On Hausdorff's moment problem in higher dimensions

Oliver Knill, *

January 4, 2000

Abstract

Denote by $\mu_n = \int x_1^{n_1} \dots x_d^{n_d} \; d\mu(x)$ the n'th moment of a Borel probability measure μ on the unit cube $I^d = [0,1]^d$ in $\mathbf{R^d}$. Generalizing results of Hausdorff, Hildebrandt and Schoenberg, we give a sufficient condition in terms of moments, that μ is absolutely continuous with respect to a second Borel measure ν on I^d.
We also review a constructive approximation of measures by atomic measures using finitely many moments and relate the Fourier theory of a measure with the moment problem.

Keywords : Moment problems
Abreviated title : Hausdorff's Moment problem.
AMS subject Classification: 44A60

1 Introduction

Moment problems occur in different mathematical contexts like in probability theory, mathematical physics, statistical mechanics, potential theory, constructive analysis or dynamical systems. In one dimension, the problem is to get information about a measure μ, if the moments $\int x^n \; d\mu(x)$ are known. For example, in probability theory, the question can occur to find properties of the law of a bounded random variable X, if the moments $\mathrm{E}[X^n]$ are known. In mathematical physics, the problem appears to make statements about the spectral type of a bounded selfadjoint operator $L : H \to H$, if the values $(\psi, L^n \psi)$ for $\psi \in H$ are known. One would for example like to see if a spectral measure μ_ψ of some random operator is absolutely continuous with respect to the density of states. A problem in complex dynamics is to get information on an equilibrium measure for a Julia set J on the real line, for which the moments are known from the asymptotic behavior of the logarithmic potential of J at infinity.
Also in higher dimensions, information about a measure μ in terms of moments $\mu_n = \int x_1^{n_1} \dots x_d^{n_d} \; d\mu(x)$ can be useful. In probability theory, the problem arrizes to determine the multi-dimensional distribution of a random vector X. In dynamical systems, the problem can occur that one wants to get information about an ergodic invariant measure μ supported by a compact invariant set K of a dynamical system $S : \mathbf{R^d} \to \mathbf{R^d}$. If $S^k(x) \in \mathbf{R^d}$ is an orbit of S, then the moment μ_n can be determined by Birkhoff's ergodic theorem as $\lim_{N \to \infty} N^{-1} \sum_{k=0}^{N-1} \prod_{i=1}^d (S^k(x))_i^{n_i}$ for μ almost all x. One would especially like to have a criterion which assures that μ is absolutely continuous with respect to Lebesgue measure or some equilibrium measure on a compact attractor of S.
An other motivation for multi dimensional moment problems is the fact that Hausdorff's solution of the moment problem provides a explicit finite dimensional approximation of a measure by finite atomic measures. While some implementations for use for encoding and

*Department of Mathematics, University of Arizona, Tucson, AZ 85721, USA

decoding of images are reportedly numerically unstable (see [2]), it should be helpful at least together with other methods. While encoding the picture into a moment sequence is time consuming, the reconstruction of the approximating measure in Corollary 2.3 is quite fast. One should also note that for simulating PDE's with discontinuous solutions, if the behavior of small moments is easy to describe, then there is a finite dimensional approximation of the real flow by evolving the moments. This is useful for some PDE's like the Bolzmann equation, where small moments are of primary physical interest.

Hausdorff's solution to the moment problem has been extended to higher dimensions by Hildebrandt and Schoenberg [7]. Hausdorff gave in one dimension also sufficient conditions for a measure μ to be absolutely continuous to Lebesgue measure ν [6]. We note here that this result generalizes to the case, when ν is a general Borel measure and that it holds also in higher dimensions. Also the property of a measure to be in $L^p(I^d)$ proved by Hausdorff in one dimension has a higher dimensional generalization. These three observations seem to be new. We hope also, that this self-contained note might be useful, because some relevant research articles are quite old and the topic seems not be well represented in newer textbooks. For more information about the multi-dimensional moment problem, especially for not compactly supported measures (which we are not addressing here) and more references see [5]. Classical references for the moment problem are [12, 3].

There is a close relation by representing a measure through moments or by Fourier transform. This relation is useful for speedy calculation of Fourier coefficients of spectral measures [9]. We review the link between moments and Fourier transform in one dimension: if $\hat{\mu}_n = \int \cos(nx) \; d\mu(x) dx$ are the Fourier coefficients of a measure μ on $[0, \pi]$, a change of variables $u = \cos(x), du = \sin(x) dx$ leads to $\hat{\mu}_n = \int T_n(u) d\nu(u) du$, where $\nu(u) = \mu(\arccos(u))/\sqrt{1-u^2}$ is a measure on $[-1, 1]$ and $T_n(u) = \cos(n \arccos(u))$ are Chebychev polynomials. The point is that each Fourier coefficient of ν is a finite sum of moments and that ν is a distorded version of μ, sharing all regularity properties with μ. In applications this is useful, if μ is a spectral measure of an operator L because then $\int p(x) \; d\mu(x) = (\psi, p(L)\psi)$ for a function p.

