SIAM J. OPTIMIZATION
Vol. 1, No. 2, pp. 191-205, May 1991

003

CONVERGENCE OF BEST ENTROPY ESTIMATES*

J. M. BORWEIN† AND A. S. LEWIS‡

Abstract. Given a finite number of moments of an unknown density $\bar{x}$ on a finite measure space, the best entropy estimate—that nonnegative density x with the given moments which minimizes the Boltzmann-Shannon entropy $I(x) := \int x \log x$—is considered. A direct proof is given that I has the Kadec property in L_1—if y_n converges weakly to $\bar{y}$ and $I(y_n)$ converges to $I(\bar{y})$, then y_n converges to $\bar{y}$ in norm. As a corollary, it is obtained that, as the number of given moments increases, the best entropy estimates converge in L_1 norm to the best entropy estimate of the limiting problem, which is simply $\bar{x}$ in the determined case. Furthermore, for classical moment problems on intervals with $\bar{x}$ strictly positive and sufficiently smooth, error bounds and uniform convergence are actually obtained.

Key words. moment problem, entropy, Kadec, partially finite program, normal convex integrand, duality

AMS(MOS) subject classifications. primary 41A46, 05C38; secondary 08A45, 28A20

1. Introduction. We shall suppose that (S, μ) is a finite measure space, and define the closed proper convex function $\phi : \mathbb{R} \to (-\infty, +\infty]$ by

$$\phi(u) := \begin{cases} u \log u, & \text{if } u > 0, \\ 0, & \text{if } u = 0, \\ +\infty, & \text{if } u < 0. \end{cases}$$

This function is a normal convex integrand [18], allowing us to define (minus) the Boltzmann-Shannon entropy $I_\phi(x) : L_1(S, \mu) \to (-\infty, +\infty]$ by

$$I_\phi(x) := \int_S \phi(x(s))\, d\mu(s). \tag{1}$$

Suppose $0 \leqq \bar{x} \in L_1(S, \mu)$ is an unknown density that we wish to estimate on the basis of a finite number of observed moments,

$$b_i = \int_S \bar{x}(s) a_i(s)\, d\mu(s), \qquad i = 1, \cdots, n,$$

where the a_i's are given functions in $L_\infty(S, \mu)$. This is a problem which commonly arises in diverse areas of physics, engineering, and statistics (see, for example, [14] and [11]). One popular technique is to choose the maximum entropy estimate—the solution of the optimization problem

$$(\mathrm{P}_n) \qquad \begin{cases} \text{minimize} & I_\phi(x) \\ \text{subject to} & \displaystyle\int_S x(s) a_i(s)\, d\mu(s) = b_i, \qquad i = 1, \cdots, n, \\ & 0 \leqq x \in L_1(S, \mu). \end{cases}$$

Attractive dual methods are available for solving the problems (P_n) computationally (see, for example, [7]).

* Received by the editors August 9, 1990; accepted for publication (in revised form) October 10, 1990. This research was partially supported by the Natural Sciences and Engineering Research Council of Canada.

† Department of Mathematics, Statistics, and Computing Science, Dalhousie University, Halifax, Nova Scotia, Canada B3H 3J5.

‡ Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1.

One measure of the effectiveness of this approach is the behavior of the optimal solution x_n of (P_n) as $n \to \infty$ (see, for example, [5] and [20]). At least when $\bar{x}$ is determined uniquely by the moment sequence $(\int \bar{x}a_i)_1^\infty$ we would hope that x_n converged to $\bar{x}$ in some sense. In [17] it was shown essentially that in this case $x_n\, d\mu$ converged to $\bar{x}\, d\mu$ weak-star as measures, while in [3] this was strengthened to the result that x_n converges weakly to $\bar{x}$ in L_1 (see also [10] and [15]). In this paper we will show that, in fact, x_n converges in L_1 norm to $\bar{x}$, and that with some further assumptions, the convergence is actually uniform.

The results break naturally into two parts. In the first we demonstrate the simple but remarkable geometric fact that, in common with the L_p norms $(1<p<\infty)$, the Boltzmann–Shannon entropy I_ϕ has the Kadec property: weak convergence of x_n to $\bar{x}$ and convergence of $I_\phi(x_n)$ to $I_\phi(\bar{x})$ implies norm convergence of x_n to $\bar{x}$. Our proof will be self-contained and straightforward. In the section which follows, we deduce the required convergence result and discuss some further implications.

In the second set of results, we begin by deriving a bound for the L_1 error in estimating $\bar{x}$ by x_n using duality techniques. Finally, when $\bar{x}$ is strictly positive and sufficiently smooth, we are able to combine this bound with ideas from the first collection of results and some standard approximation theory to show that, for classical algebraic and trigonometric moment problems on intervals, the best entropy estimates x_n converge **uniformly** to the underlying unknown density $\bar{x}$.

2. Strongly convex functions. In this section we derive the geometric property of the entropy which we will apply to the question of convergence.

LEMMA 2.1. *The Boltzmann–Shannon entropy I_ϕ defined in* (1) *is a proper, lower semicontinuous convex function, strictly convex on its domain,*

$$\operatorname{dom} I_\phi := \{x \in L_1(S, \mu) \mid I_\phi(x) < +\infty\},$$

and with weakly compact level sets, $\{x \in L_1(S, \mu) \mid I_\phi(x) \leqq \alpha\}$, for all $\alpha \in \mathbb{R}$.

Proof. (See [19].) The fact that the level sets are weakly compact follows either from the fact that the conjugate function $\phi^*(v) = e^{v-1}$ is everywhere finite, or directly from the Dunford–Pettis criterion [8]. □

The following inequality relating the so-called *I-divergence* of two probability densities with their L_1 distance appeared independently in [6], [12], and [13]. For completeness, we include a proof, following [12].

