are defined on a common probability space such that we can speak of almost sure convergence.

Our first result is a law of large numbers that identifies the limiting moment sequence to which the random moment sequence converges.

**Theorem 2.1.** *For any $l \in \mathbb{N}$ we have as $n \to \infty$*

$$(m_1^{(n)}, \ldots, m_l^{(n)}) \to \left( m_1(\mu^{\mathcal{C}}), \ldots, m_l(\mu^{\mathcal{C}}) \right) \quad a.s.,$$

*where $\mu^{\mathcal{C}}$ is a probability measure on $[0,1]$ of the form*

$$\mu^{\mathcal{C}}(dx) = \frac{1}{S_{i_k}(x)\sqrt{x(1-x)}}dx. \tag{2.1}$$

*Here $S_{i_k}$ is a polynomial of degree at most $i_k$ that is strictly positive on the interval $[0,1]$. Furthermore, $\mu^{\mathcal{C}}$ is the unique probability measure that minimizes the Kullback-Leibler divergence*

$$\mathcal{K}(\mu^0|\mu) := \begin{cases} \int \log \frac{d\mu^0}{d\mu}\, d\mu^0 & , \mu^0 \ll \mu \\ \infty & , else \end{cases} \tag{2.2}$$

*among all probability measures $\mu \in \mathcal{P}^{\mathcal{C}}([0,1])$, where $\mu^0$ is the arcsine distribution defined in* (1.1).

**Remark 2.2.**

(1) The measure $\mu^{\mathcal{C}}$ belongs to the so-called Bernstein-Szegő class on $[0,1]$ which consists of measures of the form

$$\mu(dx) = \frac{(x(1-x))^{\pm\frac{1}{2}}}{S(x)}dx,$$

where $S$ is a polynomial strictly positive on the interval $[0,1]$. They play a key role in the theory of orthogonal polynomials and possess many useful properties, e.g. explicit formulae for their orthogonal polynomials. For details and references we refer to (Szegő, 1975, §2.6). The connection between moments and orthogonal polynomials is well-known and at the roots of both theories. Theorem 2.1 shows that these important measures of orthogonal polynomials theory are also central for moment spaces in the sense that their moments provide the probabilistic barycenters of sections of the moment space. More general members of the Bernstein-Szegő class will be found in Section 3 when discussing random moment sequences over the unbounded spaces $\mathbb{R}_+$ and $\mathbb{R}$.

(2) For the unconstrained random moment problem, the probabilistic barycenter is given by the moment sequence of the arcsine measure $\mu^0$. It is the equilibrium measure on $[0,1]$ to the external field $Q_{ex} = 0$, where we recall that the equilibrium measure to an external field $Q_{ex} : E \to \mathbb{R}$ is the unique Borel probability measure $\mu$ on $E$ such that

$$\int_E \int_E \log|t-s|^{-1}d\mu(t)d\mu(s) + \int_E Q_{ex}(t)d\mu(t) \qquad (2.3)$$

is minimal, see e.g. Saff and Totik (1997). The Marchenko-Pastur distributions are (if there is no atom at 0) the equilibrium measures on $E = \mathbb{R}_+$ to $Q_{ex}(t) = \frac{t}{z_2} - \frac{z_1 - z_2}{z_2}\log t$ for some constants $0 < z_2 \leq z_1$. Likewise, the semicircle distributions are obtained as equilibrium measures on $E = \mathbb{R}$ to external fields that are quadratic polynomials with positive leading coefficient. One might thus expect $\mu^{\mathcal{C}}$ from Theorem 2.1 to be the minimizer of a constrained equilibrium problem. However, we can deduce from representation (2.1) that for constraints $\mathcal{C}$ with $\mu^0 \notin \mathcal{P}^{\mathcal{C}}([0,1])$, $\mu^{\mathcal{C}}$ is not the solution of the constrained equilibrium problem

$$\inf_{\mu \in \mathcal{P}^{\mathcal{C}}([0,1])} \int_{[0,1]} \int_{[0,1]} \log|t-s|^{-1}d\mu(t)d\mu(s). \qquad (2.4)$$

As variational calculus shows, a solution $\mu \in \mathcal{P}^{\mathcal{C}}([0,1])$ of (2.4) has to fulfill the Euler-Lagrange equations

$$2\int \log|t-s|^{-1}d\mu(s) + \sum_{j=1}^{k}\lambda_j t^{i_j} \begin{cases} = c, & t \in \text{supp}(\mu), \\ \geq c, & t \notin \text{supp}(\mu), \end{cases}$$

for some $c$ and Lagrange multipliers $\lambda_1, \ldots, \lambda_k$. Thus the moment constraint $\mathcal{C}$ leads to the appearance of polynomial external fields. Equilibrium measures to such fields are well-understood, see e.g. (Deift et al., 1998, Theorem 1.38). The support of the equilibrium measure $\mu$ consists of finitely many intervals in $[0,1]$ with non-empty interior and it has a density of the form

$$\mu(dx) = T(x) \prod_{a \in \text{HE}} \frac{1}{\sqrt{|x-a|}} \prod_{b \in \text{SE}} \sqrt{|x-b|} 1_{\text{supp}(\mu)}(x)dx.$$

Here $\text{HE} \subset \{0,1\}$ is the (possibly empty) set of the so-called hard edges and SE the set of so-called soft edges, and $T$ is a polynomial which is strictly positive on $\text{supp}(\mu)$. As an example, the minimizer of (2.4) under the constraint $\mathcal{C} = \{m_1 = c_1, m_2 = c_2\}$ is for $c_1 \in (0,1)$ and $c_2 > c_1^2$ small enough the semicircle distribution $\mu(dx) = c\sqrt{4\beta - (x-c_1)^2}1_{[a,b]}(x)dx$ for some $a,b,\beta,c > 0$. This is very different from the measure $\mu^{\mathcal{C}}$, which we will give in (2.5) below. This implies that already the supports of the equilibrium measure under a constraint and of the measure $\mu^{\mathcal{C}}$ can be very different, as the former can be arbitrarily small while the latter is always $[0,1]$. We will explain this phenomenon in Section 3.1.

(3) The Kullback-Leibler divergence $\mathcal{K}(\mu^0|\mu)$ is also called relative entropy. It is always non-negative and can be understood as a distance measure for probability distributions. Indeed, although not being a metric itself, one has Pinsker's inequality $\|\mu - \nu\|_{TV} \leq \sqrt{\mathcal{K}(\nu|\mu)/2}$ (see Csiszár (1967)) and thus convergence of the Kullback-Leibler divergence

to 0 implies convergence in the total variation distance $\|\cdot\|_{TV}$. It also appears as the rate function of a large deviations principle in Sanov's theorem. Note however, that in typical encounters of $\mathcal{K}$, the minimization is in the first argument. The connection of $\mathcal{K}$ to random moment problems was first observed in Gamboa and Lozada-Chang (2004). There the authors call it *reversed* Kullback-Leibler divergence because of the minimization in the second argument.

**Example 2.3.** The proof of Theorem 2.1 is constructive in the sense that the polynomial $S_{i_k}$ can be computed from the constraint $\mathcal{C}$. Here we provide some examples, for the full details we refer to Section 4. For $\mathcal{C}_1 := \{m_1 = c_1\}$ (fixed mean) with $c_1 \in (0,1)$, the limiting measure is

$$\mu^{\mathcal{C}_1}(dx) = \frac{c_1(1-c_1)}{\pi\sqrt{x(1-x)((1-2c_1)x+c_1^2)}}\,dx.$$

For a constraint $\mathcal{C}_2 := \{m_1 = c_1, m_2 = c_2\}$ (fixed mean and fixed variance) the admissibility condition reads $c_1^2 < c_2 < c_1$ and the limiting measure is

$$\mu^{\mathcal{C}_2}(dx) = \frac{c_1(1-c_1)(c_2-c_1^2)(c_1-c_2)}{\pi\sqrt{x(1-x)}\big((c_1-c_2)^2(x-c_1)^2+(c_2-c_1^2)^2x(1-x)\big)}\,dx. \tag{2.5}$$

Finally, if only the second moment is fixed, i.e. $\mathcal{C}_3 := \{m_2 = c_2\}$, then $\mu^{\mathcal{C}_3}$ is given by (2.5) with $c_1$ being the unique maximizer of the function

$$c_1 \mapsto \frac{(c_2-c_1^2)(c_1-c_2)}{c_1(1-c_1)}$$

on $[c_2, \sqrt{c_2}]$.

The limiting measure $\mu^{\mathcal{C}}$ allows for an effective description of the volume of $\mathcal{M}_n^{\mathcal{C}}([0,1])$. (Karlin and Studden, 1966, Theorem IV.6.2) gave an expression of the volume of the unconstrained space $\mathcal{M}_n([0,1])$ in terms of gamma functions, which can by a direct application of Stirling's formula be written as

$$\text{vol}_n(\mathcal{M}_n([0,1])) = \prod_{m=1}^{n-1}\frac{\Gamma(m)^2}{\Gamma(2m)} = 2^{-n^2}\Big(\frac{\pi e}{n}\Big)^{n/2}n^{-1/8}(1+\mathcal{O}(1)), \tag{2.6}$$

$\text{vol}_n$ denoting the $n$-dimensional Lebesgue measure. We also need to introduce some notation. Let $m_1,\dots,m_l \in \mathcal{M}_l([0,1])$. Then the possible range of $m_{l+1}$ such that $(m_1,\dots,m_{l+1}) \in \mathcal{M}_{l+1}([0,1])$ holds, is an interval, say $[m_{l+1}^-, m_{l+1}^+]$. In the next theorem, $m_{i_k+1}^{\pm}(\mu^{\mathcal{C}})$ will denote the numbers $m_{i_k+1}^{\pm}$ to the first $i_k$ moments of the measure $\mu^{\mathcal{C}}$ introduced in Theorem 2.1.

**Theorem 2.4.** *As $n \to \infty$*

$$\frac{\text{vol}_{n-k}(\mathcal{M}_n^{\mathcal{C}}([0,1]))}{\text{vol}_n(\mathcal{M}_n([0,1]))} = \frac{\big((m_{i_k+1}^+ - m_{i_k+1}^-)(\mu^{\mathcal{C}})\big)^{n-i_k}n^{k/2}(2\pi)^{k/2}/\sqrt{d^{\mathcal{C}}}}{4^{-i_k(i_k-3)/2}4^{-i_kn}(\sqrt{\pi/4})^{i_k}}(1+o(1)),$$

$$= \begin{cases} \mathcal{O}\left(e^{-cn}\right) & \text{for some } c>0 \quad, \quad \text{if } \mu^0 \notin \mathcal{P}^{\mathcal{C}}(E), \\ \mathcal{O}\left(n^{k/2}\right) & , \quad \text{if } \mu^0 \in \mathcal{P}^{\mathcal{C}}(E), \end{cases}$$

*where $d^{\mathcal{C}}$ is a constant independent of $n$.*

Theorem 2.4 can be interpreted as follows: If the constraint $\mathcal{C}$ is chosen such that the arcsine distribution does not lie in $\mathcal{P}^{\mathcal{C}}([0,1])$, then the relative volume of the constrained moment space goes to zero exponentially fast. If the arcsine distribution lies in $\mathcal{P}^{\mathcal{C}}([0,1])$, then the relative volume grows polynomially. The growth of the relative volume in the latter case reflects the fact that the volume of the (unconstrained) moment space decreases with increasing dimension. Another way to interpret this is that almost all intersections of the moment space with hyperplanes orthogonal to the standard basis vectors are "small". The only large intersections are those corresponding to a constraint in which $m_{i_j} = m_{i_j}(\mu^0)$ holds for all $1 \le j \le k$.

We now return to the probabilistic setting and turn to fluctuations of the random moment sequence around the limiting measure $\mu^{\mathcal{C}}$ on several scales. For fluctuations of order $1/\sqrt{n}$ we observe Gaussian laws.

**Theorem 2.5** (Central Limit Theorem)**.** *Let $l \geq i_k$. We have as $n \to \infty$*

$$\sqrt{n}((m_1^{(n)}, \ldots, m_l^{(n)}) - (m_1(\mu^{\mathcal{C}}), \ldots, m_l(\mu^{\mathcal{C}})) \to \mathcal{N}(0, \Sigma_l)$$

*in distribution, where $\mathcal{N}(0, \Sigma_l)$ denotes the multivariate normal distribution with mean 0 and $\Sigma_l$ is an $l \times l$ matrix of rank $l - k$ that consists of zeros in the columns and rows $i_1, \ldots, i_k$.*

We remark in passing that the condition $l \geq i_k$ is only assumed to state the result, especially the form of the matrix $\Sigma_l$, in a convenient way. Of course, the vector $(m_1, \ldots, m_l)$ also exhibits Gaussian fluctuations for any $l < i_k$. For a precise definition of the matrix $\Sigma_l$ we first need certain preliminaries like a parametrization of the moment space. For the reader's convenience, the expression of $\Sigma_l$ is given in (7.9).

For our results on fluctuations on larger scales, we first recall the notion of an LDP. Let $\mathcal{X}$ be a topological space and $\mathbb{P}_n$ a sequence of probability measures on $\mathcal{X}$, equipped with the Borel $\sigma$-field. $\mathbb{P}_n$ is said to satisfy a *large deviations principle* (LDP) with speed $a_n \to \infty$ and rate function $I$ if $I : \mathcal{X} \to [0, \infty]$ is lower semicontinuous and for all measurable $\Gamma \subset \mathcal{X}$ the inequalities

$$- \inf_{x \in \text{Int } \Gamma} I(x) \leq \liminf_{n \to \infty} \frac{1}{a_n} \log \mathbb{P}_n(\Gamma) \leq \limsup_{n \to \infty} \frac{1}{a_n} \log \mathbb{P}_n(\Gamma) \leq - \inf_{x \in \overline{\Gamma}} I(x) \tag{2.7}$$

hold. $I$ is called a *good rate function*, if its level sets $I^{-1}((-\infty, \alpha])$ are compact for all $\alpha \geq 0$. In the following, $\mathcal{X} = \mathbb{R}^l$ will always be equipped with the Euclidean topology.

The next result shows that for fluctuations between the order $1/\sqrt{n}$ and order 1, the Gaussian distribution from Theorem 2.5 still provides the leading order asymptotics. Therefore such deviations are called moderate. Throughout this paper $A^t$ denotes the transpose of the matrix $A$.

**Theorem 2.6** (Moderate Deviations Principle)**.** *Let $l \geq i_k$ and $(a_n)_n$ be a sequence with $a_n \to \infty$ and $a_n = o(n^{1/2})$ as $n \to \infty$. Then the sequence*

$$a_n((m_1^{(n)}, \ldots, m_l^{(n)}) - (m_1(\mu^{\mathcal{C}}), \ldots, m_l(\mu^{\mathcal{C}})))$$

*satisfies an LDP with speed $\frac{n}{a_n^2}$ and good rate function*

$$I(x) := \begin{cases} \frac{1}{2} x^t \Sigma_l x & , x_{i_1} = \ldots = x_{i_k} = 0 \\ \infty & , \text{else} \end{cases},$$

*where $\Sigma_l$ is the covariance matrix from Theorem 2.5.*

Note that for $l < i_k$ the moderate deviations principle (MDP) can be obtained from the previous result by an application of the contraction principle. This remark extends to all LDPs in this article.

Finally, for fluctuations of order 1 we can see that the random moment sequences satisfy a large deviations principle with a rate function that is universal for all $\mathcal{C}$ up to an additive constant. To state it, recall that given $(m_1, \ldots, m_l) \in \mathcal{M}_l([0, 1])$, the possible range of $m_{l+1}$ such that $(m_1, \ldots, m_{l+1}) \in \mathcal{M}_{l+1}([0, 1])$ holds, is an interval $[m_{l+1}^-, m_{l+1}^+]$.

**Theorem 2.7** (Large Deviations Principle)**.** *Let $l \geq i_k$ and denote*

$$I(m_1, \ldots, m_l) := \begin{cases} -\log(m_{l+1}^+ - m_{l+1}^-) & , (m_1, \ldots, m_l) \in \mathcal{M}_l^{\mathcal{C}}([0, 1]) \\ \infty & , \text{else} \end{cases}. \tag{2.8}$$

*Then $(m_1^{(n)}, \ldots, m_l^{(n)})$ satisfies an LDP with speed $n$ and good rate function $I(\,\cdot\,) - I(m_1(\mu^{\mathcal{C}}), \ldots, m_l(\mu^{\mathcal{C}}))$.*

The previous theorem shows an LDP for the first $l$ random moments. By a projective limit argument it is possible to prove an LDP for the infinite random moment sequence. Since the Hausdorff moment problem is determinate, i.e. each probability measure on a compact interval is determined uniquely by its moment sequence, we can equivalently obtain an LDP for probability measures.

**Theorem 2.8** (Functional Large Deviations Principle). *Let $(\mu^{(n)})_n$ be a sequence of random probability measures in $\mathcal{P}([0,1])$ such that the vector of corresponding moments $(m_1(\mu^{(n)}), \ldots, m_n(\mu^{(n)}))$ is uniformly distributed in $\mathcal{M}_n^{\mathcal{C}}([0,1])$ for each $n$. Then $(\mu^{(n)})_n$ satisfies an LDP with speed $n$ and good rate function*

$$I(\mu) := \begin{cases} \mathcal{K}(\mu^0 \mid \mu) - \mathcal{K}(\mu^0 \mid \mu^{\mathcal{C}}) & , \mu^0 \ll \mu \text{ and } \mu \in \mathcal{P}^{\mathcal{C}}([0,1]) \\ \infty & , \text{else} \end{cases}$$

*on the space $\mathcal{P}([0,1])$ equipped with the weak topology, where the Kullback-Leibler divergence $\mathcal{K}$ is defined in* (2.2).

**Remark 2.9.** We have left the particular choice of the distribution of $\mu^{(n)}$ open and only demand that the distribution of its first $n$ moments is uniform on the $n$-th restricted space. This leaves many possible choices for the distribution of $\mu_n$. For example, $\mu^{(n)}$ could be chosen as the upper or lower principal representation of a random moment vector $(m_1^{(n)}, \ldots, m_n^{(n)})$, see e.g. Skibinsky (1986). There is a nice constructive approach to $\mu^{(n)}$ using spectral measures of Jacobi matrices which we will discuss in Section 3.1.

## 3. General distributions on constrained moment spaces and universality

While Section 2 dealt exclusively with the uniform distribution on $\mathcal{M}_n^{\mathcal{C}}([0,1])$, we will now turn to more general distributions on $\mathcal{M}_n^{\mathcal{C}}(E)$ for $E = [0,1]$ as well as the unbounded $E = \mathbb{R}_+$ and $E = \mathbb{R}$. Particular emphasis will be on universal behavior of the random moment sequences.

The first important observation towards general distributions on $\mathcal{M}_n^{\mathcal{C}}(E)$ consists in the fact that the ordinary moments $m_1, \ldots, m_n$ are not good coordinates to define probability measures on the moment spaces $\mathcal{M}_n(E)$. Indeed, the ordinary moments are not independent but strongly dependent and moreover the possible range for the $l + 1$-th moment given the first $l$ moments decreases exponentially in $l$ (cf. Karlin and Shapley (1953)). For these reasons, it is desirable to have a new system of independent coordinates that scale with the available moment range. For the bounded moment space $\mathcal{M}_n([0,1])$, such coordinates have been introduced by Skibinsky in a series of papers Skibinsky (1967, 1968, 1969). To ease notation, let us denote a vector in bold with a subscript indicating dimension, e.g.

$$\mathbf{m}_n := (m_1, \ldots, m_n).$$

For $m_j^+ \neq m_j^-$, the $j$-th canonical moment is defined as

$$p_j := \frac{m_j - m_j^-}{m_j^+ - m_j^-}. \tag{3.1}$$

They are left undefined if $m_j^+ = m_j^-$ in which case the corresponding moment sequence lies on the boundary of $\mathcal{M}_n([0,1])$. The canonical moment simply is the relative position of the ordinary moment in the available section. In fact, this construction of the canonical moments induces a smooth bijection

$$\varphi_n^{[0,1]} : \begin{cases} (0,1)^n & \to & \text{Int}\,\mathcal{M}_n([0,1]) \\ (p_1, \ldots, p_n) & \mapsto & (m_1, \ldots, m_n) \end{cases}$$

between the open unit cube $(0,1)^n$ and the interior of the $n$-th moment space.