For example if L is a finite difference operator $a\Delta + V$ on $\mathbf{Z^d}$ and $\psi(n) = 1$ if and only if $n = 0 \in \mathbf{Z^d}$, then $\int p(x) \; d\mu(x) = p(L)_{00}$. In the case $p(x) = T_n(x)$, we have $T_n(L)_{00} = (A_{11}^n)_{00}$, where $A = \begin{pmatrix} L & -1 \\ 1 & 0 \end{pmatrix}$ is a 2x2 matrix with bounded operators as entries. In other words, the Fourier coefficients of ν are $(A_{11}^n)_{00}$ are polynomials of degree n in a.

How do we transform back? We have

$$\sum_n \hat{\mu}_n T_n(u) = \sum_n \hat{\mu}_n \cos(n \arccos(u)) = \sum_n \hat{\mu}_n \cos(nx) = \mu(x) = \mu(\arccos(u)) \; .$$

If we take N Fourier coefficients, then a slightly deformed measure of μ is approximated by a finite polynomial. By putting the measure μ on $[-b, b]$, the distortion is of the order b^3 and for smaller b, the polynomial $\sum_{n=0}^N \hat{\mu}_n T_n(u)$ approximates very well the measure μ.

This can now be generalized to higher dimensions: instead of computing the Fourier coefficients of a measure $[0, \pi]^d$, it is easier to compute the Fourier coefficients of the measure $\nu(u) du = \mu(Tu) dT^{-1}(u) dx$ on $[-1, 1]^d$, where $u = Tx, du = dT(u) dx$, with $T(x_1, \dots, x_d) = (\arccos(x_1), \dots, \arccos(x_d))$. After whatever necessary is done with the Fourier transform, we can transform back and obtain back the measure μ, rsp. a small distortion of it.

So what's the deal? Does replacing the Fourier transform by moments change anything? Yes it does: the punch line is that a small number of moments determine a large number of Fourier coefficients: while the Fourier coefficients $\int \exp(inx)\, d\mu(x)$ do not decrease in general, the moments $\int x^n\, d\mu(x)$ decay always exponentially fast if the mesure μ is contained in an interval $[-a,a]$ with $a<1$. So, only a few moments are relevant but still, we will be able to compute with them a large number of Fourier coefficients $\int T_n(x)\, d\nu(x)$ of μ.

2 Characterization of positive measures

Let μ be a bounded Borel measure on the unit cube $I^d=[0,1]^d$. We use the multi-index notation $x^n=\prod_i x_i^{n_i}$ and denote by $\mu_n=\int_{I^d} x^n\, d\mu$ the n'th moment of μ. We call the map $n\in \mathbf{N^d} \mapsto \mu_\mathbf{n}$ a configuration or, if $d=1$, a sequence. We will tacitely assume $\mu_n=0$, if at least one coordinate n_i is negative. Let e_i be the standard basis in $\mathbf{Z^d}$. Define the partial difference $(\Delta_i a)_n = a_{n-e_i}-a_n$ on configurations and write $\Delta^k=\prod_i \Delta_i^{k_i}$. (Opposit to the usual convention, we took a different sign convention in Δ, because this allows to avoid all signs in this text). By induction in $\sum_{i=1}^d n_i$, one proves the relation

$$(\Delta^k\mu)_n=\int_{I^d} x^{n-k}(1-x)^k\, d\mu \tag{1}$$

using $x^{n-e_i-k}(1-x)^k - x^{n-k}(1-x)^k = x^{n-e_i-k}(1-x)^{k+e_i}$. To improve readability, we also use notation like $\frac{k}{n}=\prod_{i=1}^n \frac{k_i}{n_i}$ or $\binom{n}{k}=\prod_{i=1}^d \binom{n_i}{k_i}$ or $\sum_{k=0}^n=\sum_{k_1=0}^{n_1}\cdots\sum_{k_d=0}^{n_d}$. We mean $n\to\infty$ in the sense that $n_i\to\infty$ for all $i=1\dots d$.