LEMMA 2.2 (a) *For* $0 < v \in \mathbb{R}$, $0 \leqq u \in \mathbb{R}$,

$$(u-v)^2 \leqq ((2u/3)+(4v/3))(u \log(u/v) - u + v).$$

(b) *For* $0 \leqq x, y \in L_1(S, d\mu)$ *with* $\int_S x(s)\, d\mu = \int_S y(s)\, d\mu = 1$,

$$\int_S x(s) \log(x(s)/y(s))\, d\mu \geqq \frac{1}{2}\left(\int_S |x(s)-y(s)|\, d\mu\right)^2.$$

Proof. (a) It is easy to check by differentiating that the function

$$\left(\frac{4}{3}+\frac{2t}{3}\right)(t \log t - t + 1) - (t-1)^2$$

is convex on $[0, +\infty)$, and attains a minimum value of 0 when $t = 1$. Putting $t := u/v$ gives (a).

(b) If $y(s) = 0$ and $x(s) > 0$ simultaneously on a nonnull set, then the left-hand side is $+\infty$ and there is nothing to prove. Therefore, assuming this does not occur, we can restrict attention to the case where $x(s)>0$ and $y(s)>0$ almost everywhere (if necessary, removing the set where $x(s)=y(s)=0$). Now set $u:=x(s)$ and $v:=y(s)$ in (a), take square-roots, and apply the Cauchy–Schwarz inequality to obtain (b), noting that $u\log(u/v)-u+v\geqq 0$, by (a). □

Notes. (1) We use the inequality (b) only in the case where $\log y\in L_\infty$, in which case there is no difficulty in defining the left-hand side in (b). If, however, we wish to be more precise, we define the left-hand side as $+\infty$ unless $x\geqq 0$; otherwise it is defined as

$$I_\phi(x)+\int_S \psi(y(s),s)\,d\mu(s),$$

where $\psi: R\times S\to(-\infty,+\infty]$ is the normal convex integrand

$$\psi(v,s):=\begin{cases}0, & \text{if } x(s)=0,\\ -x(s)\log v, & \text{if } x(s)>0, \quad v>0,\\ +\infty, & \text{if } x(s)>0, \quad v\leqq 0\end{cases}$$

(see [18]).

(2) It is easy to see from the proof that inequality (b) is strict unless $x=y$ almost everywhere.

(3) As observed in [12], the constant $\frac{1}{2}$ on the right in (b) is the best possible.

(4) We cannot hope for a similar inequality with $\|x-y\|_p^2$ replacing $\|x-y\|_1^2$ for some $p>1$ on the right. To see this, take $S=[0,1]$ with Lebesgue measure; define, for $k>1$,

$$x(s):=\begin{cases}k, & \text{for } s\in[0,1/k],\\ 0, & \text{for } s\in(1/k,1];\end{cases}$$

and let $y(s):=1$ almost everywhere. Then $\int x\log x/y=\log k$, while it is easy to see that $\|x-y\|_p^2\sim k^{2(1-1/p)}$.

Despite observation (4) above we can prove a somewhat similar result for the L_p norms.

PROPOSITION 2.3. *For $0\leqq x, y\in L_\infty(S,d\mu)$ with $\int_S x(s)\,d\mu=\int_S y(s)\,d\mu$, and $2\leqq p<+\infty$,*

$$\int_S x(s)\log(x(s)/y(s))\,d\mu\geqq(1/p(p-1))(\max\{\|x\|_\infty,\|y\|_\infty\})^{1-p}\|x-y\|_p^p.$$

Proof. Suppose $v\in(0,1]$. It is easy to check by differentiation that

$$\varphi(u):=u\log(u/v)-u+v-(1/p(p-1))|u-v|^p$$

is convex for $u\in[0,1]$ and attains a minimum value of 0 at $u=v$. Thus for $u\in[0,1]$, $v\in(0,1]$, we have

$$u\log(u/v)-u+v\geqq(1/p(p-1))|u-v|^p.$$

As before, we can restrict attention to the case where $y(s)>0$, almost everywhere. Now if we set $M:=\max\{\|x\|_\infty,\|y\|_\infty\}$, put $u:=(1/M)x(s)$ and $v:=(1/M)y(s)$ in the above, and integrate, we obtain the result. □

We are now ready to prove the geometric property of the Boltzmann–Shannon entropy I_ϕ that we will apply to the moment problem. In keeping with the terminology for normed spaces we make the following definition.

DEFINITION 2.4. Suppose X is a normed space with $f: X \to (-\infty, +\infty]$.

(a) (See [3].) The function f is **Kadec** if, whenever $y_n \to \bar{y}$ weakly in X and $f(y_n) \to f(\bar{y}) < +\infty$, it follows that $y_n \to \bar{y}$ in norm.

(b) The function f is **strongly convex** if it is Kadec and is a lower semicontinuous convex function, strictly convex on its domain $\{x \mid f(x) < +\infty\}$, with weakly compact level sets $\{x \mid f(x) \leqq \alpha\}$ for $\alpha \in \mathbb{R}$.

For example, with $X = L_p(S, \mu)$ for $1 < p < +\infty$, the norm $f(x) := \|x\|_p$ is strongly convex (see, for example, [8]). The main result of this section will be that the entropy I_ϕ is strongly convex on $L_1(S, \mu)$. By Lemma 2.1, it remains to show that I_ϕ is Kadec.

LEMMA 2.5. *Suppose $0 \leqq y_n$, $\bar{y} \in L_1(S, \mu)$, for $n = 1, 2, \cdots$, with, for each n, $\int y_n = \int \bar{y} > 0$. Suppose further that $y_n \to \bar{y}$ weakly in L_1 and that $I_\phi(y_n) \to I_\phi(\bar{y}) < +\infty$. Then $\|y_n - \bar{y}\|_1 \to 0$.*

Proof. By scaling the measure μ by a scalar factor, we lose no generality by supposing that $\int y_n = \int \bar{y} = 1$, for all n. For $m = 1, 2, \cdots$, we write $(x \wedge m)(s) := \min\{x(s), m\}$, and define

$$y^m := (1 - (1/m))\left[\left(\int \bar{y} \wedge m\right)^{-1} (\bar{y} \wedge m)\right] + (1/m)[\mu(S)^{-1}]. \tag{2}$$

Thus $\log y^m \in L_\infty$, and $\int y^m = 1$.