On the moment space $\mathcal{M}_n([0,\infty))$ the range of the $(n+1)$-th moment is a half-open interval $[m_{n+1}^-, \infty)$ and a good system of coordinates is given by

$$z_j := \frac{m_j - m_j^-}{m_{j-1} - m_{j-1}^-} \tag{3.2}$$

with $m_0 := 1$, $m_0^- := 0$. These parameters are well-defined for all moment sequences in the interior of the moment space. Indeed, they yield a smooth bijection

$$\varphi_n^{[0,\infty)} : \begin{cases} (0,\infty)^n & \to & \operatorname{Int}\mathcal{M}_n([0,\infty)) \\ (z_1, \ldots, z_n) & \mapsto & (m_1, \ldots, m_n). \end{cases}$$

Finally, the moment space $\mathcal{M}_n(\mathbb{R})$ can be parametrized by the recurrence coefficients of orthogonal polynomials. Let for a vector $\mathbf{m}_n \in \mathcal{M}_n(\mathbb{R})$, $\mu \in \mathcal{P}(\mathbb{R})$ be a measure with the first $n$ moments given by $\mathbf{m}_n$. It is well-known that to each $\mu \in \mathcal{P}(\mathbb{R})$ there is a unique sequence of monic polynomials $(P_j)_j$ with $\deg P_j = j$ that are orthogonal in $L^2(\mathbb{R}, \mu)$. If $\mu$ has a finite support then the sequence of orthogonal polynomials is finite. $P_j$ is determined by the first $2j-1$ moments of $\mu$ which shows that different measures representing a moment sequence $\mathbf{m}_n$ have the same (first) orthogonal polynomials. These polynomials satisfy a three-term recursion

$$P_j(x) = (x - \alpha_j)P_{j-1}(x) - \beta_{j-1}P_{j-2}(x), \tag{3.3}$$

where $\alpha_j \in \mathbb{R}$ and $\beta_{j-1} > 0$. Moreover, by Favard's theorem (cf. Theorem I.4.4 in Chihara (1978)) each sequence of parameters $\alpha_j \in \mathbb{R}$, $\beta_j > 0$ yield a sequence of monic orthogonal polynomials $P_j$. This induces smooth bijections

$$\varphi_{2n}^{\mathbb{R}} : \begin{cases} (\mathbb{R} \times (0,\infty))^n & \to & \operatorname{Int}\mathcal{M}_{2n}(\mathbb{R}) \\ (\alpha_1, \beta_1, \alpha_2, \ldots, \alpha_n, \beta_n) & \mapsto & (m_1, \ldots, m_{2n}) \end{cases}$$

$$\varphi_{2n+1}^{\mathbb{R}} : \begin{cases} (\mathbb{R} \times (0,\infty))^n \times \mathbb{R} & \to & \operatorname{Int}\mathcal{M}_{2n+1}(\mathbb{R}) \\ (\alpha_1, \beta_1, \alpha_2, \ldots, \alpha_n, \beta_n, \alpha_{n+1}) & \mapsto & (m_1, \ldots, m_{2n+1}) \end{cases}.$$

In order to unify notation in the three different cases $E = [0,1], \mathbb{R}_+, \mathbb{R}$, we set

$$y_j := \begin{cases} p_j & , E = [0,1] \\ z_j & , E = [0,\infty) \\ \alpha_{(j+1)/2} & , E = \mathbb{R} \text{ and } j \text{ odd} \\ \beta_{j/2} & , E = \mathbb{R} \text{ and } j \text{ even} \end{cases} \qquad D_j := \begin{cases} (0,1) & , E = [0,1] \\ (0,\infty) & , E = [0,\infty) \\ \mathbb{R} & , E = \mathbb{R} \text{ and } j \text{ odd} \\ (0,\infty) & , E = \mathbb{R} \text{ and } j \text{ even} \end{cases}, \tag{3.4}$$

so that in all three cases

$$\varphi_n^E : \begin{cases} D_1 \times \cdots \times D_n & \to & \operatorname{Int}\mathcal{M}_n(E) \\ (y_1, \ldots, y_n) & \mapsto & (m_1, \ldots, m_n) \end{cases} \tag{3.5}$$

yields a smooth bijection. Throughout this paper we will call the $y_j$ *canonical coordinates*.

We will give the Jacobians of parameterizations of the constrained moment spaces in Lemma 4.1 below. At this stage it suffices to note the remarkable result

$$\left| \det\left[ \frac{\partial \varphi_n^{[0,1]}(p_1, \ldots, p_n)}{\partial(p_1, \ldots, p_n)} \right] \right| = \prod_{j=1}^n (p_j(1 - p_j))^{n-j}, \tag{3.6}$$

which is due to (Skibinsky, 1967, p. 1f). Consequently, we see that for the uniform distribution on $\mathcal{M}_n([0,1])$, the random canonical coordinates $p_1^{(n)}, \ldots, p_n^{(n)}$ have a density on $[0,1]^n$ proportional to the r.h.s. of (3.6). In other words, the canonical coordinates are independent and $p_j^{(n)}$ has a beta$(n-j+1, n-j+1)$ distribution. As we are interested mostly in the first $l$ moments as $n \to \infty$, we have $j \ll n$ and thus the canonical moments are independent and nearly identically distributed. The class of distributions introduced in Dette et al. (2018) and adapted here to the constrained spaces, generalizes from the uniform distribution on $\mathcal{M}_n([0,1])$ in three ways. Firstly, we include distributions on the unbounded spaces $\mathcal{M}_n(\mathbb{R}_+)$ and $\mathcal{M}_n(\mathbb{R})$. Secondly, we

generalize from the beta distribution to rather arbitrary densities while keeping the two key properties of independence and (nearly) identical distribution. And lastly, we allow for different densities for even and odd coordinates. This originates in the different roles played by even and odd moments. As even moments are always positive and contain some information about the size of the support of the measure, odd moments contain information about symmetries.

We can now introduce our general class of distributions on the constrained moment spaces. Recall that $\mathcal{C}$ is always assumed to be an admissible constraint for $\mathcal{P}(E)$.

Let $V_1 : \overline{D_1} \to \mathbb{R}, \ldots, V_{i_k+2} : \overline{D_{i_k+2}} \to \mathbb{R}$ be continuous functions satisfying in the cases $E = \mathbb{R}_+, \mathbb{R}$ the integrability conditions

$$V_j(y_j) \geq \begin{cases} (2+\varepsilon)\log|y_j|, & j=1,2, & E = \mathbb{R}_+, \\ (1+\varepsilon)\log|y_j|, & j=1, & E = \mathbb{R}, \\ (2+\varepsilon)\log|y_j|, & j=2, & E = \mathbb{R}, \end{cases} \tag{3.7}$$

for $|y_i|$ large enough and some $\varepsilon > 0$. Denote by

$$\mathbf{m}_n^{\mathcal{C}} := (m_j, 1 \leq j \leq n, j \neq i_1, \ldots, i_k) \tag{3.8}$$

the vector of unconstrained moments. Finally we set for notational convenience

$$V_{i_k+2j-1} := V_{i_k+1} \text{ and } V_{2i_k+2j} := V_{i_k+2}, \quad j \geq 2$$

and define the Borel probability measure $\mathbb{P}_{n,E,V}^{\mathcal{C}}$ on $M_n^{\mathcal{C}}(E)$ by $\mathbb{P}_{n,E,V}^{\mathcal{C}}(\partial \mathcal{M}_n^{\mathcal{C}}(E)) = 0$ and on the interior of $\mathcal{M}_n^{\mathcal{C}}(E)$ by the density

$$P_{n,E,V}^{\mathcal{C}}(\mathbf{m}_n) := \frac{1}{Z_{n,E,V}^{\mathcal{C}}} \exp\Big( -n \sum_{j=1}^n V_j(y_j(\mathbf{m}_n)) \Big)$$

w.r.t. the $(n-k)$-dimensional Lebesgue measure on $\mathcal{M}_n^{\mathcal{C}}(E)$ as defined at the beginning of Section 2. Here $y_j = y_j(\mathbf{m}_n)$, $j = 1, \ldots, n$, are the canonical coordinates associated to the moment sequence $\mathbf{m}_n = (m_1, \ldots, m_n)$ and

$$Z_{n,E,V}^{\mathcal{C}} := \int_{\mathcal{M}_n^{\mathcal{C}}(E)} \exp\Big( -n \sum_{j=1}^n V_j(y_j(\mathbf{m}_n)) \Big) d\mathbf{m}_n^{\mathcal{C}} \tag{3.9}$$

is the normalizing constant. $\mathbb{P}_{n,E,V}^{\mathcal{C}}$ is for the empty constraint $\mathcal{C} = \emptyset$ determined by $V_1$ and $V_2$ and is precisely the class of distributions found in Dette et al. (2018) showing universal behavior and classical limiting measures from random matrix theory or free probability theory. In the presence of a constraint influencing the first $i_k$ moments, the results of this section will show that $\mathbb{P}_{n,E,V}^{\mathcal{C}}$ is an appropriate class of distributions to study $\mathcal{M}_n^{\mathcal{C}}(E)$. More precisely, we will find universal behavior within this class for generic functions $V_1, \ldots, V_{i_k+2}$, given by special Bernstein-Szegő measures. We remark in passing that the product form of the density of the first $i_k$ moments is not necessary for observing the universal limits and could be extended to some function of the form $\exp(-nV(m_1, \ldots, m_{i_k}))$. However, the product form is convenient for notation and the class we consider is exhaustive in the sense that any universal limit law of the extended class can be observed in the smaller class.

As it is not immediate that $0 < Z_{n,E,V}^{\mathcal{C}} < \infty$, we formulate the following lemma which is proved at the beginning of Section 4.

**Lemma 3.1.** *For $V_1, \ldots, V_{i_k+2}$ satisfying (3.7), $\mathbb{P}_{n,E,V}^{\mathcal{C}}$ is a probability measure on $\mathcal{M}_n^{\mathcal{C}}(E)$.*

For the rest of this section, we will assume that $(m_1^{(n)}, \ldots, m_n^{(n)})$ has distribution $\mathbb{P}_{n,E,V}^{\mathcal{C}}$. Our first result in this setting is a large deviations principle which holds without any further

assumptions. Define the functions $W_j : D_j \to \mathbb{R}, j = 1, \ldots, n$ by

$$W_j(y_j) := \begin{cases} V_j(y_j) - \log(y_j(1-y_j)) & , E = [0,1], \\ V_j(y_j) - \log(y_j) & , E = [0,\infty), \\ V_j(y_j) & , E = \mathbb{R}, j \text{ odd}, \\ V_j(y_j) - \log(y_j) & , E = \mathbb{R}, j \text{ even}. \end{cases}$$

**Theorem 3.2** (Large Deviations Principle). *The vector* $(m_1^{(n)}, \ldots, m_l^{(n)})$ *satisfies a large deviations principle with speed* $n$ *and good rate function* $I_l^{\mathcal{C},E}(\,\cdot\,) - \inf\limits_{\mathbf{m}_l \in \mathcal{M}_l^{\mathcal{C}}(E)} I_l^{\mathcal{C},E}(\mathbf{m}_l)$, *where*

$$I_l^{\mathcal{C},E}(m_1, \ldots, m_l) := \begin{cases} \sum\limits_{j=1}^{l} W_j(y_j) & , (m_1, \ldots, m_l) \in \text{Int}\mathcal{M}_l^{\mathcal{C}}(E) \\ \infty & , \text{else} \end{cases}. \tag{3.10}$$

Usually a large deviations principle implies a law of large numbers, provided that the rate function has a unique minimizer. In the unconstrained case, all canonical coordinates are independent and uniqueness of the minimizer of the rate function reduces to uniqueness of the minimizers of $W_1$ and $W_2$. Hence uniqueness of a minimizing moment sequence is actually a univariate problem. In contrast to that, as a consequence of the constraint $\mathcal{C}$, in general the first $i_k$ canonical coordinates are strongly dependent and thus uniqueness of a minimizer is in general a truly multivariate problem, involving simultaneously $W_1, \ldots, W_{i_k+2}$ and $\mathcal{C}$. Let us illustrate this with an example for $\mathcal{M}_n^{\mathcal{C}}([0,1])$.

**Example 3.3.** The constraint $\mathcal{C} = \{m_1 = c\}$ for some $c \in (0,1)$ does not induce any dependencies, since $m_1 = p_1$. The simplest possible, yet instructive, constraint is $\mathcal{C} = \{m_2 = c\}$ for $c \in (0,1)$. From (3.1) we deduce $p_1 = m_1$ and $p_2 = \frac{c-p_1^2}{p_1(1-p_1)}$. Thus, changing coordinates to $p_1, p_3, p_4, \ldots, p_n$, the density $P_{n,[0,1],V}^{\mathcal{C}}$ can using a computation similar to (3.6) be expressed as

$$\frac{1}{Z_{n,E,V}^{\mathcal{C}}} \exp\left[-n\left(W_1(p_1) + W_2\left(\frac{c-p_1^2}{p_1(1-p_1)}\right)\right)\right] 1_{[c,\sqrt{c}]}(p_1) \tag{3.11}$$

$$\times \exp\left[-2\log(p_1(1-p_1)) - 2\log(cp_1 + cp_1^2 - c^2 - p_1^3) + 4\log(p_1(1-p_1))\right] \tag{3.12}$$

$$\times \exp\left(-n\sum_{j=3}^{n} W_j(p_j) - j\log((p_j)(1-p_j))\right). \tag{3.13}$$

The indicator function in (3.11) stems from the fact that $m_2 = c$ implies $m_1 = p_1 \in [c, \sqrt{c}]$. We will see a general version of (3.11)-(3.13) in Corollary 4.2 below.

All terms in (3.12) are sub-leading and do not contribute much to the uniqueness problem. (3.13) factorizes and thus uniqueness of minimization over $p_3, p_4, \ldots$ reduces to $W_3$ and $W_4$ separately. However, equation (3.11) shows that the uniqueness problem for $p_1$ involves $W_1, W_2$ and $\mathcal{C}$. Note that in this simple example minimization over $p_1$ is still one-dimensional, albeit non-trivial. The constraint $\mathcal{C} = \{m_3 = c\}$ would yield dependent $p_1, p_2$, and $p_3$ could be written as a rational function of $p_1$ and $p_2$. Note also that specifying to $V_1 = \cdots = V_4 = 0$, i.e. to the uniform distribution considered in Section 2, does not simplify things much.

From the previous example it is far from obvious why the uniform distribution on $\mathcal{M}_n^{\mathcal{C}}([0,1])$ concentrates on a unique moment sequence as $n \to \infty$. This uniqueness will for general $V_1, \ldots, V_{i_k+2}$ no longer be true. Fortunately, we can prove equally strong results without requiring uniqueness of the multivariate minimization problem. The assumptions on $V_1, \ldots, V_{i_k+2}$ used lateron are formulated in the definition below. As a preparation, let

$$\mathbf{y}_n^{\mathcal{C}} := (y_j, 1 \le j \le n, j \ne i_1, \ldots, i_k). \tag{3.14}$$

be the vector of canonical coordinates corresponding to $\mathbf{m}_n^{\mathcal{C}}$. It is shown in Lemma 4.1 of Section 4 that $\mathbf{y}_n^{\mathcal{C}}$ allows for a parametrization of $\mathcal{M}_n^{\mathcal{C}}(E)$. At this stage, it suffices to note that knowing

$\mathbf{y}_n^{\mathcal{C}}$ determines $\mathbf{m}_n$ because the "missing coordinates" $y_{i_j}, j = 1, \ldots, k$ are determined by $\mathbf{y}_n^{\mathcal{C}}$ and the constraint $\mathcal{C}$.

**Definition 3.4** (Generic $V_1, \ldots, V_{i_k+2}$)**.** Let $V_j \in C^2(D_j)$, $j = 1, \ldots, i_k + 2$ satisfy (3.7), if $E = \mathbb{R}_+$ or $E = \mathbb{R}$. We call $V_1, \ldots, V_{i_k+2}$ *generic*, if the following three conditions are satisfied:

- (unique univariate minimizers) $W_{i_k+1}$ and $W_{i_k+2}$ have unique minimizers $y_{i_k+1}^* \in D_{i_k+1}$ and $y_{i_k+2}^* \in D_{i_k+2}$, respectively,
- (finitely many multivariate minimizers) $\mathbf{y}_{i_k}^{\mathcal{C}} \mapsto I_{i_k}^{\mathcal{C},E}(\mathbf{m}_{i_k}(\mathbf{y}_{i_k}^{\mathcal{C}}))$ has finitely many minimizers

$$
\mathbf{y}^{*,1}, \ldots, \mathbf{y}^{*,p} \in \prod_{\substack{j=1 \\ j \neq i_1, \ldots, i_k}}^{i_k} D_j \,,
$$

  where the set $D_j$ has been defined in (3.4) and $I_{i_k}^{\mathcal{C},E}$ in (3.10).
- (non-degeneracy of minimizers) $W_j''(y_j^*) \neq 0$, $j = i_k + 1, i_k + 2$ and

$$
\det \operatorname{Hess}^{\mathcal{C}} \left( \sum_{j=1}^{i_k} W_j \right) (\mathbf{y}^{*,q}) \neq 0, \ q = 1, \ldots, p.
$$

  Here, $\operatorname{Hess}^{\mathcal{C}}$ is the $(i_k - k) \times (i_k - k)$ dimensional Hessian matrix with respect to the variables in $\mathbf{y}_{i_k}^{\mathcal{C}}$.

**Remark 3.5.** The set of $\mathbf{y}_{i_k}^{\mathcal{C}}$ such that $I_{i_k}^{\mathcal{C},E}(\mathbf{m}_{i_k}(\mathbf{y}_{i_k}^{\mathcal{C}})) \neq \infty$, i.e. the set of potential minimizers, is an open set on which $y_{i_1}, \ldots, y_{i_k}$ are smooth functions of $\mathbf{y}_{i_k}^{\mathcal{C}}$. On this set we have $I_{i_k}^{\mathcal{C},E}(\mathbf{m}_{i_k}(\mathbf{y}_{i_k}^{\mathcal{C}})) = \sum_{j=1}^{i_k} W_j(y_j)$. It is known from Morse theory that almost all $C^2$ functions have only non-degenerate critical points, meaning that the Hessian determinant does not vanish at these points. Furthermore, there can be only finitely many such points in any compact set. From a more practical point of view, if $V_1, \ldots, V_{i_k+2}$ are not generic, then the perturbation $\tilde{V}_1, \ldots, \tilde{V}_{i_k+2}$ with $\tilde{V}_j(t) := V_j(t) + a_j t, j = 1, \ldots, i_k + 2$ is generic for almost all values $a_1, \ldots, a_{i_k+2} \in \mathbb{R}$ (Matsumoto, 2002, Theorem 2.20). In the case of unbounded $E$ and thus unbounded $D_j$, the search for minimizers can effectively be restricted to a certain compact set thanks to the integrability conditions (3.7). This justifies calling $V_1, \ldots, V_{i_k+2}$ with the properties of Definition 3.4 generic.

Of course, almost all $C^2$ functions will also have a unique multivariate minimizer. Nevertheless we find it useful to consider the more general case of several multivariate minimizers as it features a particular aspect of the universality phenomenon (see Theorem 3.6 and Remark 3.7 below).

We can now formulate the main results of this section. The first one is an analog of Theorem 2.1 in the general setting. It is instructive to briefly review one of the main results of the unconstrained case: (Dette et al., 2018, Theorems 2.1, 2.5 and 2.7) show that the first $l$ random moments from $\mathbb{P}_{n,E,V}^{\mathcal{C}}$ with $\mathcal{C} = \emptyset$ and generic $V_1, V_2$ with minimizers $y_1^*, y_2^*$ converge to the first $l$ non-random moments of the probability measure $\mu_{y_1^*, y_2^*}$, where

$$
\mu_{y_1^*, y_2^*}(dx) := \begin{cases}
\left(1 - \frac{p_1^*}{p_2^*}\right)_+ \delta_0 + \left(\frac{p_1^* + p_2^* - 1}{p_2^*}\right)_+ \delta_1 + \frac{\sqrt{(x-a)(b-x)}}{2\pi p_2^* x(1-x)} \mathbb{1}_{[a,b]}(x) dx & , \quad \text{if } E = [0,1], \\[2mm]
\left(1 - \frac{z_1^*}{z_2^*}\right)_+ \delta_0 + \frac{1}{2\pi z_2^*} \frac{\sqrt{(x-a)(b-x)}}{x} \mathbb{1}_{[a,b]}(x) dx & , \quad \text{if } E = \mathbb{R}_+, \\[2mm]
\frac{1}{2\pi \beta^*} \sqrt{(x-a)(b-x)} \mathbb{1}_{[a,b]}(x) dx & , \quad \text{if } E = \mathbb{R}.
\end{cases} \tag{3.15}
$$

Here, $(\cdot)_+ := \max(\cdot, 0)$ and $a$ and $b$ are constants depending on $y_1^*, y_2^*$ and $E$ only. To ease notation, we will with a slight abuse of notation not index the measure with $E$ and instead trust that the reader will distinguish the three cases $E = [0,1], \mathbb{R}_+, \mathbb{R}$ from the appearance of $p$'s, $z$'s or $\alpha, \beta$, respectively. The measure $\mu_{p_1^*, p_2^*}$ is called Kesten-McKay measure or free binomial

distribution as it is a free convolution power of the Bernoulli distribution. We refer to (Dette et al., 2018, p. 4f) for details. The measures $\mu_{z_1^*, z_2^*}$ and $\mu_{\alpha^*, \beta^*}$ are the Marchenko-Pastur and semicircle distributions, respectively, which are well-known in random matrix theory and free probability. The universality of the three measures is apparent as very different functions $V_1$, $V_2$ lead to the same family of measures.

Let us now formulate an analogous result in the presence of a constraint $\mathcal{C}$ that identifies the limit measures as members of the Bernstein-Szegő class plus some discrete measure. To state it, define

$$y_1^* := \begin{cases} y_{i_k+1}^*, & \text{if } i_k \text{ is even,} \\ y_{i_k+2}^*, & \text{if } i_k \text{ is odd,} \end{cases} \qquad y_2^* := \begin{cases} y_{i_k+2}^*, & \text{if } i_k \text{ is even,} \\ y_{i_k+1}^*, & \text{if } i_k \text{ is odd.} \end{cases} \tag{3.16}$$

**Theorem 3.6** (Law of Large Numbers)*. Let $V_1, \ldots, V_{i_k+2}$ be generic. Then we have as $n \to \infty$*

$$(m_1^{(n)}, \ldots, m_l^{(n)}) \to \sum_{q=1}^p w_q \delta_{\mathbf{m}_l^{*,q}}$$

*in distribution, where $w_1, \ldots, w_p > 0$, $\sum_{q=1}^p w_q = 1$ are weights and $\mathbf{m}_l^{*,q} := (m_1^{*,q}, \ldots, m_l^{*,q})$ is the vector of the first $l$ moments of a probability measure*

$$\mu_q(dx) = \mu_q^{ac}(x)dx + \mu_q^d(dx) \,.$$

*The density of the absolutely continuous part is given by*

$$\mu_q^{ac}(x) = \frac{1}{D_q(x)} \mu_{y_1^*, y_2^*}^{ac}(x) \,,$$

*where $\mu_{y_1^*, y_2^*}^{ac}$ is the density of the absolutely continuous part of the measure $\mu_{y_1^*, y_2^*}$ defined in (3.15) and $D_q$ is a polynomial of degree at most $i_k$, strictly positive on the support of $\mu_{y_1^*, y_2^*}^{ac}$. The measure $\mu_q^d$ is discrete having atoms at the positions of atoms of $\mu_{y_1^*, y_2^*}$ and at most $i_k$ extra atoms.*

**Remark 3.7.** Theorem 3.6 shows how strong the universality of (3.15) is: The limiting measure needs to fulfil the constraint $\mathcal{C}$ and, apart from atoms, the optimal measure from $\mathcal{P}^{\mathcal{C}}(E)$ is absolutely continuous w.r.t. $\mu_{y_1^*, y_2^*}$. In particular, the support of the absolutely continuous part of $\mu_q$, $q = 1, \ldots, p$ is always the one of $\mu_{y_1^*, y_2^*}$, regardless of the constraint $\mathcal{C}$! Moreover, as nicely featured in the case of several minimizers, this universality is not even broken if there are several limiting measures.