Given a continuous function $f: I^d\to\mathbf{R}$. For $n\in\mathbf{N^d}, \mathbf{n_i}>\mathbf{0}$ we define the higher dimensional *Bernstein polynomials*

$$B_n(f)(x)=\sum_{k=0}^n f(\frac{k_1}{n_1},\dots,\frac{k_d}{n_d})\binom{n}{k} x^k(1-x)^{n-k}\ .$$

These polynomials were used in [7, 12].

Lemma 2.1 (Bernstein) *In the uniform topology in $C(I^d)$, we have $B_n(f)\to f$ if $n\to\infty$.*

Proof. By Weierstrass theorem, multi-dimensional polynomials are dense in $C(I^d)$ as they separate points in $C(I^d)$. It is therefore enough to prove the claim for $f(x)=x^m=\prod_{i=1}^d x_i^{m_i}$. Because $B_n(y^m)(x)$ is the product of one dimensional Bernstein polynomials $B_n(y^m)(x)=\prod_{i=1}^d B_{n_i}(y_i^{m_i})(x_i)$, the claim follows from the result in one dimensions, which has a convenient proof using the weak law of large numbers (see for example [14]). □

Remark. Hildebrandt and Schoenberg refer for the proof of Lemma (2.1) to Bernstein's proof in one dimension. While a higher dimensional adaptation of the probabilistic proof could be done involving a stochastic process in $\mathbf{Z^d}$ with drift x_i in the i'th direction, the above given factorization argument is more elegant.

Theorem 2.2 (Hausdorff,Hildebrandt-Schoenberg) *There is a bijection between signed bounded Borel measures μ on $[0,1]^d$ and configurations μ_n for which there exists a constant C such that*

$$\sum_{k=0}^n |\binom{n}{k}(\Delta^k\mu)_n| \le C,\ \forall n\in\mathbf{N^d}\ . \tag{2}$$

A configuration μ_n belongs to a positive measure if and only if additionally to (2) one has $(\Delta^k\mu)_n\ge 0$ for all $k,n\in\mathbf{N^d}$.

Proof. (i) Because by Lemma (2.1), polynomials are dense in $C(I^d)$, there exists a unique solution to the moment problem. We show now existence of a measure μ under condition (2). For a measures μ, define for $n\in\mathbf{N^d}$ the atomic measures $\mu^{(n)}$ on I^d which have weights $\binom{n}{k}(\Delta^k\mu)_n$ on the $\prod_{i=1}^d(n_i+1)$ points $(\frac{n_1-k_1}{n_1},\dots,\frac{n_d-k_d}{n_d})\in I^d$ with $0\le k_i\le n_i$. Because

$$\begin{aligned}
\int_{I^d} x^m d\mu^{(n)}(x) &= \sum_{k=0}^n \binom{n}{k}(\frac{n-k}{n})^m(\Delta^k\mu)_n \\
&= \int_{I^d}\sum_{k=0}^n \binom{n}{k}(\frac{n-k}{n})^m x^{n-k}(1-x)^k\, d\mu(x) \\
&= \int_{I^d}\sum_{k=0}^n \binom{n}{k}(\frac{k}{n})^m x^k(1-x)^{n-k}\, d\mu(x) \\
&= \int_{I^d} B_n(y^m)(x)\, d\mu(x) \to \int_0^1 x^m\, d\mu(x)\ ,
\end{aligned}$$

we know that any signed measure μ which is an accumulation point of $\mu^{(n)}$, where $n_i\to\infty$ solves the moment problem. The condition (2) implies that the variation of the measures $\mu^{(n)}$ is bounded. By Alaoglu's theorem, there exists an accumulation point μ.

(ii) The left hand side of (2) is the variation $||\mu^{(n)}||$ of the measure $\mu^{(n)}$. Because by (i) $\mu^{(n)}\to\mu$, and μ has finite variation, there exists a constant C such that $||\mu^{(n)}||\le C$ for all n. This establishes (2).

(iii) We see that if $(\Delta^k\mu)_n\ge 0$ for all k, then the measures $\mu^{(n)}$ are all positive and therefore also the measure μ.