Now we have

$$\begin{aligned}
\frac{1}{2}\|y_n - y^m\|_1^2 &\leqq \int y_n \log(y_n/y^m), \quad \text{by Lemma 2.2,} \\
&= I_\phi(y_n) - \int y_n \log y^m \\
&\to I_\phi(\bar{y}) - \int \bar{y} \log y^m, \quad \text{as } n \to \infty, \text{ by assumption,} \\
&\leqq -\int \bar{y}\left\{(1 - (1/m)) \log\left[\left(\int \bar{y} \wedge m\right)^{-1} (\bar{y} \wedge m)\right] \cdots \right. \\
&\qquad \left. + (1/m) \log[\mu(S)^{-1}]\right\} + I_\phi(\bar{y}) \quad \text{(convexity and (2))} \\
&= (1 - (1/m))\left[\left\{\log\left(\int \bar{y} \wedge m\right)\right\}\left\{\int \bar{y}\right\} - \int \bar{y} \log(\bar{y} \wedge m)\right] \\
&\qquad + (1/m) \log[\mu(S)]\left\{\int \bar{y}\right\} + I_\phi(\bar{y}).
\end{aligned}$$

Now since $u \log u \geqq -1/e$ for any $u \geqq 0$, we have

$$-1/e \leqq \bar{y} \log(\bar{y} \wedge m) \uparrow \bar{y} \log \bar{y} \in L_1,$$

as $m \to \infty$, so we can apply the monotone convergence theorem to deduce that as $m \to \infty$, the right-hand side tends to 0. Thus we obtain

$$\lim_{m \to \infty} \overline{\lim_{n \to \infty}} \|y_n - y^m\|_1 = 0. \tag{3}$$

We also have

$$
\begin{aligned}
\|y^m-\bar y\|_1 &= \left\|(1-(1/m))\left[\left(\int \bar y\wedge m\right)^{-1}(\bar y\wedge m)\right]+(1/m)[\mu(S)^{-1}]-\bar y\right\|_1\\
&\leqq (1/m)+\left\|(1-(1/m))\left(\int \bar y\wedge m\right)^{-1}[(\bar y\wedge m)-\bar y]\right\|_1\\
&\quad+\left\|\left[(1-(1/m))\left(\int \bar y\wedge m\right)^{-1}-1\right]\bar y\right\|_1\\
&=(1/m)+(1-(1/m))\left(\int \bar y\wedge m\right)^{-1}\int[\bar y-(\bar y\wedge m)]\\
&\quad+\left[(1-(1/m))\left(\int \bar y\wedge m\right)^{-1}-1\right]\int \bar y\\
&\to 0,\quad \text{as } m\to\infty \text{ by monotone convergence.}
\end{aligned}
$$

Finally, by combining (3) and the above, we obtain

$$\overline{\lim_{n\to\infty}}\,\|y_n-\bar y\|_1\leqq \overline{\lim_{n\to\infty}}\,\|y_n-y^m\|_1+\|y^m-\bar y\|_1\to 0,\quad \text{as } m\to 0. \qquad \square$$

LEMMA 2.6. *Suppose X is a normed space, and $\alpha_n\to\bar\alpha$ in $\mathbb{R}$.*

(a) *If $w_n\to\bar w$ weakly in X, then $\alpha_n w_n\to\bar\alpha\bar w$ weakly.*

(b) *If $\|w_n-\bar w\|\to 0$, then $\|\alpha_n w_n-\bar\alpha\bar w\|\to 0$.*

Proof. The proof is elementary. $\square$

THEOREM 2.7. *The Boltzmann–Shannon entropy I_ϕ is strongly convex.*

Proof. By Lemma 2.1, we just have to show that I_ϕ is Kadec. To this end, suppose $z_n,\ \bar z\in L_1$, for $n=1,2,\cdots$; $z_n\to\bar z$ weakly in L_1; and $I_\phi(z_n)\to I_\phi(\bar z)<+\infty$. It follows that $I_\phi(z_n)<+\infty$, for all n sufficiently large, so $\bar z,\ z_n\geqq 0$.

Consider first the case where $\bar z\neq 0$, so $\int \bar z>0$. By weak convergence, $\int z_n\to\int\bar z$, so for all n sufficiently large, $\int z_n>0$, and we can define functions $\bar y:=(\int\bar z)^{-1}\bar z$ and $y_n:=(\int z_n)^{-1}z_n$. Thus $0\leqq y_n,\ \bar y\in L_1$, $\int y_n=\int\bar y$ for each n, and by Lemma 2.6, $y_n\to\bar y$ weakly. Furthermore,

$$
\begin{aligned}
I_\phi(y_n)&=\int\left\{\left(\int z_n\right)^{-1}z_n\log\left[\left(\int z_n\right)^{-1}z_n\right]\right\}\\
&=\left(\int z_n\right)^{-1}\left\{I_\phi(z_n)-\left(\log\int z_n\right)\int z_n\right\}\\
&\to\left(\int \bar z\right)^{-1}\left\{I_\phi(\bar z)-\left(\log\int \bar z\right)\int \bar z\right\}\\
&=I_\phi(\bar y).
\end{aligned}
$$

Thus Lemma 2.5 applies to show that $\|y_n-\bar y\|_1\to 0$, so by Lemma 2.6, $\|z_n-\bar z\|_1\to 0$.

Finally, suppose $\bar z=0$. Since $z_n\geqq 0$ for large n, we have, as $n\to\infty$, $\|z_n-\bar z\|_1=\int z_n\to\int\bar z=0$ by weak convergence. $\square$

If we know a priori that the sequence (z_n) in the above proof of the Kadec property is uniformly bounded, we obtain a much stronger conclusion.