**Example 3.8.** The proof of Theorem 3.6 is constructive: A recipe to determine the weights $w_q$, polynomials $D_q$ and measures $\mu_q^d$ is given in Proposition 4.3 and Poposition 4.4. As an example we consider the case $E = \mathbb{R}$ with the constraint $\mathcal{C} := \{m_1 = 0\}$, i.e. we only consider measures with mean zero. We take the functions $V_1(\alpha) := (\alpha - 1)^2$ and $V_2(\beta) := 8\beta^2$. Then $W_1(\alpha) = V_1(\alpha)$ has a unique minimizer in $\alpha^* = 1$ and $W_2(\beta) = V_2(\beta) - \log(\beta)$ in $\beta^* = \frac{1}{4}$. The limiting measure is

$$\mu_{\alpha^*, \beta^*}(dx) = \frac{\sqrt{x(2-x)}}{2\pi\left(x + \frac{1}{4}\right)} 1_{[0,2]}(x) \, dx + \frac{3}{4}\delta_{-\frac{1}{4}} \,. \tag{3.17}$$

The computation of this measure will be performed in Example 4.5 following Proposition 4.4, where the necessary preliminaries are given.

Our next result concerns Gaussian fluctuations. A remarkable fact is that in the case of several minimizers, the standardization is itself random.

**Theorem 3.9** (Central Limit Theorem)*. Let $V_1, \ldots, V_{i_k+2}$ be generic. Define*

$$\mathbf{m}_l^* := \underset{\mathbf{m}^* \in \{\mathbf{m}_l^{*,1}, \ldots, \mathbf{m}_l^{*,p}\}}{\arg\min} \|(m_1^{(n)}, \ldots, m_l^{(n)}) - \mathbf{m}^*\|,$$

where $\mathbf{m}_l^{*,q}$, $q = 1, \ldots, p$ have been introduced in Theorem 3.6 and $\|\cdot\|$ denotes the Euclidean norm. Then, as $n \to \infty$

$$Z_n := \sqrt{n}\Sigma_l(\mathbf{m}_l^*)((m_1^{(n)}, \ldots, m_l^{(n)}) - \mathbf{m}_l^*) \to \mathcal{N}(0, L) \qquad (3.18)$$

in distribution, where $L \in \mathbb{R}^{l \times l}$ is the matrix with $L_{uv} = 1$ if $u = v$ and $v \notin \{i_1, \ldots, i_k\}$ and $L_{uv} = 0$ else. The matrix $\Sigma_l \in \mathbb{R}^{l \times l}$ is given by

$$\Sigma_l(\mathbf{m}_l) := T^{\mathcal{C}}\left[\left(\mathrm{Hess}^{\mathcal{C}} \sum_{j=1}^{l} W_j\right)^{1/2}(\mathbf{y}_l^{\mathcal{C}})\left(\frac{\partial \mathbf{y}_l^{\mathcal{C}}}{\partial \mathbf{m}_l^{\mathcal{C}}}(\mathbf{m}_l)\right)^t\right],$$

where $T^{\mathcal{C}} : \mathbb{R}^{(l-k) \times (l-k)} \to \mathbb{R}^{l \times l}$ denotes insertion of rows and columns of zeros at the positions $i_1, \ldots, i_k$, $\mathrm{Hess}^{\mathcal{C}}$ denotes the Hessian with respect to the coordinates in $\mathbf{y}_l^{\mathcal{C}}$ and the variables $\mathbf{m}_l^{\mathcal{C}}$, $\mathbf{y}_l^{\mathcal{C}}$ are defined in (3.8) and (3.14), respectively.

We finish this section with a moderate deviations principle.

**Theorem 3.10** (Moderate Deviations Principle)**.** *Let* $V_1, \ldots, V_{i_k+2}$ *be generic and* $a_n \to \infty$ *be a sequence with* $a_n = o(\sqrt{n})$. *Then the sequence* $(Z_n)_n$ *in* (3.18) *satisfies a large deviations principle with speed* $\frac{n}{a_n^2}$ *and good rate function*

$$I(\mathbf{m}_l) := \begin{cases} \frac{1}{2}\|\mathbf{m}_l\|_2^2 & , m_{i_1} = \ldots = m_{i_k} = 0 \\ \infty & , else \end{cases}.$$

### 3.1. Universality and Connection to Random Matrix Theory.

There are several universal aspects of the results in this section (and Section 2). The first one is the occurrence of Gaussian fluctuations, which is generic for uniform distributions on convex bodies, see Klartag (2007). As already briefly mentioned in the introduction, it is however surprising to find Gaussian fluctuations for vectors containing only finitely many coordinates (moments). This is a sign of the strong dependence between moments.

Even more interesting is the universality of the shape of the limiting measures $\mu_q, q = 1, \ldots, p$ of Theorem 3.6. Their absolutely continuous parts are all of the form reciprocal of a polynomial times the universal $\mu_{y_1^*, y_2^*}^{ac}$ and thus depend on the constraint $\mathcal{C}$ and few values of the functions $V_1, \ldots, V_{i_k+2}$ only.

As shown in Remark 2.2, the limiting measures for constrained moment spaces are generally not equilibrium measures, in contrast to the ones obtained in unconstrained spaces. Let us now shed some more light on this phenomenon. Consider a vector $\mathbf{m}_n \in \mathcal{M}_{2n+1}(\mathbb{R})$ with the corresponding canonical coordinates, i.e. recurrence coefficients, and form the tridiagonal Jacobi matrix

$$J := \begin{pmatrix} \alpha_1 & \sqrt{\beta_1} & & & \\ \sqrt{\beta_1} & \alpha_2 & \sqrt{\beta_2} & & \\ & \sqrt{\beta_2} & \alpha_3 & \sqrt{\beta_3} & \\ & & \ddots & \ddots & \ddots & \\ & & & \sqrt{\beta_{2n-1}} & \alpha_{2n} & \sqrt{\beta_{2n}} \\ & & & & \sqrt{\beta_{2n}} & \alpha_{2n+1} \end{pmatrix}.$$

Note that $J$ is symmetric and thus diagonalizable with real eigenvalues $x_1, \ldots, x_n$ and eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$. The probability measure

$$\mu^{(n)} := \sum_{j=1}^{n} \langle \mathbf{v}_j, \mathbf{e} \rangle \delta_{x_j}$$

is called spectral measure to $J$ and the first standard basis vector $\mathbf{e}$, where $\langle \cdot, \cdot \rangle$ denotes the Euclidean scalar product. It has the property that its $l$-th moment is $m_l$ which is by the spectral

14

theorem simply the $(1,1)$-entry of $J^l$, $l = 1,\ldots,n$ (see (Simon, 2011, Chapters 1.2, 1.3)). In particular, if $(m_1^{(n)}, \ldots, m_n^{(n)})$ is uniformly distributed in $\mathcal{M}_n^{\mathcal{C}}([0,1])$, then the associated random spectral measure $\mu^{(n)}$ fulfills the assumptions of Theorem 2.8. Our laws of large numbers for random moments imply that for $(m_1^{(n)}, \ldots, m_n^{(n)})$ with distribution $\mathbb{P}_{n,E,V}^{\mathcal{C}}$, the associated random spectral measure $\mu^{(n)}$ converges weakly in distribution towards the limiting measure $c^{-1} \sum_{q=1}^p \lambda_q \mu_q$ from Theorem 3.6.

To make the connection to equilibrium measures and random matrices, we remark that in random matrix theory equilibrium measures typically occur as limits of *empirical* spectral measures

$$\nu_n := \frac{1}{n} \sum_{j=1}^n \delta_{x_j},$$

where $x_1, \ldots, x_n$ are the eigenvalues of some random $n \times n$ matrix. An important class of random matrix ensembles, the so-called $\beta$-ensembles, has eigenvalue densities proportional to

$$e^{\frac{\beta}{2} \sum_{i \neq j} \log|x_i - x_j| - n \sum_{j=1}^n Q_{ex}(x_j)} = e^{-n^2 \left( \frac{\beta}{2} \int_{\neq} \int \log|t-s|^{-1} d\nu_n(t) d\nu_n(s) + \int Q_{ex}(t) d\nu_n(t) \right)}, \quad (3.19)$$

where $\int_{\neq} \int$ means that the diagonal is excluded in the integral, $\beta > 0$ and $Q_{ex}$ is some function called external field. In view of (3.19), it is not surprising that for $n \to \infty$, the ensemble realizes eigenvalue configurations that approach the equilibrium measure which is a solution to minimizing (2.3), see e.g. (Johansson, 1998, Theorem 2.1). In fact, a large deviations principle can be obtained for this approximation of the equilibrium measure, cf. (Anderson et al., 2010, Theorem 2.6.1). For our purposes, it is important to note that the speed of this large deviations principle is $n^2$, a fact that is readily read off (3.19). Now, apparently the difference between the spectral measure connected to the random moment problem and the empirical spectral measure from random matrix theory, is the weighting of the atoms. In contrast to $\nu_n$, the weights $w_j := \langle \mathbf{v}_j, \mathbf{e}_1 \rangle$ of the spectral measure $\mu^{(n)}$ depend on the eigenvectors and are of course random. In certain cases, more can be said: For example, for $\mathbf{m}_{2n-1}^{(n)}$ uniform on $\mathcal{M}_{2n-1}([0,1])$, the eigenvalues of the associated random matrix $J$ are distributed according to (3.19) with $\beta = 4$ and $Q_{ex}(t) = 0$ for $t \in [0,1]$ and $\infty$ elsewhere, see (Killip and Nenciu, 2004, Theorem 2.2). Moreover, the weights are independent from the eigenvalues and have a Dirichlet$(2,\ldots,2)$-distribution on the simplex $\{\mathbf{w}_n \in [0,1]^n : \sum_{j=1}^n w_j = 1\}$. For $n \to \infty$, the weights concentrate around the barycenter $(1/n, \ldots, 1/n)$. However, the speed is $n$ and thus the concentration of the weights is weaker than that of the eigenvalues. More precisely, it was shown in Gamboa and Rouault (2010) that the random spectral measure corresponding to the uniform distribution on $\mathcal{M}_n([0,1])$ satisfies an LDP with speed $n$ and good rate function given by the reversed Kullback-Leibler divergence $I(\mu) = \mathcal{K}(\mu^0|\mu)$. In view of the close connection between spectral measure and the random moment problem, this also explains the occurence of the Kullback-Leibler divergence in Theorem 2.1. However, the question remains why the unconstrained equilibrium measure $\mu^0$ still appears instead of the measure solving the constrained equilibrium problem. To understand this, let us consider what happens if we force the spectral measure to fulfil a constraint $\mathcal{C}$. Deviations of the eigenvalue configuration from the energy minimizing equilibrium measure (or rather its discrete analog, the Fekete points) are far more costly than deviations of the weights from $(1/n, \ldots, 1/n)$. Therefore the weights have to change dramatically from $(1/n, \ldots, 1/n)$ to some other values in order to bend the "discretized arcsine measure" to match the constraint. This results e.g. in a limiting measure $\mu^{\mathcal{C}}$ that is supported on the full set $[0,1]$ like the arcsine measure, as observed in Remark 2.2.

## 4. Parametrizations and Proof of Law of Large Numbers

The first step towards proving the results of Sections 2 and 3 is a parametrization of the constrained space $\mathcal{M}_n^{\mathcal{C}}(E)$. Define the set $A_{i_k} \subset \prod_{j=1}^{i_k} D_j$ as

$$A_{i_k} := \left(\varphi_{i_k}^E\right)^{-1}\left(\text{relint}\,\mathcal{M}_{i_k}^{\mathcal{C}}(E)\right),$$

where $\varphi_{i_k}^E$ is the parametrization (3.5) of the unconstrained moment space and relint denotes the relative interior, i.e. $\text{relint}\,\mathcal{M}_{i_k}^{\mathcal{C}}(E) := \mathcal{M}_{i_k}^{\mathcal{C}}(E) \cap \text{Int}\,\mathcal{M}_{i_k}(E)$. Define $\tilde{A}_{i_k}$ as the projection of $A_{i_k}$ to the coordinates in $\mathbf{y}_{i_k}^{\mathcal{C}}$ and $\tilde{\mathcal{M}}_n^{\mathcal{C}}(E)$ as the projection of $\mathcal{M}_n^{\mathcal{C}}(E)$ to the coordinates in $\mathbf{m}_n^{\mathcal{C}}$. Then we have the following lemma.

**Lemma 4.1.** *For any $n \geq i_k$ the mapping*

$$\varphi_n^{E,\mathcal{C}} : \begin{cases} \tilde{A}_{i_k} \times D_{i_k+1} \times \cdots \times D_n & \rightarrow & \text{Int}\,\tilde{\mathcal{M}}_n^{\mathcal{C}}(E) \\ \mathbf{y}_n^{\mathcal{C}} & \mapsto & \mathbf{m}_n^{\mathcal{C}} \end{cases}$$

*is a $C^\infty$-diffeomorphism with Jacobian*

$$\left|\det\left[\frac{\partial \mathbf{m}_n^{\mathcal{C}}}{\partial \mathbf{y}_n^{\mathcal{C}}}\right]\right| := \left|\det\left[\frac{\partial \varphi_n^{E,\mathcal{C}}(\mathbf{y}_n^{\mathcal{C}})}{\partial \mathbf{y}_n^{\mathcal{C}}}\right]\right| = \begin{cases} \prod_{j=1}^{n}(p_j(1-p_j))^{n-j-d_{j,\mathcal{C}}} & ,E=[0,1], \\ \prod_{j=1}^{n-1} z_j^{n-j-d_{j,\mathcal{C}}} & ,E=[0,\infty), \\ \prod_{j=1}^{\lfloor n/2 \rfloor} \beta_j^{n-2j-d_{j,\mathcal{C}}} & ,E=\mathbb{R}, \end{cases}$$

*where $d_{j,\mathcal{C}} := \#\{l \mid i_l > j\}$ and $\lfloor n/2 \rfloor$ is the largest integer smaller or equal $n/2$.*

*Proof.* It is straightforward to see that $\varphi_n^{E,\mathcal{C}}$ is indeed a diffeomorphism and we only need to calculate its Jacobian. Note that the Jacobian matrix is lower triangular and thus its determinant is given by the product of the entries on the diagonal

$$\det \varphi_n^{E,\mathcal{C}} = \prod_{\substack{1 \leq j \leq n \\ j \notin \{i_1,\ldots,i_k\}}} \frac{\partial m_j}{\partial y_j}. \tag{4.1}$$

In the case $E=[0,1]$, rearranging (3.1) yields

$$m_j = p_j(m_j^+ - m_j^-) - m_j^-.$$

As $m_j^\pm$ only depends on $m_1,\ldots,m_{j-1}$, we obtain

$$\frac{\partial m_j}{\partial p_j} = m_j^+ - m_j^-$$

and the assertion follows by an application of the formula (cf. Skibinsky (1967))

$$m_l^+ - m_l^- = \prod_{j=1}^{l-1} p_i(1-p_i) \tag{4.2}$$

and rearranging the order of multiplication in (4.1).

The proof in the case $E=\mathbb{R}_+$ is analogous. Note that formula (3.2) yields

$$m_j = z_j(m_{j-1} - m_{j-1}^-) + m_j^-$$

and consequently

$$\frac{\partial m_j}{\partial z_j} = m_{j-1} - m_{j-1}^- = \prod_{l=1}^{j-1} \frac{m_l - m_l^-}{m_{l-1} - m_{l-1}^-} = z_1 \cdots z_{j-1}.$$

16

In the case $E = \mathbb{R}$, the calculation of the partial derivatives is slightly more complicated. As in the proof of Lemma 2.6 in Dette et al. (2018), the partial derivatives of the moments with respect to the recursion coefficients can be calculated as

$$\frac{\partial m_{2j-1}}{\partial \alpha_j} = \beta_1 \cdots \beta_{j-1},$$

$$\frac{\partial m_{2j}}{\partial \beta_j} = \beta_1 \cdots \beta_{j-1}.$$

The assertion again follows by plugging in these formulas into (4.1) and rearranging the order of multiplication. □

We are now in the position to prove Lemma 3.1.

*Proof of Lemma 3.1.* We have to show that (3.9) is finite, positivity is trivial. The compact case $E = [0, 1]$ is clear. From the unbounded cases we exemplarily consider $E = \mathbb{R}_+$. Changing to canonical coordinates we see that

$$Z_{n,E,V}^{\mathcal{C}} = \int_{\mathbb{R}_+^{n-k}} 1_{\tilde{A}_{i_k}}(\mathbf{z}_{i_k}^{\mathcal{C}}) \exp\left(-n \sum_{j=1}^{n} \left(V_j(z_j) - \frac{n-j-d_{j,\mathcal{C}}}{n} \log z_j\right)\right) d\mathbf{z}_n^{\mathcal{C}}.$$

Let us drop the indicator $1_{\tilde{A}_{i_k}}(\mathbf{z}_{i_k}^{\mathcal{C}})$ from the integral, thereby making it larger, and consider first the $z_{i_j}$, $j = 1, \ldots, k$. Note that by (3.7), there is a constant $c_1 \in \mathbb{R}$ such that we have for $z_{i_j} > 1$

$$V_j(z_j) - \frac{n-j-d_{j,\mathcal{C}}}{n} \log z_j \geq c_1.$$

Thus, for those $\mathbf{z}_n^{\mathcal{C}}$ with $z_{i_j} > 1$ we can use the bound

$$\exp\left(-n\left(V_{i_j}(z_{i_j}) - \frac{n-i_j-d_{i_j,\mathcal{C}}}{n} \log z_{i_j}\right)\right) \leq e^{-nc_1}.$$

By continuity we have for $0 \leq z_{i_j} \leq 1$ for some $c_2 \in \mathbb{R}$ the bound $V_{i_j}(z_{i_j}) \geq c_2$ and hence for those $\mathbf{z}_n^{\mathcal{C}}$ with $0 \leq z_{i_j} \leq 1$, we may use

$$\exp\left(-n\left(V_{i_j}(z_{i_j}) - \frac{n-i_j-d_{i_j,\mathcal{C}}}{n} \log z_{i_j}\right)\right) \leq e^{-nc_2}.$$

This shows

$$Z_{n,E,V}^{\mathcal{C}} \leq e^{-kn \min(c_1,c_2)} \int_{\mathbb{R}_+^{n-k}} \exp\left(-n \sum_{\substack{j=1 \\ j \neq i_1,\ldots,i_k}}^{n} \left(V_j(z_j) - \frac{n-j-d_{j,\mathcal{C}}}{n} \log z_j\right)\right) d\mathbf{z}_n^{\mathcal{C}},$$

which factorizes and is integrable by assumption (3.7). □

As a direct consequence of Lemma 4.1 we obtain the following

**Corollary 4.2.** $\mathbb{P}_{n,E,V}^{\mathcal{C}}$ *induces via* $\left(\varphi_n^{E,\mathcal{C}}\right)^{-1}$ *a probability measure on* $\prod_{\substack{1 \leq j \leq n \\ j \notin \{i_1,\ldots,i_k\}}} D_j$ *with density*

$$\tilde{P}_{n,E,V}^{\mathcal{C}}(\mathbf{y}_n^{\mathcal{C}}) := \frac{1}{Z_{n,E,V}^{\mathcal{C}}} \exp\left(-n \sum_{j=1}^{n} W_j(y_j)\right) 1_{\tilde{A}_{i_k}}(\mathbf{y}_{i_k}^{\mathcal{C}}) \tag{4.3}$$

$$\times \exp\left(\sum_{j=1}^{n} (W_j(y_j) - V_j(y_j))(j + d_{j,\mathcal{C}})\right) \tag{4.4}$$

*w.r.t. the $(n-k)$-dimensional Lebesgue measure.*

Note that in (4.3) and (4.4) the canonical coordinates $y_{i_1}, \ldots, y_{i_k}$ corresponding to the constrained moments $m_{i_1}, \ldots, m_{i_k}$ are functions of $\mathbf{y}_n^{\mathcal{C}}$. More precisely, $y_{i_j}$ depends on all coordinates $y_d$ from $\mathbf{y}_n^{\mathcal{C}}$ with $d < i_j$ and is determined by the requirement $m_{i_j} = c_{i_j}$, $j = 1, \ldots, k$.

We can see from (4.3) and (4.4) that the coordinates in $\mathbf{y}_{i_k}^{\mathcal{C}}$ are dependent. On the one hand, they are coupled via the indicator function. On the other hand, even apart from the indicator, in general the density does not factorize as $y_{i_1}, \ldots, y_{i_k}$ are functions of the lower canonical coordinates.

Corollary 4.2 allows for the simple but important observation that the canonical coordinates $\mathbf{y}_n^{\mathcal{C}}$ have a density of the form

$$\mathbb{P}_n(dx) = \frac{1}{Z_n} e^{-nW(x)} R(x), \tag{4.5}$$

where $W, R : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ are some measurable functions and $Z_n$ is the normalization constant.

On the level of canonical coordinates, probabilistic statements can be deduced directly from the form of the density (4.5). We will use in Section 7 a different parametrization and thus obtain functions $W$ and $R$ different from those needed here to prove the results of Section 3. To avoid duplication, we therefore state and prove a proposition valid for rather general $W$ and $R$. It might also be of independent interest.