(iv) If μ is a positive measure, then by (1)

$$\binom{n}{k}(\Delta^k\mu)_n=\binom{n}{k}\int_{I^d} x^{n-k}(1-x)^k\, d\mu(x)\ge 0\ .$$

□

Remark.
1) This proof is adapted from the proof given in [10] for $d=1$.
2) Hildebrandt and Schoenberg noted in 1933, that this result gives a *constructive proof* of Riesz representation theorem stating that the dual of $C(I^d)$ is the space of Borel measures $M(I^d)$.
Let $\delta(x)$ denote the Dirac measure located on $x\in I^d$. We extract from the proof of Theorem 2.2 the construction:

Corollary 2.3 *An explicit finite constructive approximations of a given measure μ on I^d is given for $n\in\mathbf{N^d}$ by the atomic measures*

$$\mu^{(n)}=\sum_{0\le k_i\le n_i}\binom{n}{k}(\Delta^k\mu)_n\delta((\frac{n_1-k_1}{n_1},\dots,\frac{n_d-k_d}{n_d}))\ .$$

3 Comparing two measures

Hausdorff established a criterion for absolutely continuity of a measure μ with respect to the Lebesgue measure on $[0,1]$ (see [10, 6]). This can be generalized to find a criterion for comparing two arbitrary measures and works in d dimensions.

Call a measure μ on I^d *uniformly absolutely continuous* with respect to ν, if it satisfies $\mu = f\, d\nu$ with $f \in L^\infty(I^d)$.

Corollary 3.1 *A positive probability measure μ is uniformly absolutely continuous with respect to a second probabiliy measure ν if and only if there exists a constant C such that $(\Delta^k \mu)_n \leq C \cdot (\Delta^k \nu)_n$ for all $k, n \in \mathbf{N^d}$.*

Proof. If $\mu = f\nu$ with $f \in L^\infty(I^d)$, we get using (1)

$$\begin{aligned} (\Delta^k \mu)_n &= \int_{I^d} x^{n-k}(1-x)^k \, d\mu(x) \\ &= \int_{I^d} x^{n-k}(1-x)^k \, f d\nu(x) \\ &\leq ||f||_\infty \int_{I^d} x^{n-k}(1-x)^k \, d\nu(x) \\ &= ||f||_\infty (\Delta^k \nu)_n \; . \end{aligned}$$

On the other hand, if $(\Delta^k \mu)_n \leq C(\Delta^k \nu)_n$ then $\rho_n = C(\Delta^k \nu)_n - (\Delta^k \mu)_n$ defines by Theorem 2.2 a positive measure ρ on I^d. Since $\rho = C\nu - \mu$, we have for any Borel set $A \subset I^d$ $\rho(A) \geq 0$. This gives $\mu(A) \leq C\nu(A)$ and implies that μ is absolutely continuous with respect to ν with a function f satisfying $f(x) \leq C$ almost everywhere. □

This gives especially the following higher dimensional generalization of Hausdorff's result:

Corollary 3.2 *A Borel probability measure μ on I^d is uniformly absolutely continuous with respect to Lebesgue measure on I^d if and only if $|\Delta^k \mu_n| \leq \begin{pmatrix} n \\ k \end{pmatrix} \prod_i (n_i + 1)$ for all k and n.*

Proof. Use Corollary 3.1 and $\int_{I^d} x^n \, dx = \prod_i \begin{pmatrix} n_i \\ k_i \end{pmatrix} \prod_i (n_i+1)$. □

Remark. A criterion for uniform absolutely continuity of μ with respect to Lebesgue measure in the complex plane has been given in [11] in terms of the boundedness of a shift operator with respect to some discrete kernel.

There is also a characterization of Hausdorff of L^p measures on $I^1 = [0,1]$ for $p > 2$. This has an obvious generalization to d dimensions:

Proposition 3.3 *Given a bounded positive probability measure $\mu \in M(I^d)$ and assume $1 < p < \infty$. Then $\mu \in L^p(I^d)$ if and only if there exists a constant C such that for all k, n*

$$(n+1)^{p-1} \sum_{k=0}^{n} (\Delta^k(\mu)_n \begin{pmatrix} n \\ k \end{pmatrix})^p \leq C \; . \tag{3}$$

Proof. (i) Let $\mu^{(n)}$ be the measures of Corollary 2.3. We construct first from the atomic measures $\mu^{(n)}$ absolutely continuous measures $\tilde{\mu}^{(n)} = g^{(n)} dx$ on I^d given by a function g which takes the constant value $(|\Delta^k(\mu)_n| \begin{pmatrix} n \\ k \end{pmatrix})^p \prod_{i=1}^d (n_i+1)^p$ on a cube of side lengths $1/(n_i+1)$ centered at the point $(n-k)/n \in I^d$. Because the cube has Lebesgue volume $(n+1)^{-1} = \prod_{i=1}^d (n_i+1)^{-1}$, it has the same measure with respect to both $\tilde{\mu}^{(n)}$ and $g^{(n)}dx$.

We have therefore also $g^{(n)} dx \to \mu$ weakly.