THEOREM 2.8. *Suppose $0\leqq z_n\leqq M$ almost everywhere, $z_n\to\bar z$ weakly in L_1, and $I_\phi(z_n)\to I_\phi(\bar z)$. Then for any $p<+\infty$, $\|z_n-\bar z\|_p\to 0$ as $n\to\infty$.*

Proof. We first note that since the positive cone in L_1 is closed, and hence weakly closed, it follows from the assumptions that $0\leqq\bar z\leqq M$, almost everywhere.

The proof is now exactly analogous to Lemma 2.5 and Theorem 2.7, with minor changes. We can simplify the definition of y^m in (2) to

$$y^m := (1-(1/m))\bar{y} + (1/m)[\mu(S)]^{-1},$$

and we use Proposition 2.3 in place of Lemma 2.2. The only real change is the case $\bar{z}=0$ in Theorem 2.7. For large n we know $z_n \geqq 0$, and we may as well assume that $z_n \neq 0$. Now we can assert, by Proposition 2.3, for $p \geqq 2$, large n, and some $M_1 \geqq M_2$

$$\begin{aligned}(1/p(p-1))M_1^{1-p}\left\| z_n - \int z_n \right\|_p^p &\leqq \int z_n \log\left(z_n \Big/ \int z_n \right) \\ &= I_\phi(z_n) - \left(\int z_n\right)\log\left(\int z_n\right) \to 0, \quad \text{as } n\to\infty,\end{aligned}$$

from which it follows that $\|z_n\|_p \to 0$. □

3. L_1 convergence. In this section we will apply the strong convexity of the entropy I_ϕ to deduce, in particular, that if the unknown density $\bar{x}$ is uniquely determined by its moments, $(\int \bar{x}a_i)_{i=1}^\infty$, then the optimal solution x_n of (P_n) converges in L_1 norm to $\bar{x}$. The approach will be through the following elementary result, which may be found in [3].

THEOREM 3.1. *Let X be a topological space, with a nested sequence of closed subsets, $X \supset F_1 \supset F_2 \supset \cdots$, and suppose $f: X \to (-\infty, +\infty]$ has compact level sets. Consider the optimization problems*

$$(\mathrm{Q}_n) \qquad \inf\{f(x) \mid x \in F_n\}$$

$$(\mathrm{Q}_\infty) \qquad \inf\left\{f(x) \mid x \in \bigcap_{n=1}^\infty F_n\right\}.$$

The values of (Q_n) and (Q_∞) are attained, if finite, and the value of (Q_n) increases in n to the value of (Q_∞) (finite or infinite). Suppose furthermore that x_n is optimal for (Q_n), and x_∞ is the unique optimal solution for (Q_∞), with finite value. Then $x_n \to x_\infty$.

COROLLARY 3.2. *Let X be a normed space with a nested sequence of closed, convex subsets, $X \supset F_1 \supset F_2 \supset \cdots$, and suppose $f: X \to (-\infty, +\infty]$ is strongly convex. Suppose that (Q_∞) has finite value. Then (Q_n) and (Q_∞) have unique optimal solutions (with finite value), x_n and x_∞, respectively, and $x_n \to x_\infty$ in norm.*

Proof. Existence follows from Theorem 3.1, and uniqueness is a consequence of strict convexity. Theorem 3.1 shows that $x_n \to x_\infty$ weakly, and also $f(x_n) \to f(x_\infty)$, whence $x_n \to x_\infty$ in norm, by the Kadec property. □

We recall the problems (P_n) of § 1:

$$(\mathrm{P}_n) \qquad \begin{cases} \text{minimize} & I_\phi(x) \\ \text{subject to} & \displaystyle\int_S x(s)a_i(s)\,d\mu(s) = b_i, \qquad i=1,\cdots,n, \\ & 0 \leqq x \in L_1(S,\mu). \end{cases}$$

The limiting problem is

$$(\mathrm{P}_\infty) \qquad \begin{cases} \text{minimize} & I_\phi(x) \\ \text{subject to} & \displaystyle\int_S x(s)a_i(s)\,d\mu(s) = b_i, \qquad i=1,2,\cdots, \\ & 0 \leqq x \in L_1(S,\mu). \end{cases}$$

Applying Corollary 3.2 gives the following result.

COROLLARY 3.3. *The value of* (P_n) *increases in n to the value of* (P_∞) (*finite or infinite*). *If* (P_∞) *has finite value, then* (P_n) *and* (P_∞) *have unique optimal solutions* (*with finite value*), x_n *and* x_∞, *respectively, and* $\|x_n - x_\infty\|_1 \to 0$.

Proof. The proof is by Corollary 3.2. □

Notes. (1) Assuming, as in § 1, that the b_i's are the moments of an unknown density $0 \leqq \bar{x} \in L_1$, then if $I_\phi(\bar{x}) < +\infty$ it follows that (P_∞) has finite value.

(2) If, furthermore, S is a compact metric space with Borel measure μ, and the linear span of $\{a_i \mid i = 1, 2, \cdots\}$ is dense in the continuous functions $C(S)$ (as in the classical trigonometric and algebraic moment problems), then it is easily checked that $\bar{x}$ is uniquely determined by its moments $(\int \bar{x} a_i)$, and so $\bar{x} = x_\infty$. In this case $\|x_n - \bar{x}\|_1 \to 0$.

(3) Convergence in L_1 norm is the best possible: in general, we cannot expect convergence in L_p norm for any $p > 1$. To see this, suppose $0 \leqq \bar{x} \in L_1$ with $I_\phi(\bar{x}) < +\infty$, but that $\bar{x} \notin L_p$ for any $p > 1$. Such functions are not difficult to construct (see, for example, [21]). It is well known that, under mild assumptions (see Theorem 4.2), the unique optimal solution of (P_n) is of the form

$$x_n = e^{\sum_{i=1}^n \lambda_i a_i - 1}$$

for some $\lambda \in \mathbb{R}^n$, and so $x_n \in L_\infty$ for each n. If $\|x_n - \bar{x}\|_p \to 0$, it would follow that $\bar{x} \in L_p$, which is a contradiction for $p > 1$.