Assume from here on that $W : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ and $R : \mathbb{R}^m \to \mathbb{R}_+$ are measurable and such that

$$0 < \int e^{-n_0 W(x)} R(x) \, dx < \infty \quad \text{for some } n_0 \in \mathbb{N}. \tag{4.6}$$

**Proposition 4.3** (Convergence to discrete distribution). *Let $W$ and $R$ satisfy the following conditions:*

(1) *$W$ attains its global minimum exactly in the points $\theta_1, \ldots, \theta_p$, i.e. $W(x) > W(\theta_1)$ for $x \in \mathbb{R}^m \setminus \{\theta_1, \ldots, \theta_p\}$ and $W(\theta_1) = \ldots = W(\theta_p)$.*

(2) *$W$ is twice continuously differentiable in a neighborhood of each $\theta_q$.*

(3) *For all $\varepsilon > 0$ and $M_\varepsilon := \bigcup_{q=1}^p B_\varepsilon(\theta_q)$ we have $\inf_{x \in M_\varepsilon^{\mathcal{C}}} W(x) > W(\theta_q)$, where $B_\varepsilon(\theta_q)$ denotes the open $\varepsilon$-ball around $\theta_q$.*

(4) *$\operatorname{Hess} W(\theta_q)$ is nonsingular for each $1 \leq q \leq p$.*

(5) *$R$ is continuous in all $\theta_q$ and does not vanish in all $\theta_q$ simultaneously.*

*Then $\mathbb{P}_n$ from (4.5) converges as $n \to \infty$ weakly to the distribution*

$$\mathbb{P} = \left( \sum_{q=1}^p R(\theta_q) \det \operatorname{Hess} W(\theta_q)^{-1/2} \right)^{-1} \sum_{q=1}^p R(\theta_q) \det \operatorname{Hess} W(\theta_q)^{-1/2} \delta_{\theta_q}.$$

*Proof.* We may assume without loss of generality that $W(\theta_1) = \ldots = W(\theta_p) = 0$. Let $U \subset \mathbb{R}^m$ be an arbitrary open set and $\varepsilon > 0$ so small, that the following conditions hold:

(i) If $\theta_q \in U$ holds, then also $B_\varepsilon(\theta_q) \subset U$.

(ii) $W$ is twice differentiable on $B_\varepsilon(\theta_q)$.

(iii) $\operatorname{Hess} W(x) \geq M_q$ for all $x \in B_\varepsilon(\theta_q)$ and some positive definite matrices $M_q$, where $\geq$ stands for the Löwner (partial) order, i.e. $M \geq N$ means that $M - N$ is positive semidefinite.

To see that the third condition is attainable, note that $\operatorname{Hess} W(\theta_q)$ is positive semidefinite, since $\theta_q$ is an absolute minimum of $W$. Since $\operatorname{Hess} W(\theta_q)$ is nonsingular, the matrix is even positive definite. By Weyl's inequality (see Theorem 1 in Section 6.7 of Franklin (1968)) there is a $\delta > 0$ so that $\operatorname{Hess} W(\theta_q) - \delta I_m$ is positive definite. Since $\operatorname{Hess} W$ is twice continuously differentiable

near all $\theta_q$, we may choose $\varepsilon > 0$ such that

$$\left| \frac{\partial^2 W}{\partial x_i \partial x_j}(x) - \frac{\partial^2 W}{\partial x_i \partial x_j}(\theta_q) \right| < \frac{\delta}{2m}$$

is satisfied for all $x \in B_\varepsilon(\theta_q)$ and $1 \le i, j \le m$. By Gerschgorin's theorem (see Theorem 1 in Section 6.8 of Franklin (1968)) this implies that all eigenvalues of $\operatorname{Hess} W(x) - \operatorname{Hess} W(\theta_q)$ have absolute value less than $\frac{\delta}{2}$. From Weyl's inequality we can conclude that all eigenvalues of $\operatorname{Hess} W(x) - \operatorname{Hess} W(\theta_q) + \delta I_m$ must be at least $\frac{\delta}{2}$. This means that in the Löwner order the inequalities

$$\operatorname{Hess} W(x) > \operatorname{Hess} W(\theta_q) - \delta I_m > 0$$

hold for all $x \in B_\varepsilon(\theta_q)$ and we may choose $M_i = \operatorname{Hess} W(\theta_q) - \delta I_m$ in condition (iii).

Now, with $M_\varepsilon$ from condition (3) of the proposition,

$$\mathbb{P}_n(U) \ge \frac{\sum\limits_{q=1}^p 1_U(\theta_q) \int_{B_\varepsilon(\theta_q)} e^{-nW(x)} R(x)\, dx}{\int_{(M_\varepsilon)^c} e^{-nW(x)} R(x)\, dx + \sum\limits_{q=1}^p \int_{B_\varepsilon(\theta_q)} e^{-nW(x)} R(x)\, dx}.$$

By a standard application of Laplace's method, we have as $n \to \infty$

$$\int_{B_\varepsilon(\theta_q)} e^{-nW(x)} R(x)\, dx = n^{-m/2} \left( R(\theta_q) \frac{(2\pi)^{m/2}}{\sqrt{\det \operatorname{Hess} W(\theta_q)}} + o(1) \right).$$

Furthermore, with $K := \inf\limits_{x \in (M_\varepsilon)^c} W(x) > 0$ we obtain

$$\int_{(M_\varepsilon)^c} e^{-nW(x)} R(x)\, dx \le e^{-(n-n_0)K} \int_{(M_\varepsilon)^c} e^{-n_0 W(x)} R(x)\, dx = o(n^{-m/2}),$$

implying

$$P_n(U) \ge \frac{\left(\frac{2\pi}{n}\right)^{m/2} \left( \sum\limits_{q=1}^p 1_U(\theta_q) R(\theta_q) \det \operatorname{Hess} W(\theta_q)^{-1/2} + o(1) \right)}{\left(\frac{2\pi}{n}\right)^{m/2} \left( \sum\limits_{q=1}^p R(\theta_q) \det \operatorname{Hess} W(\theta_q)^{-1/2} + o(1) \right)} \xrightarrow{n \to \infty} \mathbb{P}(U).$$

By Portmanteau's theorem, this implies $\mathbb{P}_n \xrightarrow{d} \mathbb{P}$. $\qquad \square$

We will see that applied to our random moment problem, Proposition 4.3 shows convergence of the random canonical coordinates $\mathbf{y}_n^{\mathcal{C},(n)}$ to some discrete distribution concentrated on the moment sequences of certain limiting measures. The main task now is to identify these measures from the information on their canonical coordinates. To this end, we make use of the following proposition which gives information on probability measures with eventually constant recurrence coefficients. We remark that similar results are known, see e.g. (Geronimo and Iliev, 2017, Theorem 2.1) and references therein, yet they are not sufficient for our purposes.

**Proposition 4.4.** *Let $\mu$ be a probability measure with recursion coefficients $\alpha_1, \ldots, \alpha_j$ and $\beta_1, \ldots, \beta_{j-1} > 0$, such that $\alpha_l = \alpha$ and $\beta_{l-1} = \beta > 0$ for all $l > j$. Then the Lebesgue decomposition of $\mu$ consists of an absolutely continuous part $\mu^{ac}$ and a discrete part $\mu^d$. The absolutely continuous part $\mu^{ac}$ has the density*

$$\frac{\sqrt{4\beta - (x-\alpha)^2}\,\beta_1 \cdots \beta_{j-1}}{2\pi D(x)}$$

19

on the interval $I_{\alpha,\beta} := [\alpha - 2\sqrt{\beta}, \alpha + 2\sqrt{\beta}]$, where $D(x) := P_j^2(x) - P_{j+1}(x)P_{j-1}(x)$ and $P_i$, $i \geq 0$ are the monic orthogonal polynomials corresponding to $\mu$. To leading order (with the convention $\beta_0 = 0$),

$$D(x) = (\alpha - \alpha_j)x^{2j-1} + \left[(\beta - \beta_{j-1}) + \big(2(\alpha_1 + \cdots + \alpha_{j-1}) + \alpha_j\big)(\alpha_j - \alpha)\right]x^{2j-2} + O(x^{2j-3}) \quad (4.7)$$

Moreover, $D$ is strictly positive on $\operatorname{Int} I_{\alpha,\beta}$ and possible zeros on the boundary of $I_{\alpha,\beta}$ are simple.

The discrete part $\mu^d$ is a linear combination of dirac measures

$$\mu^d = \sum_{i=1}^{2j-1} \lambda_i \delta_{x_i},$$

where the $x_i$ are the real roots of $D$ outside of $I_{\alpha,\beta}$. The weights $\lambda_i$ are possibly zero and are given as

$$\lambda_i := \lim_{x \to x_i} \left| (x - x_i) \frac{f(x)}{D(x)} \right|, \tag{4.8}$$

where the function $f$ is defined by

$$f(x) := Q_j(x)P_j(x) - Q_{j+1}(x)P_{j-1}(x) + \frac{\beta_1 \cdots \beta_{j-1}}{2}\left((x - \alpha) - z(x)\sqrt{(x-\alpha)^2 - 4\beta}\right) \tag{4.9}$$

with

$$z(x) := \begin{cases} 1 & x > \alpha + 2\sqrt{\beta} \\ -1 & x < \alpha - 2\sqrt{\beta} \end{cases}, \qquad Q_i(x) := \int \frac{P_i(x) - P_i(t)}{x - t}\, d\mu(t).$$

The polynomials $Q_i$, $i \geq 0$ in the proposition are called *secondary polynomials* and satisfy the shifted recurrence relation $Q_0(x) = 0$, $Q_1(x) = 1$ and

$$Q_i(x) = (x - \alpha_i)Q_{i-1}(x) - \beta_{i-1}Q_{i-1}(x), \quad i \geq 2. \tag{4.10}$$

*Proof.* By a classical result due to Markov, the Stieltjes transform $S_\mu(z)$ of $\mu$,

$$S_\mu(z) := \int_E \frac{1}{z - t} d\mu(t), \quad z \in \mathbb{C}_+ := \{z \in \mathbb{C} : \operatorname{Re} z > 0\},$$

admits a continued fraction expansion in terms of the recurrence coefficients,

$$S_\mu(z) = \frac{1}{\left| z - \alpha_1 \right.} - \frac{\beta_1}{\left| z - \alpha_2 \right.} - \cdots - \frac{\beta_{j-1}}{\left| z - \alpha_j \right.} - \beta S_{\alpha,\beta}(z),$$

see e.g. (Dette et al., 2018, Lemma 4.1) for a simple proof. Here $S_{\alpha,\beta}(z)$ is the Stieltjes transform of the measure with constant recurrence coefficients $\alpha$, $\beta$. Writing the convergent

$$\frac{1}{\left| z - \alpha_1 \right.} - \frac{\beta_1}{\left| z - \alpha_2 \right.} - \cdots - \frac{\beta_{j-1}}{\left| z - \alpha_j \right.} = \frac{A_j}{B_j}$$

as a single fraction, it is a standard argument to see that $B_j$, $j \geq 0$, are polynomials of degree $j$ satisfying the same recursion as the monic orthogonal polynomials $P_j$, $j \geq 0$ w.r.t. $\mu$. Comparison of coefficients then gives $B_j = P_j$. Likewise one finds that $A_j$, $j \geq 0$ satisfies the recursion (4.10) and concludes $A_j = Q_j$. This gives

$$S_\mu(z) = \frac{Q_j(z) - \beta S_{\alpha,\beta}(z)Q_{j-1}(z)}{P_j(z) - \beta S_{\alpha,\beta}(z)P_{j-1}(z)}. \tag{4.11}$$

Expanding $S_{\alpha,\beta}(z)$ shows that it satisfies the recursion

$$S_{\alpha,\beta}(z) = \frac{1}{z - \alpha - \beta S_{\alpha,\beta}(z)},$$

which yields $S_{\alpha,\beta}(z) = \frac{z - \alpha - \sqrt{(z-\alpha)^2 - 4\beta}}{2\beta}$. In this representation, the square root $\sqrt{y}$ of a complex number $y \in \mathbb{C} \setminus \mathbb{R}_+$ is defined to be the unique solution of the equation $x^2 = y$ with a positive

20

imaginary part. With this definition, the square root appearing in $S_{\alpha,\beta}$ can be continuously extended to the real line via

$$\lim_{\substack{z \to x \\ z \in \mathbb{C}^+}} \sqrt{(z-\alpha)^2 - 4\beta} = \begin{cases} -\sqrt{(x-\alpha)^2 - 4\beta} & \text{,if } x < \alpha - 2\sqrt{\beta} \\ i\sqrt{4\beta - (x-\alpha)^2} & \text{,if } x \in [\alpha - 2\sqrt{\beta}, \alpha + 2\sqrt{\beta}] \\ \sqrt{(x-\alpha)^2 - 4\beta} & \text{,if } x > \alpha + 2\sqrt{\beta} \end{cases} . \quad (4.12)$$

The latter expression then identifies $S_{\alpha,\beta}$ via the Stieltjes inversion formula

$$\frac{\mu(dx)}{dx} = -\frac{1}{\pi} \lim_{y \to 0} \operatorname{Im} S_\mu(x+iy) \quad (4.13)$$

as the Stieltjes transform of the semicircle distribution $\mu_{\alpha,\beta}$ of (3.15) on $I_{\alpha,\beta} = [\alpha - 2\sqrt{\beta}, \alpha + 2\sqrt{\beta}]$.

Observing (4.11) and (4.12), we can see that $S_\mu$ can be extended to the interior of the interval $I_{\alpha,\beta}$ as

$$S_\mu(x) = \frac{Q_j(x) - \frac{1}{2}(x-\alpha)Q_{j-1}(x) + \frac{i}{2}\sqrt{4\beta - (x-\alpha)^2}Q_{j-1}(x)}{P_j(x) - \frac{1}{2}(x-\alpha)P_{j-1}(x) + \frac{i}{2}\sqrt{4\beta - (x-\alpha)^2}P_{j-1}(x)} \quad (4.14)$$

for any $x$ in which the denominator does not vanish. The only points in which the imaginary part of the denominator vanishes are the roots of the polynomial $P_{j-1}$. However, in these points the real part of the denominator is given by $P_j(x)$, which does not vanish as the roots of two consecutive orthogonal polynomials strictly interlace. Consequently, $S_\mu$ can be continuously extended to Int $I_{\alpha,\beta}$ and thus it admits a density on this interval by (4.13). To compute it, we start with the imaginary part of (4.14), expanding the fraction with the complex conjugate of the denominator and thus obtaining the new denominator

$$\begin{aligned} D(x) &:= (P_j(x) - \beta S_{\alpha,\beta}(x)P_{j-1}(x))\overline{(P_j(x) - \beta S_{\alpha,\beta}(x)P_{j-1}(x))} \\ &= \left(P_j(x) - \frac{1}{2}(x-\alpha)P_{j-1}(x)\right)^2 + \frac{1}{4}\left(4\beta - (x-\alpha)^2\right)P_{j-1}^2(x) \quad (4.15) \\ &= P_j^2(x) - (x-\alpha)P_j(x)P_{j-1}(x) + \beta P_{j-1}^2(x) \\ &= P_j^2(x) - P_{j+1}(x)P_{j-1}(x). \end{aligned}$$

Calculating the corresponding numerator we arrive at

$$\operatorname{Im} S_\mu(x) = -\frac{1}{2D(x)}\sqrt{4\beta - (x-\alpha)^2}\Big(P_{j-1}(x)Q_j(x) - Q_{j-1}(x)P_j(x)\Big)$$

for any $x \in (\alpha - 2\sqrt{\beta}, \alpha + 2\sqrt{\beta})$. We can further simplify using the Christoffel-Darboux formula (Theorem 4.5 in Chihara (1978))

$$\begin{aligned} P_{j-1}(x)Q_j(x) - Q_{j-1}(x)P_j(x) &= \int \frac{P_{j-1}(t)P_j(x) - P_j(t)P_{j-1}(x)}{x-t} \, d\mu(t) \\ &= h_{j-1} \int \sum_{i=0}^{j-1} \frac{P_i(x)P_i(t)}{h_i} \, d\mu(t) = h_{j-1} \quad (4.16) \end{aligned}$$

with $h_i := \int P_i^2(t) \, d\mu(t) = \beta_1 \cdots \beta_i$. Consequently, $\mu$ has a density on the interval Int $I_{\alpha,\beta}$ that is given by

$$\frac{\sqrt{4\beta - (x-\alpha)^2}\beta_1 \cdots \beta_{j-1}}{2\pi D(x)}.$$

Next, we turn to a more thorough investigation of $D(x)$. First of all, from the representation (4.15) we can conclude that the only possible roots of $D$ on $I_{\alpha,\beta}$ must lie on the boundary of the interval. If $x_0$ is such a root, then

$$P_j(x_0) - \frac{1}{2}(x_0 - \alpha)P_{j-1}(x_0) = 0$$

must hold. Since $P_j$ and $P_{j-1}$ have no common roots, this implies $P_{j-1}(x_0) \neq 0$. But then $x_0$ can only be a simple root of $D$ by (4.15). This also shows $\mu(\{\alpha \pm 2\sqrt{\beta}\}) = 0$ by the Stieltjes inversion formula for atoms

$$\mu(\{x\}) = -\lim_{y \to 0} y \operatorname{Im} S_\mu(x + iy). \tag{4.17}$$

Furthermore, an induction argument shows

$$P_j(x) = x^j - (\alpha_1 + \cdots + \alpha_j)x^{j-1} + \left(\sum_{1 \leq i < l \leq j} \alpha_i \alpha_l - \sum_{i=1}^{j-1} \beta_i\right) x^{j-2} + O(x^{j-3}),$$

$$P_j^2(x) = x^{2j} - 2(\alpha_1 + \cdots + \alpha_j)x^{2j-1}$$
$$+ \left(2 \sum_{1 \leq i < l \leq j} \alpha_i \alpha_l - 2 \sum_{i=1}^{j-1} \beta_i + (\alpha_1 + \cdots + \alpha_j)^2\right) x^{2j-2} + O(x^{2j-3}),$$

$$P_j(x)P_{j-1}(x) = x^{2j-1} - (2(\alpha_1 + \cdots + \alpha_{j-1}) + \alpha_j)x^{2j-2}$$
$$+ \left(\sum_{1 \leq i < l \leq j-1} \alpha_i \alpha_l - \sum_{i=1}^{j-2} \beta_i + \sum_{1 \leq i < l \leq j} \alpha_i \alpha_l - \sum_{i=1}^{j-1} \beta_i\right.$$
$$+ (\alpha_1 + \cdots + \alpha_j)(\alpha_1 + \cdots + \alpha_{j-1})\Bigg) x^{2j-3} + O(x^{2j-4}).$$

Plugging these formulas into (4.15) proves (4.7).

Observing (4.12), we can continuously extend the Stieltjes transform to $\mathbb{R} \setminus I_{\alpha,\beta}$ via

$$\frac{Q_j(x) - \frac{1}{2}(x - \alpha)Q_{j-1}(x) + \frac{1}{2}z(x)\sqrt{(x - \alpha)^2 - 4\beta}Q_{j-1}(x)}{P_j(x) - \frac{1}{2}(x - \alpha)P_{j-1}(x) + \frac{1}{2}z(x)\sqrt{(x - \alpha)^2 - 4\beta}P_{j-1}(x)} \tag{4.18}$$

in any point $x$ that is not a root of the denominator. As this expression is real, the absolutely continuous part vanishes outside of $I_{\alpha,\beta}$. Let us now examine whether $\mu$ admits any point masses in the roots of the denominator. If $x_0$ is such a root, then an expansion of the polynomial terms in (4.11) combined with the Stieltjes inversion formula (4.17) shows

$$\mu(\{x_0\}) = \lim_{x \to x_0} |(x - x_0)S_\mu(x)|,$$

where $S_\mu(x)$ is the extension of the Stieltjes transform to the real line (4.18). It only remains to show that the representation (4.8) of the discrete part of $\mu$ in the formulation of the theorem holds. To this end, note that by expanding the fraction in (4.18) with the term

$$P_j(x) - \frac{1}{2}(x - \alpha)P_{j-1}(x) - \frac{1}{2}z(x)\sqrt{(x - \alpha)^2 - 4\beta}P_{j-1}(x)$$

the denominator becomes $D(x)$. Using the recursion (3.3) of the orthogonal polynomials and the Christoffel-Darboux formula (4.16) shows

$$(Q_j(x) - \frac{1}{2}(x - \alpha)Q_{j-1}(x) + \frac{1}{2}z(x)\sqrt{(x - \alpha)^2 - 4\beta}Q_{j-1}(x))$$

$$\cdot (P_j(x) - \frac{1}{2}(x - \alpha)P_{j-1}(x) - \frac{1}{2}z(x)\sqrt{(x - \alpha)^2 - 4\beta}P_{j-1}(x))$$

$$= Q_j(x)P_j(x) - \frac{1}{2}(x - \alpha)(Q_j(x)P_{j-1}(x) + Q_{j-1}(x)P_j(x))$$

$$- z(x)\frac{1}{2}(Q_j(x)P_{j-1}(x) - Q_{j-1}(x)P_j(x))\sqrt{(x - \alpha)^2 - 4\beta} + \beta Q_{j-1}(x)P_{j-1}(x)$$

$$= Q_j(x)P_j(x) + \frac{1}{2}(x - \alpha)(Q_j(x)P_{j-1}(x) - Q_{j-1}(x)P_j(x)) - (x - \alpha)Q_j(x)P_{j-1}(x)$$

$$- z(x)\frac{h_{j-1}}{2}\sqrt{(x - \alpha)^2 - 4\beta} + \beta Q_{j-1}(x)P_{j-1}(x)$$

$$= Q_j(x)P_j(x) - Q_{j+1}(x)P_{j-1}(x) + \frac{h_{j-1}}{2}\left((x - \alpha) - z(x)\sqrt{(x - \alpha)^2 - 4\beta}\right) = f(x). \qquad \square$$

**Example 4.5** (Continuation of Example 3.8). It remains to compute the measure in (3.17) from the information that the sequence of random moments $(m_1^{(n)}, \ldots, m_l^{(n)})$ converges in probability to the sequence of moments $(m_1(\mu), \ldots, m_l(\mu))$, where the measure $\mu$ is uniquely determined by having recursion coefficients $\alpha_1 = 0$, $\alpha_j = 1$ and $\beta_{j-1} = \frac{1}{4}$ for all $j > 1$. In light of Theorem 4.4, we calculate the orthogonal polynomials $P_j$ as well as the secondary polynomials $Q_j$ up to order 2 as

$$P_0(x) = 1, \qquad\qquad\qquad\qquad\qquad Q_0(x) = 0,$$
$$P_1(x) = x - \alpha_1 = x, \qquad\qquad\qquad Q_1(x) = 1,$$
$$P_2(x) = (x - \alpha_2)P_1(x) - \beta_1 P_0(x) = x^2 - x - \frac{1}{4}, \qquad Q_2(x) = x - \alpha_2 = x - 1.$$

Consequently, the polynomial $D$ is given by

$$D(x) = P_1^2(x) - P_2(x)P_0(x) = x + \frac{1}{4}.$$

Therefore $\mu$ has a possible point mass in $-\frac{1}{4}$. In order to calculate the weight of the point mass, we determine

$$f(x) = 1 + \frac{1}{2}(x - 1 - z(x)\sqrt{x^2 - 2x}).$$

This yields the weight of the point mass as

$$\lim_{x \to -\frac{1}{4}} \left|\left(x + \frac{1}{4}\right)\frac{f(x)}{x + \frac{1}{4}}\right| = f\left(-\frac{1}{4}\right) = \frac{3}{4}.$$

Hence the measure $\mu$ is given by (3.17).