(ii) Assume $\mu = f dx$ with $f \in L^p$. Because $g^{(n)} dx \to f dx$ in the weak topology for measures, we have $g^{(n)} \to f$ weakly in L^p. But then, there exists a constant C such that $||g^{(n)}||_p \leq C$ and this is equivalent to (3).

(iii) On the other hand, assumption (3) means that $||g^{(n)}||_p \leq C$, where $g^{(n)}$ was constructed in (i). Since the unit-ball in the reflexive Banach space $L^p(I^d)$ is weakly compact for $p \in (0,1)$, a subsequence of $g^{(n)}$ converges to a function $g \in L^p$. This implies that a subsequence of $g^{(n)} dx$ converges as a measure to $g dx$ which is in L^p and which is equal to μ by the uniqueness of the moment problem (Weierstrass). □

Remark. For more on the reconstruction of the measure in the L^2 case, see [15].

We end with some discussion.

An obvious question is to find a necessary and sufficient condition in terms of the moments, such that μ is L^p-absolutely continuous with respect to a second measure ν.

Analogue criterions for generalized moments might be simpler in some circumstances: for example, in one dimension, if x^n is replaced by the Chebychev polynomial $T_n(x)$ of the first kind, the moments $\int T_n(x) \, d\mu$ become the Fourier coefficients of a measure ν on $\mathbf{T}$. This measure ν is defined by $\nu(A) = \mu(\{x = (z + \overline{z})/2 \in [0,1] \mid z \in A\})$. We are not aware of a criterion in terms of Fourier coefficients which assure that μ is L^1-absolutely continuous with respect to ν. The problem is already subtle, if ν is the Lebesgue measure (see [8]). While it is by Plancherel's theorem easy to check for L^2-absolutely continuity with respect to Lebesgue measure, the decision weather μ is L^2-absolutely continuous with respect to a general second measure ν seems not to be simple also in terms of Fourier coefficients.
While some fractal properties of a measure can be deduced from the Fourier coefficients (see the references in the survey article [13]), it would be interesting to estimate from the moments μ_n the Hausdorff dimension of μ, which is the infimum of the Hausdorff dimensions of Borel sets S satisfying $\mu(I^d \setminus S) = 0$. For inverse problems for fractal measures obtained as attractors of linear iterated function systems in one dimension see [1]. Also interesting would be to determine from the moments, whether the potential theoretical α energy of μ is finite. One could translate corresponding results for Fourier coefficients (see [8]) into conditions for moments, but there might be simpler (only sufficient) conditions on the moments, which assure finite α-energy and so a lower bound α for the Hausdorff dimension using a result in [4].

References

[1] S. Abenda. Inverse problem for one-dimensional fractal measures via iterated function systems and the moment method. *Inverse problems*, 6:885–596, 1990.

[2] S. Abenda and G. Turchetti. Inverse problem for fractal sets on the real line via the moment problem. *Il Nuovo Cimento*, 104:213–227.

[3] N.I. Akhiezer. *The classical moment problem and some related questions in analysis*. University Mathematical Monographs. Hafner publishing company, New York, 1965.

[4] K.J. Falconer. *The geometry of fractal sets*. Cambridge University Press, Cambridge, 1985.

[5] B. Fuglede. The multidimensional moment problem. *Expo. Math.*, 1:47–65, 1983.

[6] F. Hausdorff. Summationsmethoden und Momentfolgen, I,II. *Math. Z.*, 9:74–109, 280–299, 1921.

[7] T.H. Hildebrandt and I.J. Schoenberg. On linear functional operations and the moment problem for a finite interval in one or several dimensions. *Annals of Mathematics*, 34:317–328, 1933.

[8] J-P. Kahane and R. Salem. *Ensembles parfaits et séries trigonométriques*. Hermann, 1963.

[9] O. Knill. A remark on quantum dynamics. *Helvetica Physica Acta*, 71:233–241, 1998.

[10] I.P. Natanson. *Constructive theory of functions*. Translation series. United states atomic energy commission, 1949. State Publishing House of Technical-Theoretical Literature.

[11] M. Putinar. A two-dimensional moment problem. *J. Func. Anal.*, 80:1–8, 1980.

[12] J.A. Shohat and J.D. Tamarkin. *The problem of moments*, volume 1 of *Mathematical Surveys*. ANS, New York, 1943.

[13] R.S. Strichartz. Self-similarity in harmonic analysis. *The journal of Fourier analysis and Applications*, 1:1–37, 1994.

[14] D.W. Stroock. *Probability theory, an analytic view*. Cambridge University Press, 1993.

[15] G.A. Viano. Solution of the Hausdorff moment problem by the use of Pollaczek polynomials. *J. Math. Anal. and Appl.*, pages 410–427, 1991.