(4) On the other hand, if the x_n's are uniformly bounded, then we can apply Theorem 2.8 in place of the Kadec property to deduce that $\|x_n - \bar{x}\|_p \to 0$ as $n \to \infty$, for every $p < +\infty$. Unfortunately it is unclear how we might know uniform boundedness of $(x_n)_1^\infty$ a priori. We return to this question of stronger convergence in the next section.

In some estimation problems it is natural to suppose that the unknown density $\bar{x}$ is bounded above by some known constant $0 < K \in \mathbb{R}$ (see, for example, [7]). In this case it may be appropriate to modify the Boltzmann-Shannon entropy $I_\phi(x)$ to $I_\phi(x) + I_\phi(K - x)$, thereby incorporating this information. We then arrive at the following modified problems:

$$(\mathrm{P}_n^K) \qquad \begin{cases} \text{minimize} & I_\phi(x) + I_\phi(K - x) \\ \text{subject to} & \displaystyle\int_S x(s) a_i(s)\, d\mu(s) = b_i, \qquad i = 1, \cdots, n, \\ & x \in L_1(S, \mu), \end{cases}$$

and the limiting problem

$$(\mathrm{P}_\infty^K) \qquad \begin{cases} \text{minimize} & I_\phi(x) + I_\phi(K - x) \\ \text{subject to} & \displaystyle\int_S x(s) a_i(s)\, d\mu(s) = b_i, \qquad i = 1, 2, \cdots, \\ & x \in L_1(S, \mu). \end{cases}$$

The following proposition concerning strong convexity is useful in this context.

PROPOSITION 3.4. *Let X be a normed space with* $f, g : X \to (-\infty, +\infty]$. *Suppose f is strongly convex and g is convex, lower semicontinuous, and bounded below. Then* $f + g$ *is strongly convex.*

Proof. Clearly $f+g$ is lower semicontinuous and convex, since f and g are, and is strictly convex on its domain, since f is. Suppose $g \geqq M$. Then the level set

$$\{x | (f+g)(x) \leqq \alpha\} \subset \{x | f(x) \leqq \alpha - M\},$$

and is closed, so therefore it is weakly compact. Finally, the fact that $f+g$ is Kadec follows from Theorem 6.5 in [3]. □

From Corollary 3.2 we immediately deduce that if (P_∞^K) has finite value, then the unique optimal solution x'_n of (P_n^K) converges in L_1 norm to the unique optimal solution x'_∞ of (P_∞^K) (and corresponding comments to Notes 1 and 2 following Corollary 3.3 hold). However, we can prove a stronger result.

THEOREM 3.5. *The value of* (P_n^K) *increases in n to the value of* (P_∞^K) *(finite or infinite). If* (P_∞^K) *has finite value, then* (P_∞^K) *and* (P_n^K) *have unique optimal solutions (with finite value)* x'_∞ *and* x'_n*, respectively, and* $\|x'_n - x'_\infty\|_p \to 0$, *as* $n \to \infty$, *for every* $p < +\infty$.

Proof. Since $\phi(u) \geqq -1/e$ for all u, $I_\phi(K-x)$ is bounded below (and certainly is convex and lower semicontinuous). Therefore, by Proposition 3.4, $I_\phi(x) + I_\phi(K-x)$ is strongly convex, so we can apply Theorem 3.1 and Corollary 3.2 to deduce the first assertions and the fact that $\|x'_n - x'_\infty\|_1 \to 0$. Thus by lower semicontinuity, $\underline{\lim}_{n\to\infty} I_\phi(x'_n) \geqq I_\phi(x'_\infty)$. However, we also know that

$$\begin{aligned}\overline{\lim_{n\to\infty}}\, I_\phi(x'_n) &= \overline{\lim}\,(I_\phi(x'_\infty) + I_\phi(K - x'_\infty) - I_\phi(K - x'_n)) \\ &= I_\phi(x'_\infty) + I_\phi(K - x'_\infty) - \underline{\lim}_{n\to\infty}\, I_\phi(K - x'_n) \\ &\leqq I_\phi(x'_\infty) + I_\phi(K - x'_\infty) - I_\phi(K - x'_\infty) \\ &= I_\phi(x'_\infty),\end{aligned}$$

again by lower semicontinuity.

Thus $I_\phi(x'_n) \to I_\phi(x'_\infty)$, and Theorem 2.8 now gives the result. □

4. Error bounds and uniform convergence. In the last section we saw that the unique optimal solution x_n of the problem (P_n) converged in L_1 norm to the unique optimal solution of the limiting problem (P_∞) (which in the determined case is exactly the unknown density $\bar{x}$). In this section we will demonstrate how, in more special circumstances, we can provide bounds on the L_1 error between x_n and $\bar{x}$. In classical cases this in turn allows us to prove that when $\bar{x}$ is strictly positive and sufficiently smooth, x_n actually must converge uniformly to $\bar{x}$. This of course is the most desirable result in practice.

In order to accomplish this, we use a combination of ideas from the previous sections and results from classical approximation theory to investigate the relationship between (P_n) and its dual problem. We therefore begin by summarizing what is known in general about this duality (see, for example, [2]). Recall that the primal problem is

$$(\mathrm{P}_n) \qquad \begin{cases} \text{minimize} & I_\phi(x) \\ \text{subject to} & \displaystyle\int_S (x-\bar{x})a_i\,d\mu = 0, \qquad i = 1, \cdots, n, \\ & 0 \leqq x \in L_1(S, \mu). \end{cases}$$

The corresponding dual problem is then

$$(\mathrm{D}_n) \qquad \begin{cases} \text{maximize} & \displaystyle\int_S \left(\bar{x}\left[\sum_{i=1}^n \lambda_i a_i\right] - e^{[\sum_{i=1}^n \lambda_i a_i]-1}\right) d\mu \\ \text{subject to} & \lambda \in \mathbb{R}^n. \end{cases}$$

The following weak duality result is elementary.