We are now in the position to prove Theorem 3.6.

*Proof of Theorem 3.6.* After a change of variables by means of Lemma 4.1, an application of Proposition 4.3 shows the convergence on the level of canonical coordinates. The convergence result can then be transferred back to the ordinary moments with the continuous mapping theorem.

It remains to show the claimed representation of $\mu_q$. The measure $\mu_q$ is uniquely determined by having $\mathbf{y}^{*,q}$ as first $i_k$ canonical coordinates, as well as

$$y_j = \begin{cases} y_1^* & , j \text{ odd} \\ y_2^* & , j \text{ even} \end{cases}$$

for all $j > i_k$, where $y_1^*$ and $y_2^*$ were defined in (3.16). We have to distinguish the three cases of $E$.

**Case $E = \mathbb{R}$:** In this case the respresentation is a direct consequence of Proposition 4.4.

**Case $E = \mathbb{R}_+$:** In the case $E = \mathbb{R}_+$, (Dette et al., 2018, p. 18) shows that the recursion coefficients of the orthogonal polynomials corresponding to $\mu_q$ are given in terms of the canonical coordinates as

$$\alpha_j = z_{2j-2} + z_{2j-1},$$
$$\beta_j = z_{2j-1}z_{2j},$$

and moreover, we have

$$z_{2j+1} = -\frac{P_{j+1}(0)}{P_j(0)}. \tag{4.19}$$

Since the odd and even parameters $z_j$ are constant for all $j > i_k$, the $\alpha_j$ and $\beta_{j-1}$ are constant for all $j > l := \lfloor \frac{i_k+3}{2} \rfloor$ and we may apply Proposition 4.4. Observing the expansion (4.7) of the polynomial $D(x)$ from Proposition 4.4 for $i_k$ even and odd respectively, we can see that the degree of $D$ is at most $i_k + 1$. To prove that the absolutely continuous part of the measure is of the claimed form

$$\mu_q^{ac}(x) = \frac{1}{D_q(x)}\mu_{z_1^*,z_2^*}^{ac}(x)$$

for some polynomial $D_q$ of degree at most $i_k$, we have to show $D(0) = 0$ in order to factor out the linear polynomial $p(x) = x$. By (4.19),

$$P_s(0) = (-1)^s \prod_{j=0}^{s-1} \left( -\frac{P_{j+1}(0)}{P_j(0)} \right) = (-1)^s \prod_{j=0}^{s-1} z_{2j+1}, \tag{4.20}$$

which yields

$$D(0) = P_l^2(0) - P_{l-1}(0)P_{l+1}(0) = \left( \prod_{j=0}^{l-2} z_{2j+1}^2 \right) z_{2l-1}(z_{2l-1} - z_{2l+1}) = 0,$$

since $z_{2l-1} = z_{2l+1}$. Factoring out the polynomial $p(x) = x$ yields the desired result for the absolutely continuous part with the polynomial $D_q(x) := D(x)/x$ which is of degree at most $i_k$ and is strictly positive at 0 since the root of $D$ in 0 is simple by Proposition 4.4.

For the discrete part of the measure, let $x_0$ be a root of $D(x)$, i.e. a number satisfying $P_l^2(x_0) = P_{l-1}(x_0)P_{l+1}(x_0)$. Since two consecutive orthogonal polynomials have no common roots, $x_0$ is neither a root of $P_{l-1}$, $P_l$ or $P_{l+1}$. We then have

$$\left( Q_l(x_0)P_l(x_0) - Q_{l+1}(x_0)P_{l-1}(x_0) \right) P_{l+1}(x_0)$$
$$= Q_l(x_0)P_l(x_0)P_{l+1}(x_0) - Q_{l+1}(x_0)P_l^2(x_0)$$
$$= -P_l(x_0)\left( Q_{l+1}(x_0)P_l(x_0) - Q_l(x_0)P_{l+1}(x) \right) = -P_l(x_0)h_l,$$

where we have used the identity (4.16) in the last line. Recalling the definition of the function $f$ in (4.9) we therefore obtain

$$f(x_0) = -\frac{P_l(x_0)h_l}{P_{l+1}(x_0)} + \frac{\beta_1 \cdots \beta_{l-1}}{2}\left( x - \alpha - z(x)\sqrt{(x-\alpha)^2 - 4\beta} \right). \tag{4.21}$$

Plugging in the value $x_0 = 0$ and using (4.20) as well as $\alpha = z_1^* + z_2^*$, $\beta = z_1^* z_2^*$ and $h_j = \beta_1 \cdots \beta_j$ we obtain

$$
\begin{aligned}
f(0) &= \frac{h_l}{z_{2l+1}} + \frac{h_{l-1}}{2}\left(-z_1^* - z_2^* + \sqrt{(z_1^* + z_2^*)^2 - 4z_1^* z_2^*}\right) \\
&= h_{l-1}\left(\frac{z_{2l-1} z_{2l}}{z_{2l+1}} + \frac{1}{2}\left(-z_1^* - z_2^* + |z_1^* - z_2^*|\right)\right) \\
&= h_{l-1}(z_2^* - z_1^*)_+ \,,
\end{aligned}
$$

where we have used that $z_{2l-1} = z_{2l+1} = z_1^*$ and $z_{2l} = z_2^*$ (note that $l$ was chosen such that $2l - 1 \geq i_k + 1$).

Recall that the weight of the point mass of $\mu_q$ in zero can be calculated via the formula

$$
\lambda = \lim_{x \to 0}\left| x \frac{f(x)}{D(x)} \right|.
$$

We now have to consider three separate cases:

(1) If $\mu_{z_1^*, z_2^*}$ has an atom at 0, then $z_2^* > z_1^*$ must hold. In this case, we have $f(0) > 0$ and consequently $\lambda > 0$. This means that $\mu_q$ has a point mass at zero, as well as up to $i_k$ further point masses at the roots of $D$ outside of zero.
(2) If $\mu_{z_1^*, z_2^*}$ has no atom at 0 and 0 is not a simple root of $D$, then $D$ has at most $i_k$ distinct roots and therefore up to $i_k$ possible point masses.
(3) If $\mu_{z_1^*, z_2^*}$ has no atom at 0 and 0 is a simple root of $D$, then $z_1^* \leq z_2^*$, $f(0) = 0$ and $D'(0) \neq 0$ holds. We therefore obtain

$$
\lambda = \lim_{x \to 0}\left| x \frac{f(x)}{D(x)} \right| = \left| \frac{f(0)}{D'(0)} \right| = 0 \,.
$$

Consequently, $\mu_q$ has no point mass in zero and up to $i_k$ further point masses in the roots of $D$ outside of zero.

Common to all three cases is that $\mu_q$ has at most $i_k$ point masses more than $\mu_{z_1^*, z_2^*}$, which yields the desired result.

**Case $E = [0, 1]$:** In the case $E = [0, 1]$, arguments as above show that the recursion coefficients $\alpha_j$ and $\beta_{j-1}$ are constant for all $j > l := \lfloor \frac{i_k+4}{2} \rfloor$. Similar calculations using (see Dette and Nagel (2012))

$$
\begin{aligned}
\alpha_j &= q_{2j-3} p_{2j-2} + q_{2j-2} p_{2j-1}, \\
\beta_j &= q_{2j-2} p_{2j-1} q_{2j-1} p_{2j},
\end{aligned}
\tag{4.22}
$$

where $q_j := 1 - p_j$ and $p_{-1} := p_0 := 0$, and formula (4.20) then show that $D$ is a polynomial of degree at most $i_k + 2$, satisfying $D(0) = 0$. Furthermore, it is well-known that the monic orthogonal polynomial $P_s$ of order $s$ is given by

$$
P_s(x) = (\Delta_{s-1})^{-1}
\begin{pmatrix}
m_0 & m_1 & \cdots & m_{s-1} & 1 \\
m_1 & m_2 & \cdots & m_s & x \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
m_s & m_{s+1} & \cdots & m_{2s-1} & x^s
\end{pmatrix}
$$

25

with $\Delta_{s-1} = \det\left((m_{a+b})_{a,b=0}^{s-1}\right)$ and $m_0 := 1$, see e.g. (Deift, 1999, p. 38). This yields

$$P_2(1) = (\Delta_{s-1})^{-1}\begin{pmatrix} m_0 & m_1 & \ldots & m_{s-1} & 1 \\ m_1 & m_2 & \cdots & m_s & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ m_s & m_{s+1} & \cdots & m_{2s-1} & 1 \end{pmatrix}$$

$$= (\Delta_{s-1})^{-1}\begin{pmatrix} m_0 - m_1 & m_1 - m_2 & \ldots & m_{s-1} - m_s & 0 \\ m_1 - m_2 & m_2 - m_3 & \cdots & m_s - m_{s+1} & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ m_{s-1} - m_2 & m_s - m_{s+1} & \cdots & m_{2s-2} - m_{2s-1} & 0 \\ m_s & m_{s+1} & \cdots & m_{2s-1} & 1 \end{pmatrix}$$

$$= (\Delta_{s-1})^{-1} \det\left((m_{a+b} - m_{a+b+1})_{a,b=0}^{s-1}\right).$$

Combined with (Dette and Studden, 1997, Theorem 1.4.10 and Corollary 1.4.6) this results in

$$P_s(1) = \prod_{j=0}^{s-1} q_{2j}q_{2j+1} \tag{4.23}$$

and consequently

$$D(1) = \left(\prod_{j=0}^{l-2} q_{2j}^2 q_{2j+1}^2\right) q_{2l-2}q_{2l-1}(q_{2l-2}q_{2l-1} - q_{2l}q_{2l+1}) = 0\,.$$

Factoring out $x(1-x)$ in $D$ proves the representation of the absolutely continuous part.

For the discrete part of $\mu_q$, we observe that by (4.21), (4.20) and (4.23) we have

$$f(0) = h_{l-1}(p_2^* - p_1^*)_+,$$

$$f(1) = h_{l-1}\left(-p_1^* p_2^* + \frac{1}{2}\left(1 - (p_1^* + p_2^* - 2p_1^* p_2^*) - |1 - (p_1^* + p_2^*)|\right)\right)$$

$$= h_{l-1}(p_1^* + p_2^* - 1)_+.$$

Arguments similar to the case $E = \mathbb{R}_+$ show that $\mu_q$ has point masses in the point masses of $\mu_{p_1^*,p_2^*}$, as well as up to $i_k$ additional point masses. $\qquad\square$

## 5. Proof of the LDP

In this section, we will prove Theorem 5.2. As before, we start with a general proposition. We note in passing that the following statement is stronger than a usual LDP in the sense that it establishes the rate function for any measurable set as an actual limit that is described as an essential infimum, instead of just lower and upper bounds as in a usual LDP. This is formulated in the following lemma.

**Lemma 5.1** (Large deviation principle). *Let $\mathbb{P}_n$ be a sequence of probability measures on $\mathbb{R}^m$ and $a_n \to \infty$ a sequence such that*

$$\lim_{n\to\infty} \frac{1}{a_n} \log \mathbb{P}_n(\Gamma) = -\operatorname*{ess\,inf}_{x\in\Gamma} S(x)$$

*holds for all measurable sets $\Gamma \subset \mathbb{R}^m$ and some function $S : \mathbb{R}^m \to [0,\infty]$. Then $\mathbb{P}_n$ satisfies an LDP with speed $a_n$ and rate function*

$$I(x) := \operatorname*{ess\,lim\,inf}_{y\to x} S(y) = \sup\{\operatorname*{ess\,inf}_{y\in U} S(y) \mid x \in U, U \text{ open}\}.$$

*Proof.* It suffices to prove three assertions:

(1) $I$ is lower semicontinuous

(2) $\inf_{x\in G} I(x) \geq \operatorname*{ess\,inf}_{x\in G} S(x)$ holds for all open sets $G$.

(3) $\inf_{x\in F} I(x) \leq \operatorname*{ess\,inf}_{x\in F} S(x)$ holds for all closed sets $F$.

Let $x_0 \in \mathcal{X}$ and $K \leq I(x_0)$ be arbitrary with $K < \infty$. Choose $U$ open so that $\operatorname*{ess\,inf}_{y\in U} S(y) \geq K - \varepsilon$ holds. This implies for all $x \in U$ the inequality

$$I(x) = \sup\{\operatorname*{ess\,inf}_{y\in V} S(y) \mid x \in V, V \text{ open}\} \geq \operatorname*{ess\,inf}_{y\in U} S(y) \geq K - \varepsilon.$$

Now the lower semicontinuity follows, since for $I(x_0)$ finite, we may choose $K = I(x_0)$ and for infinite $I(x_0)$ by taking the limit $K \to \infty$.

Next, let $G$ be an open set with $x \in G$. Then we have

$$I(x) = \sup\{\operatorname*{ess\,inf}_{y\in U} S(y) \mid x \in U, U \text{ open}\} \geq \operatorname*{ess\,inf}_{y\in G} S(y)$$

The assertion (2) now follows by taking the infimum over all $x \in G$ on the left-hand side.

Finally, let $F$ be closed, fix an arbitrary finite $K \leq \inf_{y\in F} I(y)$ and let $\varepsilon > 0$. By the definition of $I$ we can find for each $x \in F$ an open set $U_x \ni x$ such that

$$\operatorname*{ess\,inf}_{y\in U_x} S(y) \geq K - \varepsilon.$$

As $\{U_x\}_{x\in F}$ is an open covering of the closed set $F$ and $\mathbb{R}^m$ is $\sigma$-compact, we can find a countable subcovering $U_{x_1}, U_{x_2}, \ldots$ for $F$. This implies

$$\operatorname*{ess\,inf}_{x\in F} S(x) \geq \inf_{n\in\mathbb{N}} \operatorname*{ess\,inf}_{x\in U_{x_n}} S(x) \geq K - \varepsilon,$$

and thus we get $\operatorname*{ess\,inf}_{x\in F} S(x) \geq K$. The result now follows by choosing $K := \inf_{y\in F} I(y)$ if the infimum or else letting $K \to \infty$. $\square$

**Proposition 5.2** (Large deviations). *Let $\mathbb{P}_n$ be given by* (4.5) *satisfying* (4.6) *and assume that $W : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ and $R : \mathbb{R}^m \to [0, \infty])$ are measurable functions such that*

*(1) $W$ is essentially lower bounded,*

*(2) $R$ is almost surely strictly positive.*

*Then $\mathbb{P}_n$ satisfies for all measurable $\Gamma \subset \mathbb{R}^m$*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}_n(\Gamma) = -\operatorname*{ess\,inf}_{x\in\Gamma} S(x),$$

*where*

$$S(x) := W(x) - \operatorname*{ess\,inf}_{y\in\mathbb{R}^m} W(y),$$

*Proof.* We will first show that

$$\lim_{n\to\infty} \frac{1}{n} \log \int_\Gamma e^{-nW(x)} R(x) \, dx = -\operatorname*{ess\,inf}_{x\in\Gamma} W(x)$$

holds for all measurable sets $\Gamma \subset \mathbb{R}^m$. Set $\alpha := \operatorname*{ess\,inf}_{x\in\Gamma} W(x)$, and note that the case $\alpha = \infty$ is trivial. If $\alpha < \infty$ we have $\Gamma \subset \{W \geq \alpha\} \cup N$ for some Lebesgue nullset $N$. This implies with $n_0$ from (4.6)

$$\limsup_{n\to\infty} \frac{1}{n} \log \int_\Gamma e^{-nW(x)} R(x) \, dx \leq \limsup_{n\to\infty} \frac{1}{n} \log \int_{\{W\geq\alpha\}} e^{-n_0 W(x)} R(x) e^{-(n-n_0)\alpha} \, dx$$

$$\leq \limsup_{n\to\infty} \left( \frac{\log C}{n} - \alpha \frac{n - n_0}{n} \right) = -\alpha.$$

Let $\varepsilon > 0$ be arbitrary. By definition of $\alpha$ we have $\mathrm{vol}_m(\Gamma \cap \{W \leq \alpha + \varepsilon\}) > 0$. Since $R$ is almost surely strictly positive, this implies

$$0 < \int_{\Gamma \cap \{W \leq \alpha + \varepsilon\}} e^{-n_0 W(x)} R(x) \, dx < \infty,$$

from which we can conclude

$$\liminf_{n \to \infty} \frac{1}{n} \log \int_\Gamma e^{-nW(x)} R(x) \, dx$$
$$\geq \liminf_{n \to \infty} \frac{1}{n} \log \int_{\Gamma \cap \{W \leq \alpha + \varepsilon\}} e^{-n_0 W(x)} e^{-(n-n_0)(\alpha + \varepsilon)} R(x) \, dx = -(\alpha + \varepsilon).$$

Now let $\varepsilon \to 0$ and combine this equation with (5) to conclude that (5) holds. This yields the desired result, as

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}_n(\Gamma) = \lim_{n \to \infty} \left\{ \frac{1}{n} \log \int_\Gamma e^{-nW(x)} R(x) \, dx - \frac{1}{n} \log \int_{\mathbb{R}^m} e^{-nW(x)} R(x) \, dx \right\}$$
$$= -\operatorname*{ess\,inf}_{x \in \Gamma} W(x) + \operatorname*{ess\,inf}_{y \in \mathbb{R}^m} W(y) = -\operatorname*{ess\,inf}_{x \in \Gamma} S(x).$$

$\square$

*Proof of Theorem 3.2.* This theorem is a consequence of the previously obtained results and follows in three simple steps. Firstly, change coordinates from $(m_1^{(n)}, \ldots, m_l^{(n)})$ to the vector of unconstrained canonical coordinates $\mathbf{y}_l^{\mathcal{C},(n)}$ by means of Corollary 4.2. Secondly, apply Proposition 5.2 and Lemma 5.1 to obtain an LDP for the vector of canonical coordinates. The last step then is to transfer the LDP from canonical coordinates to ordinary coordinates using the contraction principle.

The compactness of level sets of $I$ is clear for $E = [0, 1]$ and follows for unbounded $E$ from the integrability conditions (3.7). $\square$

## 6. Proofs of the CLT and MDP

In this section we prove the central limit theorem and the moderate deviations principle of Section 3. As before, for later use we first give results for the general density $\mathbb{P}_n$ of the form (4.5), satisfying (4.6).

**Proposition 6.1** (Convergence to normal distribution)**.** *Let the same conditions and notations as in Proposition 4.3 hold and let $X_n$ be a random variable having distribution with density $\mathbb{P}_n$ from (4.5). Let $\theta_n^*$ be such that*

$$\theta_n^* \in \arg\min\{\|x - X_n\| \mid x \in \{\theta_1, \ldots, \theta_p\}\}$$

*with the standard Euclidean norm $\|\cdot\|$ and*

$$Y_n := \operatorname{Hess} W(\theta_n^*)^{1/2} \sqrt{n}(X_n - \theta_n^*).$$

*If $N \sim \mathcal{N}(0, 1)$, then as $n \to \infty$*

$$d_{TV}(Y_n, N) \to 0,$$

*where $d_{TV}$ denotes the total variation distance*

$$d_{TV}(X, Y) := \sup_{A \in \mathcal{B}(\mathbb{R}^m)} |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)|.$$

*Proof.* Choose $\varepsilon > 0$ so that the conditions

    (1) $B_\varepsilon(\theta_q) \cap B_\varepsilon(\theta_{q'}) = \emptyset$ for $q \neq q'$,
    (2) for all $x \in B_\varepsilon(\theta_q)$ we have $\operatorname{Hess} W(x) \geq M_q$ for some postive definite matrices $M_q$,

are satisfied and let $A \subset \mathbb{R}^m$ be a Borel set. Then

$$|\mathbb{P}\left(Y_n \in A\right) - \mathbb{P}\left(N \in A\right)| \leq |\mathbb{P}\left(Y_n \in A, \|X_n - \theta_n^*\| < \varepsilon\right) - \mathbb{P}\left(N \in A\right)| + \mathbb{P}\left(\|X_n - \theta_n^*\| \geq \varepsilon\right).$$

By Proposition 4.3 and Portmanteau's theorem, the last summand converges to zero. Set

$$D := \sum_{q=1}^{p} R(\theta_q) \det \operatorname{Hess} W(\theta_q)^{-1/2} \qquad \text{and} \qquad d_n := D^{-1} \left(\frac{n}{2\pi}\right)^{m/2} \int e^{-nW(x)} R(x)\, dx.$$

Then $d_n \to 1$ holds by calculations analogous to the ones used to prove Proposition 4.3. To simplify notation, define

$$f_n^q(y) := \exp(-nW(\theta_q + \operatorname{Hess} W(\theta_q)^{-1/2} n^{-1/2} y)) R(\theta_q + \operatorname{Hess} W(\theta_q)^{-1/2} n^{-1/2} y),$$

$$S_n^q := \operatorname{Hess} W(\theta_q)^{1/2} n^{1/2} B_\varepsilon(0),$$

$$K_n^q := \operatorname{Hess} W(\theta_q)^{-1/2} n^{-1/2} A$$

and introduce the decomposition

$$\exp(-y^t y/2) = D^{-1} \sum_{q=1}^{p} R(\theta_q) \det \operatorname{Hess} W(\theta_q)^{-1/2} \exp(-y^t y/2) 1_{S_n^q}(y)$$

$$+ D^{-1} \sum_{q=1}^{p} R(\theta_q) \det \operatorname{Hess} W(\theta_q)^{-1/2} \exp(-y^t y/2) 1_{(S_n^q)^c}(y). \tag{6.1}$$

Finally, note that

$$\{Y_n \in A, \|X_n - \theta_n^*\| < \varepsilon\} = \bigcup_{q=1}^{p} \{X_n - \theta_q \in K_n^q \cap B_\varepsilon(0)\}.$$

Combining all this yields

$$|\mathbb{P}\left(Y_n \in A, \|X_n - \theta_n^*\| < \varepsilon\right) - \mathbb{P}\left(N \in A\right)|$$

$$= \left| \sum_{q=1}^{p} \mathbb{P}\left(X_n - \theta_q \in K_n^q \cap B_\varepsilon(0)\right) - \mathbb{P}\left(N \in A\right) \right|$$

$$= \left| \frac{n^{m/2}}{(2\pi)^{m/2} d_n D} \sum_{q=1}^{p} \int_{K_n^q \cap B_\varepsilon(0)} \exp(-nW(\theta_q + x)) R(\theta_q + x)\, dx - \frac{1}{(2\pi)^{m/2}} \int_A \exp(-x^t x/2)\, dx \right|$$

$$= (2\pi)^{-m/2} \left| \int_A \sum_{q=1}^{p} d_n^{-1} D^{-1} f_n^q(y) 1_{S_n^q}(y) \det \operatorname{Hess} W(\theta_q)^{-1/2} - \exp(-y^t y/2)\, dy \right|.$$

By the decomposition (6.1) we therefore obtain

$$|\mathbb{P}\left(Y_n \in A, \|X_n - \theta_n^*\| < \varepsilon\right) - \mathbb{P}\left(N \in A\right)|$$

$$\leq (2\pi)^{-m/2} \sum_{q=1}^{p} D^{-1} \det \operatorname{Hess} W(\theta_q)^{-1/2} \left| \int_A \left(d_n^{-1} f_n^q(y) - \exp(-y^t y/2) R(\theta_q)\right) 1_{S_n^q}(y)\, dy \right|$$

$$+ D^{-1} \sum_{q=1}^{p} R(\theta_q) \det \operatorname{Hess} W(\theta_q)^{-1/2} P(N \notin S_n^q)$$

$$\leq (2\pi)^{-m/2} \sum_{q=1}^{p} D^{-1} \det \operatorname{Hess} W(\theta_q)^{-1/2} \int_{S_n^q} \left| d_n^{-1} f_n^q(y) - \exp(-y^t y/2) R(\theta_q) \right| dy$$

$$+ D^{-1} \sum_{q=1}^{p} R(\theta_q) \det \operatorname{Hess} W(\theta_q)^{-1/2} \mathbb{P}\left(N \notin S_n^q\right).$$

Note that this bound is uniform in $A$. Clearly $\mathbb{P}(N \notin S_n^q) \to 0$ as $n \to \infty$. On $S_n^q$ the pointwise convergence

$$d_n^{-1} f_n^q(y) \to \exp(-y^t y / 2) R(\theta_q)$$

holds and the claimed result follows from the dominated convergence theorem, using the dominating functions

$$R(\theta_q) \left( \left( \sup_n d_n^{-1} \right) \exp(-y^t M_q y / 2) + \exp(-y^t y / 2) \right), \qquad q = 1, \ldots, p.$$

$\square$

*Proof of Theorem 3.9.* After a change of coordinates by Corollary 4.2, an application of Proposition 6.1 yields a central limit theorem for the canonical coordinates,

$$\sqrt{n} (\mathrm{Hess}^{\mathcal{C}} \sum_{j=1}^{l} W_j)^{1/2} (\mathbf{y}_l^*)(\mathbf{y}_l^{\mathcal{C},(n)} - \mathbf{y}_l^*) \xrightarrow{d} \mathcal{N}(0, I_{l-k}),$$

where the first $i_k - k$ coordinates of $\mathbf{y}_l^*$ are given by

$$\underset{y \in \{\mathbf{y}^{*,1}, \ldots, \mathbf{y}^{*,p}\}}{\arg\min} \|\mathbf{y}_l^{\mathcal{C},(n)} - y\|$$

and the remaining coordinates for $j > i_k$ are

$$(\mathbf{y}_l^*)_j = \begin{cases} y_1^* & , j \text{ odd} \\ y_2^* & , j \text{ even} \end{cases}.$$

To transfer this CLT for $\mathbf{y}_l^{\mathcal{C},(n)}$ to one for $\mathbf{m}_l^{\mathcal{C},(n)}$, we use a variant of the delta-method, which is a common technique in mathematical statistics:

Let $(X_n)_n$, $(Y_n)_n$ be two sequences of random variables with $X_n, Y_n \in \mathbb{R}^m$ and $\sqrt{n}(X_n - Y_n) \to N \sim \mathcal{N}(0, \Sigma)$ in distribution as $n \to \infty$. Let $f : \mathbb{R}^m \to \mathbb{R}^m$ be a continuously differentiable function. Assume moreover that $Y_n$ is almost surely contained in an $n$-independent set $B$ on which $y \mapsto Df(y)$ is uniformly continuous and $\|Df(y)\|_{op} \geq c > 0$ for some $c$ and $y \in B$, $\|\cdot\|_{op}$ denoting the operator norm. Then $\sqrt{n} Df(Y_n)^{-1}(f(X_n) - f(Y_n))$ is asymptotically $\mathcal{N}(0, \Sigma)$-distributed. To see this, let $f_j$ denote the $j$-th component of the vector-valued $f$, $j = 1, \ldots, m$ and note that by the mean value theorem,

$$\sqrt{n}(f_j(X_n) - f_j(Y_n)) = \langle \nabla f_j(\xi_{j,n}), \sqrt{n}(X_n - Y_n) \rangle$$

for some $\xi_{j,n}$ in the line segment $[X_n, Y_n]$. Since $\sqrt{n}(X_n - Y_n)$ is asymptotically normal, we have $\xi_{j,n} - Y_n \to 0$ in probability. Let $M_n$ be the matrix built from the rows $\nabla f_j(\xi_{j,n})$, $j = 1, \ldots, m$. By the uniform continuity, $Df(M_n) - Df(Y_n) \to 0$ in probability and the asymptotic normality of $\sqrt{n} Df(Y_n)^{-1}(f(X_n) - f(Y_n))$ follows from $\|Df^{-1}\|_{op} \leq c^{-1}$ on $B$ and Slutsky's theorem.

In our case, $X_n := \mathbf{y}_l^{\mathcal{C},(n)}$, $Y_n := \mathbf{y}_l^*$ and $f := \varphi_l^{E,\mathcal{C}}$. The assumptions on $Df$ are easily verified for $B$ being the set of the $p$ possible values of $\mathbf{y}_l^*$. There is a small subtlety to consider: As the minimizers $\mathbf{y}_l^*$ and $\mathbf{m}_l^*$ are determined via $\mathbf{y}_l^{\mathcal{C},(n)}$ and $\mathbf{m}_l^{(n)}$ on different spaces, not necessarily $\varphi_l^{E,\mathcal{C}}(\mathbf{y}_l^*) = \mathbf{m}_l^*$ holds. However, by the continuity of $\varphi_l^{E,\mathcal{C}}$ the minimizers are identical whenever $\mathbf{m}_l^{(n)}$ is in a sufficiently small neighborhood $U$ of $\{\mathbf{m}_l^{*,1}, \ldots, \mathbf{m}_l^{*,p}\}$. Due to Theorem 3.6 we have $\mathbb{P}(\mathbf{m}_l^{(n)} \in U) \to 1$, which implies the stated result.

$\square$

Let us now turn to moderate deviations. The general result is

**Proposition 6.2** (Moderate Deviations)**.** *With the assumptions and notation of Proposition 6.1 for any sequence $a_n \to \infty$ with $a_n = o(\sqrt{n})$, the sequence of random variables*

$$Z_n := \mathrm{Hess}\, W(\theta_n^*)^{1/2} a_n (X_n - \theta_n^*)$$

*satisfies with* $b_n := \frac{n}{a_n^2}$

$$\lim_{n\to\infty} \frac{1}{b_n} \log \mathbb{P}(Z_n \in \Gamma) = - \operatorname*{ess\,inf}_{x\in\Gamma} \frac{1}{2} \|x\|_2^2$$

*for any measurable set* $\Gamma$.

*Proof.* Let $S_\varepsilon^q$ denote the ellipsoid

$$S_\varepsilon^q := \{x \in \mathbb{R}^m \mid \|\operatorname{Hess} W(\theta_q)^{1/2}(x - \theta_q)\| < \varepsilon\} = \theta_q + \operatorname{Hess} W(\theta_q)^{-1/2} B_\varepsilon(0). \qquad (6.2)$$

As in the proof of Proposition 6.1, let $0 < \varepsilon < 1$ be so small that the following conditions hold:

(1) $S_\varepsilon^q \cap S_\varepsilon^{q'} = \emptyset$ for $q \neq q'$,

(2) $\theta_q = \arg\min_{y \in S_\varepsilon^q} W(y)$ holds for all $1 \leq q \leq p$,

(3) for all $x \in S_\varepsilon^q$ we have $\operatorname{Hess} W(x) \geq M_q$ for some positive definite matrices $M_q$,

(4) there is a constant $K > 0$ such that $R(x) \leq K$ holds for all $x \in S_\varepsilon := \bigcup_{q=1}^p S_\varepsilon^q$,

(5) $\inf_{y \in S_\varepsilon^q} R(y) > 0$ holds for some $1 \leq q \leq p$.

For an arbitrary Borel set $\Gamma$ we use the decomposition

$$\mathbb{P}(Z_n \in \Gamma) = \sum_{q=1}^p \mathbb{P}\left(\operatorname{Hess} W(\theta_q)^{1/2} a_n(X_n - \theta_q) \in \Gamma, X_n \in S_\varepsilon^q\right)$$
$$+ \mathbb{P}(Z_n \in \Gamma, X_n \notin S_\varepsilon). \qquad (6.3)$$

Observing that

$$\max_{j=1}^N \frac{1}{b_n} \log A_j \leq \frac{1}{b_n} \log\left(\sum_{j=1}^N A_j\right) \leq \max_{j=1}^N \frac{1}{b_n} \log A_j + \frac{\log N}{b_n}, \qquad (6.4)$$

holds for any $A_1, \ldots, A_N \geq 0$, we see that only the largest of the probabilities in (6.3) matter. As will become apparent in the course of the proof, the normalization constant of the density $\exp(-nW(x))R(x)$ satisfies

$$\lim_{n\to\infty} \frac{1}{b_n} \log \int a_n^m e^{-nW(x)} R(x)\, dx = 0, \qquad (6.5)$$

which means that on the exponential scale of a large deviations principle, the integration constant roughly equals $a_n^{-m}$. In order to deal with $\mathbb{P}\left(\operatorname{Hess} W(\theta_q)^{1/2} a_n(X_n - \theta_q) \in \Gamma, X_n \in S_\varepsilon^q\right)$, we are therefore interested in approximations for the term

$$a_n^m \int_{S_\varepsilon^q} \mathbb{1}_\Gamma(\operatorname{Hess} W(\theta_q) a_n(x - \theta_q)) \exp(-nW(x)) R(x)\, dx. \qquad (6.6)$$

Set $\alpha := \operatorname*{ess\,inf}_{x\in\Gamma} \|x\|$, then we eventually want to prove $\lim_{n\to\infty} \frac{1}{b_n} \log \mathbb{P}(Z_n \in \Gamma) = -\alpha^2/2$. The case $\alpha = \infty$ corresponds to a nullset and the claimed result obviously holds. In the case $\alpha < \infty$ we start with an upper bound of (6.6). Define

$$\lambda_q^- := \lambda_q^-(\varepsilon) := \inf_{y \in S_\varepsilon^q} \lambda_{\min}\left(\operatorname{Hess} W(\theta_q)^{-1/2} \operatorname{Hess} W(y) \operatorname{Hess} W(\theta_q)^{-1/2}\right)$$

$$\lambda_q^+ := \lambda_q^+(\varepsilon) := \sup_{y \in S_\varepsilon^q} \lambda_{\max}\left(\operatorname{Hess} W(\theta_q)^{-1/2} \operatorname{Hess} W(y) \operatorname{Hess} W(\theta_q)^{-1/2}\right),$$

where $\lambda_{\min}(M)$ and $\lambda_{\max}(M)$ denote the smallest and largest eigenvalue of a symmetric square matrix $M$, respectively. By continuity $\lambda_q^-$ and $\lambda_q^+$ both converge to 1 as $\varepsilon \to 0$ and by Taylor's theorem we have for all $x \in S_\varepsilon^q$

$$W(x) =$$

$$\frac{1}{2}(\operatorname{Hess} W(\theta_q)^{1/2} x)^t \left(\operatorname{Hess} W(\theta_q)^{-1/2} \operatorname{Hess} W(\theta_q + \eta(x - \theta_q)) \operatorname{Hess} W(\theta_q)^{-1/2}\right)(\operatorname{Hess} W(\theta_q)^{1/2} x)$$

for some $\eta = \eta(x) \in (0,1)$, where we used our assumption $W(\theta_q) = 0$. Since $x \in S_\varepsilon^q$ we have $x = \theta_q + \operatorname{Hess} W(\theta_q)^{-1/2} v$ for some vector $v$ with $\|v\| \leq \varepsilon$. Therefore

$$\theta_q + \eta(x - \theta_q) = \theta_q + \operatorname{Hess} W(\theta_q)^{-1/2}(\eta v) \in S_\varepsilon^q$$

and we can conclude

$$\lambda_q^-(\varepsilon)\frac{\|\operatorname{Hess} W(\theta_q)^{1/2}(x - \theta_q)\|^2}{2} \leq W(x) \leq \lambda_q^+(\varepsilon)\frac{\|\operatorname{Hess} W(\theta_q)^{1/2}(x - \theta_q)\|^2}{2}.$$

This yields with $K$ from (4)

$$a_n^m \int_{S_\varepsilon^q} 1_\Gamma(\operatorname{Hess} W(\theta_q)^{1/2} a_n(x - \theta_q)) \exp(-nW(x)) R(x)\, dx$$

$$\leq a_n^m \int 1_\Gamma(\operatorname{Hess} W(\theta_q)^{1/2} a_n(x - \theta_q)) \exp\left(-n\lambda_q^- \|\operatorname{Hess} W(\theta_q)^{1/2}(x - \theta_q)\|^2/2\right) K\, dx$$

$$= \frac{a_n^m K}{n^{m/2}\sqrt{\det \operatorname{Hess} W(\theta_q)}} \int 1_\Gamma(a_n t/\sqrt{n}) \exp\left(-\lambda_q^- \|t\|^2/2\right) dt$$

$$\leq \frac{b_n^{-m/2} K}{\sqrt{\det \operatorname{Hess} W(\theta_q)}} \int 1_{[n\alpha^2/(2a_n^2),\infty)}(\|t\|^2/2) \exp\left(-(1 - \varepsilon)\lambda_q^- n\alpha^2/(2a_n^2) - \varepsilon\lambda_q^- \|t\|^2/2\right) dt$$

$$\leq \frac{b_n^{-m/2} K}{\sqrt{\det \operatorname{Hess} W(\theta_q)}} \exp\left(-(1 - \varepsilon)\lambda_q^- b_n\alpha^2/2\right) \left(\frac{2\pi}{\varepsilon\lambda_q^-}\right)^{m/2}.$$

Consequently, it follows

$$\limsup_{n\to\infty} \frac{1}{b_n} \log\left(a_n^m \int_{S_\varepsilon^q} 1_\Gamma(\operatorname{Hess} W(\theta_q)^{1/2} a_n(x - \theta_q)) e^{-nW(x)} R(x)\, dx\right)$$

$$\leq -(1 - \varepsilon)\lambda_q^- \frac{\alpha^2}{2}. \tag{6.7}$$

To find a lower bound of (6.6), let $n$ be so large that $\frac{\alpha + \varepsilon}{a_n} < \varepsilon$ holds. Then

$$a_n^m \int_{S_\varepsilon^q} 1_\Gamma(\operatorname{Hess} W(\theta_q)^{1/2} a_n(x - \theta_q)) e^{-nW(x)} R(x)\, dx$$

$$\geq a_n^m \inf_{y \in S_\varepsilon^q} R(y) \int 1_\Gamma(\operatorname{Hess} W(\theta_q)^{1/2} a_n x) 1_{[0,\varepsilon)}(\|\operatorname{Hess} W(\theta_q)^{1/2} x\|) e^{-n\lambda_q^+ \|\operatorname{Hess} W(\theta_q)^{1/2} x\|^2/2}\, dx$$

$$\geq \inf_{y \in S_\varepsilon^q} R(y) \frac{b_n^{-m/2}}{\sqrt{\det \operatorname{Hess} W(\theta_q)}} \int 1_\Gamma(a_n t/\sqrt{n}) 1_{[0,\alpha+\varepsilon)}(\|a_n t/\sqrt{n}\|) \exp\left(-\lambda_q^+ \|t\|^2/2\right) dt$$

$$\geq \inf_{y \in S_\varepsilon^q} R(y) \frac{b_n^{-m/2}}{\sqrt{\det \operatorname{Hess} W(\theta_q)}} \exp\left(-\lambda_q^+ n(\alpha + \varepsilon)^2/(2a_n^2)\right) \cdot \operatorname{vol}_m\left(\sqrt{b_n} \cdot (\Gamma \cap B_{\alpha+\varepsilon}(0))\right)$$

$$= \inf_{y \in S_\varepsilon^q} R(y) \left(\det \operatorname{Hess} W(\theta_q)\right)^{-1/2} \exp\left(-\lambda_q^+ b_n(\alpha + \varepsilon)^2/2\right) \cdot \operatorname{vol}_m\left(\Gamma \cap B_{\alpha+\varepsilon}(0)\right).$$

By definition of $\alpha$, $\Gamma \cap B_{\alpha+\varepsilon}(0)$ has positive volume. Therefore we get

$$\liminf_{n\to\infty} \frac{1}{b_n} \log\left(a_n^m \int_{S_\varepsilon^q} 1_\Gamma(\operatorname{Hess} W(\theta_q)^{1/2} a_n(x - \theta_q)) e^{-nW(x)} R(x)\, dx\right) \geq -\lambda_q^+ \frac{(\alpha + \varepsilon)^2}{2}, \tag{6.8}$$

if $\inf_{y \in S_\varepsilon^q} R(y) > 0$ and $-\infty$ else.

For the term $\mathbb{P}\left(Z_n \in \Gamma, X_n \notin S_\varepsilon\right)$ in (6.3) we obtain

$$\limsup_{n\to\infty} \frac{1}{b_n} \log\left(a_n^m \int_{(S_\varepsilon)^c} e^{-nW(x)} R(x)\, dx\right)$$

$$\leq \limsup_{n\to\infty} \frac{1}{b_n} \log\left(a_n^m \int \exp(-n_0 W(x)) R(x)\, dx \exp\left(-(n - n_0) \inf_{y \notin S_\varepsilon} W(y)\right)\right),$$

where $n_0$ is the fixed number from (4.6). Thus

$$\limsup_{n\to\infty} \frac{1}{b_n} \log\left(a_n^m \int_{(S_\varepsilon)^c} e^{-nW(x)} R(x)\, dx\right)$$

$$\leq \limsup_{n\to\infty} \frac{a_n^2}{n}\left(\log(a_n^m) - (n-n_0)\inf_{y\notin S_\varepsilon} W(y)\right)$$

$$\leq \limsup_{n\to\infty} \frac{a_n^2\left(m\log(a_n) - n/2 \inf_{y\notin S_\varepsilon} W(y)\right)}{n} + \limsup_{n\to\infty} -a_n^2(1-n_0/n)\inf_{y\notin S_\varepsilon} W(y) = -\infty. \ (6.9)$$

In order to see the last equality, note that the first term is negative for sufficiently large $n$ and the second term diverges to $-\infty$. In view of (6.4) we conclude that $\mathbb{P}\left(Z_n \in \Gamma, X_n \notin S_\varepsilon\right)$ is negligible.