PROPOSITION 4.1. *The value of* (P_n) *is greater than or equal to the value of* (D_n).

In order to claim equality between the values of the primal and dual problems, we need a constraint qualification:

(CQ) There exists an $\hat{x} \in L_1$, feasible for (P_n) with finite value, and with $\hat{x}(s) > 0$ a.e.

In practice, it is frequently the case that the constraint functions $\{a_1, \cdots, a_n\}$ are *pseudo-Haar* (in other words linearly independent on nonnull sets). For example, this is the case when the a_i's are linearly independent and analytic on a compact interval with Lebesgue measure (which covers the classical moment problems). In this case, providing that $\bar{x}$ is nonzero with finite value, (CQ) holds.

When the constraint qualification holds, we get a strong duality result.

THEOREM 4.2. *Suppose* (CQ) *holds. Then both* (P_n) *and* (D_n) *attain their values, which are equal. If* λ^n *is optimal for* (D_n), *then the unique optimal solution of* (P_n) *is*

$$x_n := e^{\sum_{i=1}^n \lambda_i^n a_i - 1}.$$

All these results may be found in [2].

We define the constant E_n associated with the problem (P_n) to measure how well it is possible to approximate $1 + \log \bar{x}$ uniformly with a linear combination of the a_i's, $i = 1, \cdots, n$.

DEFINITION 4.3. For each n, E_n is defined to be $+\infty$ unless $\log \bar{x} \in L_\infty$, in which case $E_n := \min\{\|\sum_{i=1}^n \lambda_i a_i - 1 - \log \bar{x}\|_\infty \mid \lambda \in \mathbb{R}^n\}$.

Using this constant we can now give a lower bound on the value of (D_n) (and therefore of (P_n)). We need the following lemma.

LEMMA 4.4. *Suppose* $\beta > 0$.

(a) *If* $|u| \leqq \beta$, *then* $1 + u \leqq e^u \leqq 1 + u + e^\beta \beta^2/2$.

(b) *If* $|v - w| \leqq \beta$, *then* $|e^v - e^w| \leqq \beta(1 + e^\beta \beta/2)e^w$.

Proof. (a) This part follows by Taylor's theorem, and convexity.

(b) $-\beta(1 + e^\beta\beta/2) < -\beta \leqq v - w \leqq e^{v-w} - 1 \leqq v - w + e^\beta \beta^2/2 \leqq \beta(1 + e^\beta \beta/2)$, by applying (a) twice. Thus $|e^{v-w} - 1| \leqq \beta(1 + e^\beta\beta/2)$, and the result now follows. □

THEOREM 4.5. *For every* n *we have*

$$I_\phi(\bar{x}) \geqq V(\mathrm{P}_n) \geqq V(\mathrm{D}_n) \geqq I_\phi(\bar{x}) - \frac{1}{2} e^{En} E_n^2 \int \bar{x},$$

where E_n *is given by Definition* 4.3 *and* $V(\cdot)$ *denotes value.*

Proof. The first inequality follows from the fact that $\bar{x}$ is always feasible for (P_n), while the second is Proposition 4.1. We need only check the last for $\log \bar{x} \in L_\infty$. Since span $\{a_1, \cdots, a_n\}$ is finite-dimensional, there exists $\bar{\lambda} \in \mathbb{R}^n$ attaining the minimum in Definition 4.3, so

$$\left| \sum_{i=1}^n \bar{\lambda}_i a_i - 1 - \log \bar{x} \right| \leqq E_n \quad \text{a.e.}$$

Applying Lemma 4.4(a) now gives

$$e^{\sum_{i=1}^n \bar{\lambda}_i a_i - 1 - \log \bar{x}} \leqq 1 + \sum_{i=1}^n \bar{\lambda}_i a_i - 1 - \log \bar{x} + \frac{1}{2} e^{E_n} E_n^2,$$

so multiplying by $\bar{x}$ (which is nonnegative) gives

$$e^{\sum_{i=1}^n \bar{\lambda}_i a_i - 1} - \bar{x} \sum_{i=1}^n \bar{\lambda}_i a_i \leqq -\bar{x} \log \bar{x} + \frac{1}{2} e^{E_n} E_n^2 \bar{x}.$$

Integrating now gives the result. □

The following assumption holds in most of the cases in which we are interested.

ASSUMPTION. $a_1 \equiv 1$.

PROPOSITION 4.6. *Suppose* $a_1 \equiv 1$. *Whenever* (P_n) *has finite value, denote its (unique) optimal solution by* x_n.

(a) *If* x^0 *is feasible for* (P_n) *with finite value, then* $\int x_n \log x_n \leqq \int x^0 \log x_n$. *In particular, if* (P_m) *has finite value, with* $m \geqq n$, *then we have that* $\int x_n \log x_n \leqq \int x_m \log x_n$, *and if* $I_\phi(\bar{x}) < +\infty$, *then it follows that* $\int x_n \log x_n \leqq \int \bar{x} \log x_n$.

(b) *Suppose* (CQ) *holds for* (P_n). *If* x^0 *is feasible for* (P_n), *then we have* $\int x_n \log x_n = \int x^0 \log x_n$. *In particular,* $\int x_n \log x_n = \int \bar{x} \log x_n$, *and if* (P_m) *has finite value, with* $m \geqq n$, *then* $\int x_n \log x_n = \int x_m \log x_n$.

Proof. Since ϕ is convex, it is easy to check that if $u > 0$,

$$(1/\nu)\{\phi(u+\nu w) - \phi(u)\} \downarrow (\log u + 1)w \tag{4}$$

as $\nu \downarrow 0$. Then since x_n is optimal for (P_n), we have

$$\begin{aligned} 0 &\leqq \lim_{\nu\downarrow 0} (1/\nu)\{I_\phi(x_n + \nu(x^0 - x_n)) - I_\phi(x_n)\} \\ &= \int (\log x_n + 1)(x^0 - x_n) = \int (x^0 - x_n) \log x_n, \end{aligned}$$

by the monotone convergence theorem (observing the fact that when $\nu = 1$, the integrand in the first inequality is integrable), providing $x_n > 0$ almost everywhere, which gives (a) in this case.