It remains to show (6.5). Using (6.4) in combination with (6.7) and (6.9) for $\Gamma = \mathbb{R}^m$ yields

$$\limsup_{n\to\infty} \frac{1}{b_n} \log \int a_n^m e^{-nW(x)} R(x)\, dx$$

$$= \max\left\{\left(\max_{q=1}^{p} \limsup_{n\to\infty} \frac{1}{b_n} \log \int_{S_\varepsilon^q} a_n^m e^{-nW(x)} R(x)\, dx\right),\right.$$

$$\left. \limsup_{n\to\infty} \frac{1}{b_n} \log \int_{(S_\varepsilon)^c} a_n^m e^{-nW(x)} R(x)\, dx\right\} \leq 0.$$

On the other hand, we obtain from (6.8) and (6.4) that

$$\liminf_{n\to\infty} \frac{1}{b_n} \log \int a_n^m e^{-nW(x)} R(x)\, dx \geq \max_{q=1}^{p} \liminf_{n\to\infty} \frac{1}{b_n} \log \int_{S_\varepsilon^q} e^{-nV(x)} R(x)\, dx$$

$$\geq \max_{q=1}^{p}\left\{-\lambda_q^+ \frac{\varepsilon^2}{2}\right\}.$$

Letting $\varepsilon \to 0$, (6.5) follows. Combining (6.7), (6.4) and (6.5) gives

$$\limsup_{n\to\infty} \frac{1}{b_n} \log \mathbb{P}\left(Z_n \in \Gamma\right) \leq \limsup_{n\to\infty} \frac{1}{b_n} \log\left(\mathbb{P}\left(X_n \notin S_\varepsilon\right) + \sum_{q=1}^{p} \mathbb{P}\left(Z_n \in \Gamma, X_n \in S_\varepsilon^q\right)\right)$$

$$\leq \limsup_{n\to\infty} \max\left\{\left(\max_{q=1}^{p} \frac{1}{b_n} \log \int_{S_\varepsilon^q} a_n^m 1_\Gamma(\operatorname{Hess} W(\theta_q)^{1/2} a_n(x-\theta_q)) e^{-nW(x)} R(x)\, dx\right),\right.$$

$$\left. \frac{1}{b_n} \log \int_{(S_\varepsilon)^c} a_n^m e^{-nW(x)} R(x)\, dx\right\}$$

$$+ \limsup_{n\to\infty} \frac{\log(p+1)}{b_n} + \limsup_{n\to\infty}\left\{-\frac{1}{b_n} \log \int a_n^m e^{-nW(x)} R(x)\, dx\right\}$$

$$\leq \max_{q=1}^{p}\left\{-(1-\varepsilon)\lambda_q^- \frac{\alpha^2}{2}\right\}.$$

Letting $\varepsilon \to 0$ we now get

$$\limsup_{n\to\infty} \frac{1}{b_n} \log \mathbb{P}\left(Z_n \in \Gamma\right) \leq -\frac{\alpha^2}{2}. \tag{6.10}$$

Similarly, (6.8), (6.4) and (6.5) yield

$$\liminf_{n\to\infty}\frac{1}{b_n}\log\mathbb{P}\left(Z_n\in\Gamma\right)\geq\liminf_{n\to\infty}\frac{1}{b_n}\log\left(\sum_{j=1}^{n}\mathbb{P}\left(Z_n\in\Gamma,X_n\in S_\varepsilon^q\right)\right)$$

$$\geq\max_{q=1}^{p}\liminf_{n\to\infty}\frac{1}{b_n}\log\int_{S_\varepsilon^q}a_n^m\mathbf{1}_\Gamma(\operatorname{Hess}W(\theta_q)^{1/2}a_n(x-\theta_q))e^{-nW(x)}R(x)\,dx$$

$$\geq\max_{q=1}^{p}\left\{-\lambda_q^+\frac{(\alpha+\varepsilon)^2}{2}\right\}.$$

Letting $\varepsilon\to 0$ we thus obtain

$$\liminf_{n\to\infty}\frac{1}{b_n}\log\mathbb{P}\left(Z_n\in\Gamma\right)\geq-\frac{\alpha^2}{2}. \tag{6.11}$$

Combining (6.10) and (6.11) now finally shows

$$\lim_{n\to\infty}\frac{1}{b_n}\log\mathbb{P}\left(Z_n\in\Gamma\right)=-\frac{\alpha^2}{2}=-\frac{1}{2}\left(\operatorname*{ess\,inf}_{x\in\Gamma}\|x\|\right)^2=-\operatorname*{ess\,inf}_{x\in\Gamma}\frac{\|x\|^2}{2}.$$

$\square$

*Proof of Theorem 3.10.* After a change of coordinates by Corollary 4.2, an application of Proposition 6.2 and Lemma 5.1 yields the MDP for

$$Y_n:=a_nH(\mathbf{y}_l^*)\cdot(\mathbf{y}_l^{\mathcal{C},(n)}-\mathbf{y}_l^*)$$

with good rate function $I(x)=-\frac{1}{2}\|x\|^2$, speed $b_n=\frac{n}{a_n^2}$ and

$$H(\mathbf{y}_l):=\left(\operatorname{Hess}^{\mathcal{C}}\sum_{j=1}^{l}W_j\right)^{1/2}(\mathbf{y}_l).$$

To transfer this to an MDP for the ordinary moments, we first argue that we have the MDP with same rate function and same speed for $Y_n$ under the conditioned probability measures

$$\mathbb{P}^q:=\mathbb{P}(\cdot|\mathbf{y}_l^*=\mathbf{y}_{l,q}^*),\quad q=1,\ldots,p,$$

where $\mathbf{y}_{l,1}^*,\ldots,\mathbf{y}_{l,p}^*$ are the $p$ possible values of $\mathbf{y}_l^*$ and $\mathbb{P}$ denotes the underlying probability measure. To see this, note that under $\mathbb{P}^q$ we have

$$Y_n=Y_n^q:=a_nH(\mathbf{y}_{l,q}^*)(\mathbf{y}_l^{\mathcal{C},(n)}-\mathbf{y}_{l,q}^*)=a_n\left(H(\mathbf{y}_{l,q}^*)\mathbf{y}_l^{\mathcal{C},(n)}-H(\mathbf{y}_{l,q}^*)\mathbf{y}_{l,q}^*\right).$$

By Theorem 3.6 there are $0<c_1<c_2$ such that for $n$ large enough and all $q=1,\ldots,p$

$$c_1\leq\mathbb{P}(\mathbf{y}_l^*=\mathbf{y}_{l,q}^*)\leq c_2.$$

The upper bound in (2.7) for $Y_n^q$ follows for $n$ large from

$$\mathbb{P}^q(Y_n^q\in\Gamma)\leq c_1^{-1}\mathbb{P}(Y_n\in\Gamma).$$

For the lower bound, for $n$ large enough

$$\mathbb{P}^q(Y_n^q\in\Gamma)\geq c_2^{-1}\,\mathbb{P}(Y_n\in\Gamma,Y_n\in S_\varepsilon^q)$$

with $S_\varepsilon^q$ from (6.2) and $\varepsilon$ chosen as in the proof of Proposition 6.2. It was shown in (6.3), (6.5) and (6.8) that

$$\liminf_{n\to\infty}\frac{1}{b_n}\log\mathbb{P}(Y_n\in\Gamma,Y_n\in S_\varepsilon^q)=-\frac{(\operatorname{ess\,inf}_{x\in\Gamma}\|x\|^2+\varepsilon)^2}{2}$$

and the lower bound follows from letting $\varepsilon\to 0$ and Lemma 5.1.

Our aim is to transfer the MDPs for $Y_n^q$, $q = 1, \ldots, p$ to one for $Z_n$ from Theorem 3.10. Choosing

$$f^q(x) := H(\mathbf{y}_{l,q}^*) \Big[ \frac{\partial \mathbf{y}_l^{\mathcal{C}}}{\partial \mathbf{m}_l^{\mathcal{C}}}(\mathbf{y}_{l,q}^*) \Big]^t \varphi_l^{E,\mathcal{C}} \Big( (H(\mathbf{y}_{l,q}^*))^{-1} \cdot x \Big)$$

we can apply the delta-method for large deviations (Gao and Zhao, 2011, Theorem 3.1) to obtain an LDP for the sequence

$$Z_n^q := a_n \Big( f^q \big( H(\mathbf{y}_{l,q}^*) \mathbf{y}_l^{\mathcal{C},(n)} \big) - f^q \big( H(\mathbf{y}_{l,q}^*) \mathbf{y}_{l,q}^* \big) \Big)$$

$$= a_n H(\mathbf{y}_{l,q}^*) \Big[ \frac{\partial \mathbf{y}_l^{\mathcal{C}}}{\partial \mathbf{m}_l^{\mathcal{C}}}(\mathbf{y}_{l,q}^*) \Big]^t \Big( \mathbf{m}_l^{\mathcal{C},(n)} - \varphi_l^{E,\mathcal{C}}(\mathbf{y}_{l,q}^*) \Big)$$

under $\mathbb{P}^q$, with good rate function $I$ and speed $b_n$. The next step is to extend the MDP to

$$Z_n' := a_n H(\mathbf{y}_l^*) \Big[ \frac{\partial \mathbf{y}_l^{\mathcal{C}}}{\partial \mathbf{m}_l^{\mathcal{C}}}(\mathbf{y}_l^*) \Big]^t \Big( \mathbf{m}_l^{\mathcal{C},(n)} - \varphi_l^{E,\mathcal{C}}(\mathbf{y}_l^*) \Big)$$

under $\mathbb{P}$. In order to show this LDP, we estimate for an arbitrary measurable set $\Gamma$

$$c_1 \sum_{q=1}^p \mathbb{P}^q(Z_n^q \in \Gamma) \leq \mathbb{P}\Big( Z_n' \in \Gamma \Big) \leq c_2 \sum_{q=1}^p \mathbb{P}^q(Z_n^q \in \Gamma) \,.$$

Therefore the LDP for $Z_n'$ holds by (6.4). The last step in the proof is to augment $Z_n'$ to $Z_n$ by inserting zeros at at positions $i_1, \ldots, i_k$. Let $T$ denote this operation. The MDP for the sequence $Z_n$ then follows from observing that by Theorem 3.2 we have

$$\limsup_{n \to \infty} \frac{1}{b_n} \log \mathbb{P}(Z_n \neq T(Z_n')) \leq \limsup_{n \to \infty} \frac{1}{b_n} \log \mathbb{P}(\varphi_l^{E,\mathcal{C}}(\mathbf{y}_l^*) \neq \mathbf{m}_l^{*,\mathcal{C}}) = -\infty \,,$$

where $\mathbf{m}_l^{*,\mathcal{C}}$ denotes the vector of unconstrained moments in $\mathbf{m}_l^*$. This proves the asymptotic equivalence of $T(Z_n')$ and $Z_n$ and concludes the proof. $\qquad \square$

## 7. Proofs of the results in the uniform case

The uniform distribution on the moment space $\mathcal{M}_n^{\mathcal{C}}([0,1])$ is a special case of the distribution $\mathbb{P}_{n,[0,1],V}^{\mathcal{C}}$ defined in Section 3. It is however difficult to prove genericity of $V_1 = \cdots = V_{i_k+2} \equiv 0$ for *any* constraint $\mathcal{C}$ instead of just generic constraints. We will therefore consider a special parametrization of $\mathcal{M}_n^{\mathcal{C}}([0,1])$ that allows to exploit concavity properties of the density of the canonical moments.

**Lemma 7.1.** *Recall the definition of $\mathbf{m}_{i_k}^{\mathcal{C}}$ in (3.8) and denote by $\tilde{\mathcal{M}}_l^{\mathcal{C}}([0,1])$ the projection of $\mathcal{M}_l^{\mathcal{C}}([0,1])$ to the coordinates in $\mathbf{m}_l^{\mathcal{C}}$. For any $n \geq i_k$ the mapping*

$$\varphi_n^{\mathcal{C}} : \left\{ \begin{array}{ccc} \text{Int } \tilde{\mathcal{M}}_{i_k}^{\mathcal{C}} \times (0,1)^{n-i_k} & \to & \text{Int } \tilde{\mathcal{M}}_n^{\mathcal{C}}(E) \\ \big( \mathbf{m}_{i_k}^{\mathcal{C}}, p_{i_k+1}, \ldots, p_n \big) & \mapsto & \mathbf{m}_n^{\mathcal{C}} \end{array} \right. \tag{7.1}$$

*is a $C^\infty$-diffeomorphism with Jacobian*

$$\det \varphi_n^{\mathcal{C}} = \prod_{j=1}^n (p_j(1-p_j))^{n-\max(j,i_k)} \,.$$

*Proof.* The Jacobian matrix of $\varphi_n^{\mathcal{C}}$ is lower triangular with determinant

$$\det \varphi_n^{\mathcal{C}} = \prod_{\substack{l=1 \\ l \neq i_1, \ldots, i_k}}^{i_k} \frac{\partial m_l}{\partial m_l} \prod_{l=i_k+1}^n \frac{\partial m_l}{\partial p_l} = \prod_{l=i_k+1}^n \prod_{j=1}^{l-1} p_j(1-p_j) = \prod_{j=1}^n (p_j(1-p_j))^{n-\max(j,i_k)} \,,$$

where we used the previously applied formula (4.2) again. $\qquad \square$

In order to identify the limiting measures we need a specialization of Proposition 4.4.

**Corollary 7.2.** *Let $p_1, \ldots, p_r \in (0,1)$. The probability measure $\mu$ on $[0,1]$ corresponding to the infinite sequence of canonical moments $p_1, \ldots, p_r, \frac{1}{2}, \frac{1}{2}, \ldots$ is absolutely continuous with density*

$$\frac{\prod_{i=1}^{r} p_i q_i}{\pi R_r(x) \sqrt{x - x^2}},$$

*where $R_r(x)$ is a polynomial of degree at most $r$ that is strictly positive on the interval $[0,1]$. The coefficient of $x^r$ in $R_r$ is $1 - 2p_r$ and all coefficients depend continuously on $p_1, \ldots, p_r$. Furthermore, $R_r$ can be expressed as*

$$R_r(x) = \left( \frac{(P_j(x) - \frac{1}{2}(x - \frac{1}{2})P_{j-1}(x))^2}{x(1-x)} + \frac{1}{4}P_{j-1}^2(x) \right) \cdot \begin{cases} 16 & , \text{ if } r \text{ is even} \\ 4 & , \text{ if } r \text{ is odd} \end{cases},$$

*where $P_1, P_2, \ldots$ are the monic orthogonal polynomials corresponding to $\mu$ and $j := \left\lfloor \frac{r+4}{2} \right\rfloor$.*

*Proof.* By (4.22) we have $\alpha_l = \alpha = \frac{1}{2}$ for all $l > \frac{r+3}{2}$ and $\beta_{l-1} = \beta = \frac{1}{16}$ for all $l > \frac{r+4}{2}$. By Proposition 4.4 the measure $\mu$ therefore admits a density on the interval $[0,1]$ that is given by

$$\frac{\sqrt{x(1-x)}\beta_1 \cdots \beta_{j-1}}{2\pi D(x)},$$

where $j := \left\lfloor \frac{r+4}{2} \right\rfloor$ and

$$D(x) = \left( P_j(x) - \frac{1}{2}(x - \frac{1}{2})P_{j-1}(x) \right)^2 + \frac{1}{4}x(1-x)P_{j-1}^2(x).$$

Considering separately the cases where $r$ is odd and even, respectively, we obtain from (4.22)

$$\beta_1 \cdots \beta_{j-1} = p_{2j-2} \prod_{i=1}^{2j-3} p_i q_i = \prod_{i=1}^{r} p_i q_i \cdot \begin{cases} \frac{1}{8} & , r \text{ even} \\ \frac{1}{2} & , r \text{ odd} \end{cases}.$$

With the same arguments as in the proof of Theorem 3.6 we find that $D$ has zeros in 0 and 1 and the assertion follows from setting

$$R_r(x) := \frac{D(x)}{x(1-x)} \cdot \begin{cases} 16 & , r \text{ even} \\ 4 & , r \text{ odd} \end{cases}.$$

It remains to see that the degree of $R_r$ is at most $r$ and that the coefficient of $x^r$ is $1 - 2p_r$. If $r$ is odd, then by (4.7) the degree of $R_r(x)$ is at most $2j - 3 = r$ and the coefficient of $x^r$ in $R_r(x)$ is given by

$$4\left( \alpha_j - \frac{1}{2} \right) = 4\left( q_r p_{r+1} + q_{r+1}p_{r+2} - \frac{1}{2} \right) = 1 - 2p_r.$$

If $r$ is even, then by (4.7) the degree of $R_r(x)$ is at most $2j - 3 = r + 1$ and the coefficient of $x^{r+1}$ in $R_r(x)$ is given by

$$16\left( \alpha_j - \frac{1}{2} \right) = 16\left( q_{r+1}p_{r+2} + q_{r+2}p_{r+3} - \frac{1}{2} \right) = 0.$$

Consequently, the degree of $D(x)$ is at most $r$ and the coefficient of $x^r$ is given by

$$16\left( \beta_{j-1} - \frac{1}{16} \right) = 16\left( q_r p_{r+1} q_{r+1} p_{r+2} - \frac{1}{16} \right) = 1 - 2p_r.$$

$\square$

A key observation is contained in the following

**Lemma 7.3.** *The map*

$$W_n : \begin{cases} \text{Int } \mathcal{M}_n([0,1]) & \to & \mathbb{R} \\ (m_1, \ldots, m_n) & \mapsto & \log(4^n (m_{n+1}^+ - m_{n+1}^-)) \end{cases} \tag{7.2}$$

*is strictly concave and the Hessian matrix is negative definite in every point.*

*Proof.* Recall that a function $f : \Omega \to \mathbb{R}$ on a convex set $\Omega$ is called *strongly concave with modulus $c > 0$* if

$$f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda)f(y) + \frac{c}{2}\lambda(1-\lambda)\|x-y\|_2^2$$

holds for all $x, y \in \Omega$ and $\lambda \in [0,1]$. If $\Omega \subset \mathbb{R}^n$ and $f$ is twice differentiable, then this is equivalent to $\text{Hess } f(x) \leq -cI_n$ for all $x \in \Omega$, cf. (Hiriart-Urruty and Lemaréchal, 2001, Theorem 4.3.1)). We will show that for each point $m \in \text{Int } \mathcal{M}_n([0,1])$ there is a neighborhood $U$ in which $W_n$ is strongly concave with a positive modulus. Then by the mentioned equivalence, negative definiteness of $\text{Hess } W_n(m)$ will follow.

Let $m, m' \in \text{Int } \mathcal{M}_n([0,1])$ be two moment vectors in the interior of the $n$-th moment space. Let $\mu$ and $\mu'$ denote the probability measures with first $n$ moments given by $m$ and $m'$, respectively and whose canonical moments $p_i$ are $\frac{1}{2}$ for $i > n$. By Corollary 7.2, $\mu$ and $\mu'$ admit densities $f$ and $g$ w.r.t. the arcsine measure $\mu^0$. To proceed we use the remarkable identity

$$\mathcal{K}(\mu^0 \mid \mu) = -\sum_{i=1}^{\infty} \log(4p_i(1-p_i)) \tag{7.3}$$

between a measure $\mu$ on $[0,1]$ and its corresponding sequence of canonical moments. Identity (7.3) is a so-called sum rule, given in (Gamboa et al., 2016, p. 523). It can be deduced from Szegő's theorem (Gamboa et al. (2016)), for an interesting historical survey we refer to (Simon, 2011, p.29).

By (4.2) and (7.3), $W_n$ can be written in terms of the Kullback-Leibler divergence as

$$W_n(m) = \sum_{i=1}^{n} \log(4p_iq_i) = \sum_{i=1}^{\infty} \log(4p_iq_i) = -\mathcal{K}(\mu^0 \mid \mu),$$

$$W_n(m') = \sum_{i=1}^{n} \log(4p_iq_i) = \sum_{i=1}^{\infty} \log(4p_iq_i) = -\mathcal{K}(\mu^0 \mid \mu').$$

Let $\lambda \in [0,1]$ be arbitrary and denote by $\nu$ the measure whose first $n$ moments are given by $\lambda m + (1-\lambda)m'$ and which has canonical moments $p_i = \frac{1}{2}$ for $i > n$. Since $\nu$ and $\lambda\mu + (1-\lambda)\mu'$ agree on the first $n$ moments, they also have the same first $n$ canonical moments. This yields

$$W_n(\lambda m + (1-\lambda)m') = -\mathcal{K}(\mu^0|\nu) = \sum_{i=1}^{\infty} \log(4p_i(1-p_i)) \geq -\mathcal{K}(\mu^0|\lambda\mu + (1-\lambda)\mu').$$

To see the last inequality, we observe that the canonical moment $p_i$ of the measure $\lambda\mu+(1-\lambda)\mu'$ is for $i > n$ in general not $\frac{1}{2}$ and $\log(4p_i(1-p_i)) \leq 0$ with equality only for $p_i = \frac{1}{2}$.

Now note that for any constant $c > 0$ the function $x \mapsto \log(x)$ is strongly concave on the interval $(0, c]$ with modulus $\frac{1}{c^2}$, i.e. for all $x, y \in (0, c]$ we have

$$\log(\lambda x + (1-\lambda)y) \geq \lambda \log(x) + (1-\lambda)\log(y) + \frac{\lambda(1-\lambda)}{2c^2}(x-y)^2.$$

This yields

$$W_n(\lambda m + (1-\lambda)m') \geq -\mathcal{K}(\mu^0|\lambda\mu + (1-\lambda)\mu') = \int_0^1 \frac{\log(\lambda f(x) + (1-\lambda)g(x))}{\pi\sqrt{x-x^2}}\, dx$$

$$\geq \lambda W_n(m) + (1-\lambda)W_n(m') + \lambda(1-\lambda) \cdot \frac{1}{2}\int_0^1 \frac{(f(x)-g(x))^2}{\max\{f^2(x), g^2(x)\}\pi\sqrt{x-x^2}}\, dx.$$

By Hölder's inequality we have

$$\left(\int_0^1 \frac{x^m(f(x)-g(x))}{\pi\sqrt{x-x^2}}\,dx\right)^2 \le \left(\int_0^1 \frac{|f(x)-g(x)|}{\max\{f(x),g(x)\}\sqrt{\pi}\sqrt[4]{x-x^2}} \cdot \frac{\max\{f(x),g(x)\}}{\sqrt{\pi}\sqrt[4]{x-x^2}}\,dx\right)^2$$

$$\le \int_0^1 \frac{(f(x)-g(x))^2}{\max\{f^2(x),g^2(x)\}\pi\sqrt{x-x^2}}\,dx \int_0^1 \frac{\max\{f^2(x),g^2(x)\}}{\pi\sqrt{x-x^2}}\,dx.$$

If $f_\lambda$ denotes the density of $\mu$ with respect to the Lebesgue-measure, then

$$f_\lambda(x) = \frac{d\mu}{d\lambda}(x) = \frac{d\mu}{d\mu^0}(x)\cdot\frac{d\mu^0}{d\lambda}(x) = \frac{f(x)}{\pi\sqrt{x-x^2}}.$$

This results in

$$\int_0^1 \frac{f^2(x)}{\pi\sqrt{x-x^2}}\,dx = \int_0^1 (\pi\sqrt{x-x^2}f_\lambda(x))\cdot f_\lambda(x)\,dx \le \int_0^1 \frac{1}{R_n(x)}f_\lambda(x)\,dx,$$

where $R_n$ is the polynomial from Corollary 7.2. By the same corollary, we know that $R_n(x)$ is a strictly positive polynomial with coefficients depending on $p_1,\ldots,p_n$ in a continuous way and therefore locally in $p_1,\ldots,p_n$ we can bound $R_n(x)$ from below by some constant $K_n$ depending on $n$ and $p_1,\ldots,p_n$. Since $f_\lambda$ is a density, we have

$$\int_0^1 \frac{1}{R_n(x)}f_\lambda(x)\,dx \le \frac{1}{K_n}.$$

Combining these inequalities we get

$$W_n(\lambda m + (1-\lambda)m') \ge \lambda W_n(m) + (1-\lambda)W_n(m') + \lambda(1-\lambda)\cdot\frac{1}{2K_n n}\|m-m'\|_2^2,$$

i.e. $W_n$ is (locally) strongly concave. $\qquad\square$

Before giving the proofs of the main results of Section 2, a few words on the logic structure of the proofs are in order. We start with the LDP which will be used in the proof of the law of large numbers (Theorem 2.1) lateron. The limiting measure $\mu^{\mathcal{C}}$ was introduced in Section 2 in Theorem 2.1 but as we already need it in the proof of the LDP, we will define it as the unique minimizer of the Kullback-Leibler divergence $\mathcal{K}(\mu^0|\cdot)$ over $\mathcal{P}^{\mathcal{C}}([0,1])$. In particular, in the following proof we show that this minimization problem indeed has a unique solution. The convergence of the random moments to the moments of the minimizing measure and the representation of the minimizer will then be given in the proof of Theorem 2.1.

*Proof of Theorem 2.7.* Applying the parametrization of Lemma 7.1 to $(m_1^{(n)},\ldots,m_n^{(n)})$, we obtain a random vector $(\mathbf{m}_{i_k}^{\mathcal{C},(n)}, p_{i_k+1}^{(n)},\ldots,p_n^{(n)})$, where $p_j^{(n)}$ is beta$(n-j+1,n-j+1)$-distributed, the density of $\mathbf{m}_{i_k}^{\mathcal{C},(n)}$ is proportional to

$$\exp(-(n-i_k)\sum_{i=1}^{i_k}\log(p_i(1-p_i))) = \exp(-(n-i_k)\log(m_{i_k+1}^+ - m_{i_k+1}^-)), \qquad (7.4)$$

and we have used formula (4.2) for the last identity. Applying Theorem 5.2 and Lemma 5.1, we obtain an LDP for the coordinates $(\mathbf{m}_{i_k}^{\mathcal{C},(n)}, p_{i_k+1}^{(n)},\ldots,p_n^{(n)})$ with good rate function

$$I_1(\mathbf{m}_{i_k}^{\mathcal{C}}, p_{i_k+1},\ldots,p_l) := -\log(m_{i_k+1}^+ - m_{i_k+1}^-) - \sum_{j=i_k+1}^{l}\log(p_j(1-p_j)) + K, \qquad (7.5)$$

where $K$ is a constant that guarantees the infimum of $I_1$ being zero. By the contraction principle and equation (4.2) the vector $(m_1^{(n)},\ldots,m_l^{(n)})$ therefore satisfies an LDP with good rate function

$$I_2(m) := I(m) + K, \qquad (7.6)$$

where $I$ is defined in (2.8). From (7.5) we see that

$$K = \log(m_{i_k+1}^+(\mu^*) - m_{i_k+1}^-(\mu^*)) - (l - i_k)\log 4, \tag{7.7}$$

where $\mu^*$ is a measure with first $i_k$ canonical moments $p_1(\mu^*), \ldots, p_{i_k}(\mu^*)$ minimizing

$$(p_1, \ldots, p_{i_k}) \mapsto -\log(m_{i_k+1}^+(\mu^*) - m_{i_k+1}^-(\mu^*))$$

and higher canonical moments $p_j = \frac{1}{2}$ for all $j > i_k$. It remains to show that there is a unique such $\mu^*$ and that $\mu^* = \mu^{\mathcal{C}}$, where

$$\mu^{\mathcal{C}} := \underset{\mu \in \mathcal{P}^{\mathcal{C}}([0,1])}{\arg\min} \mathcal{K}(\mu^0|\mu).$$

For the uniqueness, note that since $\mathbf{m}_{i_k} \mapsto -\log(m_{i_k+1}^+ - m_{i_k+1}^-)$ is $+\infty$ on the boundary of $\mathcal{M}_{i_k}^{\mathcal{C}}([0,1])$ and strictly convex on the interior by Lemma 7.2, it has a unique minimizer which is attained on $\mathrm{Int}\,\mathcal{M}_{i_k}^{\mathcal{C}}([0,1])$. To see $\mu^* = \mu^{\mathcal{C}}$, recall that whether a measure $\mu \in \mathcal{P}([0,1])$ fulfills constraint $\mathcal{C}$ or not, is determined by the first $i_k$ canonical moments $p_1, \ldots, p_{i_k}$ only. Thus by (7.3), a measure $\mu \in \mathcal{P}^{\mathcal{C}}([0,1])$ minimizing $\mathcal{K}(\mu^0|\cdot)$ has to have canonical moments $p_j = \frac{1}{2}$ for all $j > i_k$ and hence

$$\mathcal{K}(\mu^0|\mu) = -\sum_{j=1}^{i_k} \log(4p_j(1 - p_j)) = -\log(m_{i_k+1}^+(\mu) - m_{i_k+1}^-(\mu)) - i_k \log 4. \tag{7.8}$$

This implies $\mu^* = \mu^{\mathcal{C}}$ and finishes the proof. $\qquad\square$

*Proof of Theorem 2.8.* By (7.6), (7.7), (7.3) and (7.8), for any $l$ the sequence $(m_1(\mu^{(n)}), \ldots, m_l(\mu^{(n)}))$ satisfies an LDP with good rate function

$$I(\mathbf{m}_l) = \mathcal{K}(\mu^0 \mid \mu(\mathbf{m}_l)) - \mathcal{K}(\mu^0 \mid \mu^{\mathcal{C}}),$$

where $\mathbf{m}_l$ is an $l$-dimensional vector of moments and $\mu(\mathbf{m}_l)$ is the measure that has $\mathbf{m}_l$ as its first $l$ moments and canonical moments $p_j = \frac{1}{2}$ for all $j > l$. By the Dawson-Gärtner theorem (Dembo and Zeitouni, 1998, Theorem 4.6.1) we then obtain an LDP for the infinite sequence of moments $(m_1(\mu^{(n)}), m_2(\mu^{(n)}), \ldots)$ with good rate function

$$I(\mathbf{m}_\infty) = \sup_{l \geq 0} \mathcal{K}(\mu^0 \mid \mu(\mathbf{m}_l)) - \mathcal{K}(\mu^0 \mid \mu^{\mathcal{C}}),$$

where $\mathbf{m}_\infty$ is the infinite vector of moments containing the $\mathbf{m}_l$'s. By (7.3) the term

$$\mathcal{K}(\mu^0 \mid \mu(\mathbf{m}_l)) = -\sum_{j=1}^{l} \log(4p_i q_i)$$

is increasing in $l$ and we therefore obtain

$$I(\mathbf{m}_\infty) = \sup_{l \geq 0} \mathcal{K}(\mu^0 \mid \mu(\mathbf{m}_l)) - \mathcal{K}(\mu^0 \mid \mu^{\mathcal{C}}) = \mathcal{K}(\mu^0 \mid \mu(\mathbf{m}_\infty)) - \mathcal{K}(\mu^0 \mid \mu^{\mathcal{C}}),$$

where $\mu(\mathbf{m}_\infty)$ is the measure with moments $\mathbf{m}_\infty$. The theorem now follows from applying the contraction principle to transfer the LDP from the sequence $(m_1(\mu^{(n)}), m_2(\mu^{(n)}), \ldots)$ to the measure $\mu^{(n)}$. This is possible since $\mu^{(n)}$ is compactly supported and hence uniquely determined by its moments. $\qquad\square$

*Proof of Theorem 2.1.* Recall from the proof of Theorem 2.7 that $\mathbf{m}_l^{(n)}$ satisfies an LDP with good rate function $I - I(\mathbf{m}_l(\mu^{\mathcal{C}}))$, where $\mu^{\mathcal{C}}$ has been defined as the unique minimizer of $\mathcal{K}(\mu^0|\cdot)$ over $\mathcal{P}^{\mathcal{C}}([0,1])$. The convergence stated in Theorem 2.1 now follows from an application of the Borel-Cantelli lemma. The claimed representation of $\mu^{\mathcal{C}}$ is then provided by Corollary 7.2. $\quad\square$

Next, we give the computation of the measures in Example 2.3.

**Example 7.4** (Continuation of Example 2.3)**.** We will consider the different constraints $\mathcal{C}_1$, $\mathcal{C}_2$ and $\mathcal{C}_3$ separately.

**Case $\mathcal{C}_1 = \{m_1 = c_1\}$:** As seen in the proof of Theorem 2.1, the measure $\mu^{\mathcal{C}_1}$ is uniquely determined by having canonical moments $p_1 = c_1$ and $p_j = \frac{1}{2}$ for all $j > 1$. We may thus apply Corollary 7.2 to obtain the representation of $\mu^{\mathcal{C}_1}$. By formula (4.22) the recursion parameters of $\mu^{\mathcal{C}_1}$ are given by $\alpha_1 = c_1$, $\alpha_2 = \frac{3-2c_1}{4}$, $\beta_1 = \frac{c_1(1-c_1)}{2}$ and $\alpha_j = \frac{1}{2}$, $\beta_{j-1} = \frac{1}{16}$ for all $j > 2$. The orthogonal polynomials of $\mu^{\mathcal{C}_1}$ up to order 2 can hence be calculated as

$$P_0(x) = 1$$
$$P_1(x) = x - c_1$$
$$P_2(x) = (x - \alpha_2)P_1(x) - \beta_1 P_0(x) = x^2 - \frac{3 + 2c_1}{4}x + \frac{c_1}{4}$$

and we obtain

$$R_1(x) = 4\left(\frac{P_2(x) - \frac{1}{2}(x - \frac{1}{2})P_1(x))^2}{x(1-x)} + \frac{1}{4}P_1^2(x)\right) = x(1-x) + (x - c_1)^2.$$

The stated form of the measure now follows from Corollary 7.2.

**Case $\mathcal{C}_2 = \{m_1 = c_1, m_2 = c_2\}$:** As in the first case, the measure $\mu^{\mathcal{C}_2}$ is uniquely determined by having canonical moments $p_1 = c_1$, $p_2 = \frac{m_2 - m_2^-}{m_2^+ - m_2^-} = \frac{c_2 - c_1^2}{c_1(1-c_1)}$ and $p_j = \frac{1}{2}$ for all $j > 2$. Analogously to the first case we calculate

$$R_2(x) = (1 - p_2)^2(x - p_1)^2 + p_2^2 x(1-x) = \frac{(c_1 - c_2)^2(x - c_1)^2 + (c_2 - c_1^2)^2 x(1-x)}{c_1^2(1-c_1)^2}$$

and the representation of the measure follows again from Corollary 7.2. Note here that

$$p_1 q_1 p_2 q_2 = c_1(1 - c_1)\frac{c_2 - c_1^2}{c_1(1-c_1)}\frac{c_1 - c_2}{c_1(1-c_1)} = \frac{(c_2 - c_1^2)(c_1 - c_2)}{c_1(1-c_1)}.$$

**Case $\mathcal{C}_3 = \{m_2 = c_2\}$:** As seen in the proof of Theorem 2.1, the moment $m_1$ of $\mu^{\mathcal{C}_3}$ maximizes the moments range $m_3^+ - m_3^-$ under all measures satisfying $\mathcal{C}_3$. Recalling (4.2), we have for any measure with $m_2 = c_2$

$$m_3^+ - m_3^- = p_1 q_1 p_2 q_2 = \frac{(c_2 - m_1^2)(m_1 - c_2)}{m_1(1 - m_1)},$$

which (as a function of $m_1$) has a unique maximizer $m_1^*$ by Lemma 7.3. Therefore the canonical moments of $\mu^{\mathcal{C}_3}$ are given by $p_1 = m_1^*$, $p_2 = \frac{c_2 - (m_1^*)^2}{m_1^*(1 - m_1^*)}$ and $p_j = \frac{1}{2}$ for all $j > 2$. The assertion now follows from the calculations in the case $\mathcal{C}_2 = \{m_1 = m_1^*, m_2 = c_2\}$.

*Proof of Theorem 2.5.* Recall that the coordinates $(\mathbf{m}_{i_k}^{\mathcal{C},(n)}, p_{i_k+1}^{(n)}, \ldots, p_l^{(n)})$ have distributions $p_j^{(n)} \sim \text{beta}(n - j + 1, n - j + 1)$ and $\mathbf{m}_{i_k}^{\mathcal{C},(n)}$ has a density proportional to

$$\exp(n \log(m_{i_k+1}^+ - m_{i_k+1}^-))(m_{i_k+1}^+ - m_{i_k+1}^-)^{-i_k} \mathbb{1}_{\text{Int } \mathcal{M}_{i_k}^{\mathcal{C}}([0,1])}(m).$$

With Lemma 7.3 the theorem follows by an application of Proposition 6.1 and the delta-method analogously to the proof of Theorem 3.9. Note that the minimizer of the map $m \mapsto -\log(m_{i_k+1}^+ - m_{i_k+1}^-)$ is unique and given by the moments $(m_1(\mu^{\mathcal{C}}), \ldots, m_{i_k}(\mu^{\mathcal{C}}))$ by the same arguments as in the proof of Theorem 2.1. The covariance matrix $\Sigma_l$ is given as

$$\Sigma_l := T^{\mathcal{C}} \tilde{\Sigma}_l, \tag{7.9}$$

$$\tilde{\Sigma}_l := \mathcal{D}^t \, \text{diag}\left(\text{Hess}^{\mathcal{C}}(W_{i_k})(\mathbf{m}_{i_k}(\mu^{\mathcal{C}})), \frac{1}{8}, \ldots, \frac{1}{8}\right)\mathcal{D},$$

$$\mathcal{D} := D\varphi_l^{\mathcal{C}}\left(\mathbf{m}_{i_k}(\mu^{\mathcal{C}}), \frac{1}{2}, \ldots, \frac{1}{2}\right),$$

where $T^{\mathcal{C}} : \mathbb{R}^{(l-k)\times(l-k)} \to \mathbb{R}^{l\times l}$ denotes the transformation of an $(l-k) \times (l-k)$ matrix by inserting rows and columns of zeros at the positions $i_1, \ldots, i_k$. The maps $\varphi_l^{\mathcal{C}}$ and $W_{i_k}$ are defined in (7.1) and (7.2) and $\mathrm{Hess}^{\mathcal{C}}$ denotes the Hessian with respect to the coordinates in $\mathbf{m}_{i_k}^{\mathcal{C}}$. $\qquad\square$

*Proof of Theorem 2.6.* The proof follows by an application of Proposition 6.2 analogous to the proof of Theorem 2.5. Note that the delta-method for large deviations must be applied similar to the proof of Theorem 3.10. $\qquad\square$

*Proof of Proposition 2.4.* By Lemma 7.1 and (7.4), the volume of the restricted space is given by

$$\mathrm{vol}_{n-k}\left(\mathcal{M}_n^{\mathcal{C}}([0,1])\right)$$

$$= \int\limits_{\mathcal{M}_{i_k}^{\mathcal{C}}([0,1])} \exp\left(-(n-i_k)\log(m_{i_k+1}^+ - m_{i_k+1}^-)\right) dm \prod_{j=i_k+1}^{n} B(n-j+1, n-j+1),$$

where $B(a,b) := \int_0^1 x^{a-1}(1-x)^{b-1}\,dx$ is the beta function. By an application of Laplace's method and the minimization property of $\mu^C$, the first factor satisfies

$$\int\limits_{\mathcal{M}_{i_k}^{\mathcal{C}}([0,1])} \exp\left((n-i_k)\log(m_{i_k+1}^+ - m_{i_k+1}^-)\right) dm$$

$$= n^{-(i_k-k)/2}\left((m_{i_k+1}^+ - m_{i_k+1}^-)(\mu^{\mathcal{C}})\right)^{n-i_k}\left(\frac{(2\pi)^{(i_k-k)/2}}{\sqrt{d^{\mathcal{C}}}} + o(1)\right),$$

where the constant $d^{\mathcal{C}}$ is the determinant of the Hessian of the map $m \mapsto -\log(m_{i_k+1}^+ - m_{i_k+1}^-)$ with respect to the coordinates $\mathbf{m}_{i_k}^{\mathcal{C}}$, evaluated in $(m_1(\mu^{\mathcal{C}}), \ldots, m_{i_k}(\mu^{\mathcal{C}}))$.

For the second factor, observe that by (2.6)

$$\prod_{j=i_k+1}^{n} B(n-j+1, n-j+1) = \mathrm{vol}_n\left(\mathcal{M}_n([0,1])\right)\prod_{j=1}^{i_k} B(n-j+1, n-j+1)^{-1}$$

holds and another application of Laplace's method yields

$$B(n-j+1, n-j+1) = n^{-1/2} 4^{-(n-j)}\left(\frac{\sqrt{\pi}}{2} + o(1)\right).$$

Combining these equations then gives

$$\frac{\mathrm{vol}_{n-k}\left(\mathcal{M}_n^{\mathcal{C}}([0,1])\right)}{\mathrm{vol}_n\left(\mathcal{M}_n([0,1])\right)} = n^{k/2}\left(4^{i_k}(m_{i_k+1}^+ - m_{i_k+1}^-)(\mu^{\mathcal{C}})\right)^{n-i_k}\left(\frac{2^{i_k^2+(i_k-k)/2}}{\sqrt{d^{\mathcal{C}}}\pi^{k/2}} + o(1)\right).$$

$\qquad\square$

## References

Akemann, G., Baik, J., and Di Francesco, P., editors (2011). *The Oxford handbook of random matrix theory*. Oxford University Press, Oxford.

Anderson, G., Guionnet, A., and Zeitouni, O. (2010). *An introduction to random matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge.

Chang, F. C., Kemperman, J. H. B., and Studden, W. J. (1993). A normal limit theorem for moment sequences. *Ann. Probab.*, 21(3):1295–1309.

Chihara, T. S. (1978). *An introduction to orthogonal polynomials*. Gordon and Breach Science Publishers, New York-London-Paris. Mathematics and its Applications, Vol. 13.

Csiszár, I. (1967). Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.*, 2:299–318.

Deift, P., Kriecherbauer, T., and McLaughlin, K. T.-R. (1998). New results on the equilibrium measure for logarithmic potentials in the presence of an external field. *J. Approx. Theory*, 95(3):388–475.

Deift, P. A. (1999). *Orthogonal polynomials and random matrices: a Riemann-Hilbert approach*, volume 3 of *Courant Lecture Notes in Mathematics*. New York University, Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI.

Dembo, A. and Zeitouni, O. (1998). *Large deviations techniques and applications*, volume 38 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition.

Dette, H. and Nagel, J. (2012). Distributions on unbounded moment spaces and random moment sequences. *Ann. Probab.*, 40(6):2690–2704.

Dette, H. and Studden, W. J. (1997). *The theory of canonical moments with applications in statistics, probability, and analysis*. Wiley Series in Probability and Statistics: Applied Probability and Statistics. John Wiley & Sons, Inc., New York. A Wiley-Interscience Publication.

Dette, H., Tomecki, D., and Venker, M. (2018). Universality in random moment problems. *Electron. J. Probab.*, 23:23 pp.

Franklin, J. N. (1968). *Matrix theory*. Prentice-Hall, Inc., Englewood Cliffs, N.J.

Gamboa, F. and Lozada-Chang, L.-V. (2004). Large deviations for random power moment problem. *Ann. Probab.*, 32(3B):2819–2837.

Gamboa, F., Nagel, J., and Rouault, A. (2016). Sum rules via large deviations. *J. Funct. Anal.*, 270(2):509–559.

Gamboa, F. and Rouault, A. (2010). Canonical moments and random spectral measures. *J. Theoret. Probab.*, 23(4):1015–1038.

Gao, F. and Zhao, X. (2011). Delta method in large deviations and moderate deviations for estimators. *Ann. Statist.*, 39(2):1211–1240.

Geronimo, J. S. and Iliev, P. (2017). Bernstein-Szegő measures, Banach algebras, and scattering theory. *Trans. Amer. Math. Soc.*, 369(8):5581–5600.

Hiriart-Urruty, J.-B. and Lemaréchal, C. (2001). *Fundamentals of convex analysis*. Grundlehren Text Editions. Springer-Verlag, Berlin. Abridged version of *it Convex analysis and minimization algorithms. I* [Springer, Berlin, 1993; MR1261420 (95m:90001)] and *it II* [ibid.; MR1295240 (95m:90002)].

Johansson, K. (1998). On fluctuations of eigenvalues of random Hermitian matrices. *Duke Math. J.*, 91(1):151–204.

Karlin, S. and Shapley, L. S. (1953). Geometry of moment spaces. *Mem. Amer. Math. Soc.*, No. 12:93.

Karlin, S. and Studden, W. J. (1966). *Tchebycheff systems: With applications in analysis and statistics*. Pure and Applied Mathematics, Vol. XV. Interscience Publishers John Wiley & Sons, New York-London-Sydney.

Killip, R. and Nenciu, I. (2004). Matrix models for circular ensembles. *Int. Math. Res. Not.*, (50):2665–2701.

Klartag, B. (2007). A central limit theorem for convex sets. *Invent. Math.*, 168(1):91–131.

Kreĭn, M. G. and Nudel'man, A. A. (1977). *The Markov moment problem and extremal problems.* American Mathematical Society, Providence, R.I. Ideas and problems of P. L. Čebyšev and A. A. Markov and their further development, Translated from the Russian by D. Louvish, Translations of Mathematical Monographs, Vol. 50.

Matsumoto, Y. (2002). *An introduction to Morse theory*, volume 208 of *Translations of Mathematical Monographs.* American Mathematical Society, Providence, RI. Translated from the 1997 Japanese original by Kiki Hudson and Masahico Saito, Iwanami Series in Modern Mathematics.

Saff, E. and Totik, V. (1997). *Logarithmic potentials with external fields*, volume 316 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer-Verlag, Berlin. Appendix B by Thomas Bloom.

Simon, B. (2011). *Szegő's theorem and its descendants.* M. B. Porter Lectures. Princeton University Press, Princeton, NJ. Spectral theory for $L^2$ perturbations of orthogonal polynomials.

Skibinsky, M. (1967). The range of the $(n + 1)$th moment for distributions on $[0, 1]$. *J. Appl. Probability*, 4:543–552.

Skibinsky, M. (1968). Extreme $n$th moments for distributions on $[0, 1]$ and the inverse of a moment space map. *J. Appl. Probability*, 5:693–701.

Skibinsky, M. (1969). Some striking properties of binomial and beta moments. *Ann. Math. Statist.*, 40:1753–1764.

Skibinsky, M. (1986). Principal representations and canonical moment sequences for distributions on an interval. *J. Math. Anal. Appl.*, 120(1):95–118.

Szegő, G. (1975). *Orthogonal polynomials.* American Mathematical Society, Providence, R.I., fourth edition. American Mathematical Society, Colloquium Publications, Vol. XXIII.

[*]Department of Mathematics, Ruhr University Bochum, 44780 Bochum, Germany
*E-mail address*: holger.dette@ruhr-uni-bochum.de

[†]Department of Mathematics, Ruhr University Bochum, 44780 Bochum, Germany
*E-mail address*: dominik.tomecki@ruhr-uni-bochum.de

[†]Department of Mathematics, Ruhr University Bochum, 44780 Bochum, Germany
*E-mail address*: martin.venker@ruhr-uni-bochum.de