In view of Theorem 4.2, this is all we will use. However, in point of fact a more precise argument shows that $x_n(s) = 0$ implies $x^0(s) = 0$ almost everywhere (see [4]), allowing us to restrict the range of integration to $\{s \mid x_n(s) > 0\}$.

To see (b) we simply have to rewrite $\log x_n$ using the known form of the solution from Theorem 4.2. $\square$

By combining the above result with the weak duality bound in Theorem 4.5, and using the inequality in Lemma 2.2, we obtain a bound on the L_1 error of x_n from $\bar{x}$ in terms of the approximation error E_n of Definition 4.3. Ignoring the case $\bar{x} = 0$, we lose no generality (scaling if necessary) in assuming $\int \bar{x} = 1$.

THEOREM 4.7. *Suppose* $a_1 \equiv 1$, $\int \bar{x} = 1$, *and* $I_\phi(\bar{x}) < +\infty$. *Then the optimal solution* x_n *satisfies*

$$\|x_n - \bar{x}\|_1 \leqq E_n\, e^{E_n/2}, \tag{5}$$

where E_n *is given by Definition* 4.3.

Furthermore, if $\bar{x} \in L_\infty$ *and the sequence* $(x_n)_1^\infty$ *is uniformly bounded in* L_∞, *then given any* $2 \leqq p < +\infty$,

$$\|x_n - \bar{x}\|_p^p \leqq K_p E_n^2\, e^{E_n}, \tag{6}$$

where the constant K_p *is independent of* n.

Proof. By Proposition 4.6,

$$\int \bar{x} \log x_n \geqq \int x_n \log x_n = V(\mathrm{P}_n) \geqq \int \bar{x} \log \bar{x} - \frac{1}{2}\, e^{E_n} E_n^2$$

by Theorem 4.5. Applying Lemma 2.2(b) gives

$$\frac{1}{2}\|x_n - \bar{x}\|_1^2 \leqq \int \bar{x} \log(\bar{x}/x_n) \leqq \frac{1}{2} e^{E_n} E_n^2,$$

and hence the first result. The second part follows by using Proposition 2.3 in place of Lemma 2.2. $\square$

Thus we see that for large n, if we can approximate $\log \bar{x}$ uniformly with a certain error by linear combinations of the a_i's, $i=1,\cdots,n$, then x_n approximates $\bar{x}$ with error no worse (asymptotically) in the L_1 norm. In the last part of this section we shall see that when E_n is decreasing sufficiently quickly, as happens typically when $\bar{x}$ is sufficiently smooth, this actually forces uniform convergence of x_n to $\bar{x}$, due to the known form of x_n from Theorem 4.2.

We know that, by definition, $E_n \to 0$ exactly when $1+\log \bar{x}$ lies in the closed span of the a_i's in L_∞. It follows from (5) that $\bar{x}$ is therefore uniquely determined by this fact and its moment sequence.

COROLLARY 4.8. *Suppose* $a_1 \equiv 1$, $\log x^1$ *and* $\log x^2$ *lie in* $\operatorname{cl}\operatorname{span}\{a_1, a_2, \cdots\}$, *and* $\int (x^1 - x^2)a_i = 0$, *for* $i=1,2,\cdots$. *Then* $x^1 = x^2$.

Proof. Clearly $0 \neq x^1$, $x^2 \in \operatorname{dom} I_\phi$, and without loss of generality we may assume that $\int x^1 = \int x^2 = 1$. If we set $\bar{x} := x^1$ in (P_n), then the corresponding sequence of optimal solutions $x_n^1 \to x^1$ in L_1, by (5). Similarly, setting $\bar{x} := x^2$ shows that the optimal solutions $x_n^2 \to x^2$. However, since x^1 and x^2 have identical moments, $x_n^1 = x_n^2$, for each n, so $x^1 = x^2$. □

DEFINITION 4.9. For each $n=1,2,\cdots$,

$$\Delta_n := \max\{\|f\|_\infty / \|f\|_1 \mid 0 \neq f \in \operatorname{span}\{a_1,\cdots,a_n\}\}.$$

We assume that at least one a_i is not identically zero, whence it is clear from positive homogeneity and compactness that Δ_n is well defined for large n with $0 < \Delta_n < +\infty$, and Δ_n is nondecreasing in n.

LEMMA 4.10. *For any* $f \in \operatorname{span}\{a_1,\cdots,a_n\}$,

$$\|f\|_\infty \leqq -\log(1 - \Delta_n \|e^f - 1\|_1),$$

where we interpret $\log(u) = -\infty$ *if* $u \leqq 0$.

Proof. We first claim that for any $u \geqq -M$, with $M > 0$,

$$|e^u - 1| \geqq M^{-1}(1 - e^{-M})|u|. \tag{7}$$

To see this, note that for $u \geqq 0$, $e^u - 1 \geqq u$, and $e^{-M} \geqq 1 - M$ by convexity, so $e^u - 1 \geqq u \geqq M^{-1}(1-e^{-M})u$, as required. On the other hand, for $u \in [-M, 0]$, by convexity we have

$$\begin{aligned}\exp\{u\} &= \exp\{(-u/M)(-M) + (1-(-u/M))0\}\\ &\leqq (-u/M)\exp(-M) + (1-(-u/M))\exp(0)\\ &= (-u/M)\,e^{-M} + 1 + (u/M),\end{aligned}$$

so we obtain $1 - e^u \geqq -M^{-1}(1-e^{-M})u$, which gives (7).

Now to prove the lemma, we can suppose $f \neq 0$. Then from (7) we obtain, since $f \geqq -\|f\|_\infty$ almost everywhere, $|e^f - 1| \geqq \|f\|_\infty^{-1}(1 - e^{-\|f\|_\infty})|f|$ almost everywhere, so integrating shows

$$\|e^f - 1\|_1 \geqq \|f\|_\infty^{-1}(1-e^{-\|f\|_\infty})\|f\|_1 \geqq \Delta_n^{-1}(1 - e^{-\|f\|_\infty})$$

by Definition 4.9. The result now follows. □

LEMMA 4.11. *With* E_n *defined as in Definition* 4.3, *there exists a* p_n *in* $\operatorname{span}\{a_1,\cdots,a_n\}$, *satisfying* $\|p_n - 1 - \log \bar{x}\|_\infty = E_n$, *and hence*

$$\|e^{p_n - 1} - \bar{x}\|_1 \leqq E_n\left(1 + \frac{1}{2}E_n e^{E_n}\right)\int \bar{x}.$$

Proof. The first statement is just the definition of E_n. Now applying Lemma 4.4(b) gives $|e^{p_n-1} - \bar{x}| \leqq E_n(1 + \frac{1}{2}E_n e^{E_n})\bar{x}$ almost everywhere, and integrating gives the result. □

We are now ready to obtain an estimate for the uniform error of the optimal solution x_n from $\bar{x}$.

THEOREM 4.12. *Suppose* $a_1 \equiv 1$, $\int \bar{x} = 1$, *and* $\log \bar{x} \in L_\infty$ (*or equivalently, for some* $\delta > 0$, $\delta \leqq \bar{x} \leqq \delta^{-1}$ *almost everywhere*). *Then with* E_n *and* Δ_n *defined in Definitions* 4.3 *and* 4.9, *the unique optimal solution of* (P_n),

$$x_n := e^{\sum_{i=1}^n \lambda_i^n a_i - 1},$$

has the property that

$$\|\log x_n - \log \bar{x}\|_\infty \leqq E_n - \log\{1 - (\operatorname{ess\,inf} \bar{x})^{-1} \Delta_n E_n\, e^{E_n}(1 + e^{E_n/2} + (E_n/2)\, e^{E_n})\}.$$

Proof. Since $\bar{x} > 0$ almost everywhere, (CQ) is satisfied, so x_n has the form given by Theorem 4.2, and

$$\|x_n - \bar{x}\|_1 \leqq E_n\, e^{E_n/2}, \tag{8}$$

by Theorem 4.7.

By Lemma 4.11, there exists p_n in span $\{a_1, \cdots, a_n\}$ with

$$\|p_n - 1 - \log \bar{x}\|_\infty = E_n, \tag{9}$$

$$\|e^{p_n - 1} - \bar{x}\|_1 \leqq E_n(1 + \tfrac{1}{2} E_n\, e^{E_n}). \tag{10}$$

If we write $q_n := \sum_{i=1}^n \lambda_i^n a_i$, we obtain from (8) and (10),

$$\begin{aligned} E_n(1 + \tfrac{1}{2} E_n\, e^{E_n} + e^{E_n/2}) &\geqq \|e^{q_n - 1} - e^{p_n - 1}\|_1 \\ &= \int e^{p_n - 1} |e^{q_n - p_n} - 1| \\ &\geqq \|e^{q_n - p_n} - 1\|_1 \operatorname{ess\,inf}(e^{p_n - 1}) \\ &\geqq \|e^{q_n - p_n} - 1\|_1\, e^{-E_n} \operatorname{ess\,inf} \bar{x}, \end{aligned}$$

by (9). Thus we obtain

$$\|e^{q_n - p_n} - 1\|_1 \leqq E_n\, e^{E_n}(1 + \tfrac{1}{2} E_n\, e^{E_n} + e^{E_n/2})(\operatorname{ess\,inf} \bar{x})^{-1}.$$

We now apply Lemma 4.10, observing that $-\log(1 - \Delta_n u)$ is increasing in u:

$$\begin{aligned} \|q_n - p_n\|_\infty &\leqq -\log(1 - \Delta_n \|e^{q_n - p_n} - 1\|_1) \\ &\leqq -\log\{1 - (\operatorname{ess\,inf} \bar{x})^{-1} \Delta_n E_n\, e^{E_n}(1 + \tfrac{1}{2} E_n\, e^{E_n} + e^{E_n/2})\}. \end{aligned}$$

This, in combination with (9), gives the result. □

The error estimate in the above result shows that, providing, as n increases, that Δ_n is increasing at a slower rate than E_n is decreasing, then $\log x_n$ must converge to $\log \bar{x}$ in $\|\cdot\|_\infty$. More precisely, we have the following result.

COROLLARY 4.13. *Suppose* $a_1 \equiv 1$ *and* $\log \bar{x} \in L_\infty$. *Suppose further that* $\Delta_n E_n \to 0$ *as* $n \to \infty$, *where* E_n *and* Δ_n *are given by Definitions* 4.3 *and* 4.9, *respectively. Then the unique optimal solution of* (P_n), x_n, *converges in* $\|\cdot\|_\infty$ *to* $\bar{x}$ *as* $n \to \infty$. *In fact,* $\|x_n - \bar{x}\|_\infty = O(\Delta_n E_n)$, *as* $n \to \infty$.

Proof. Since $\bar{x} \neq 0$, we can, without loss of generality, scale μ so that $\int \bar{x}\, d\mu = 1$. Since Δ_n is nondecreasing in n, we have that $E_n \to 0$, so Theorem 4.12 shows that

$$\begin{aligned} \overline{\lim_{n\to\infty}} \|\log x_n - \log \bar{x}\|_\infty / \Delta_n E_n &\leqq \left(\lim_{n\to\infty} \Delta_n^{-1} \right) + 2(\operatorname{ess\,inf} \bar{x})^{-1} \\ &\leqq \mu(S) + 2(\operatorname{ess\,inf} \bar{x})^{-1}, \end{aligned}$$

since $\Delta_n \geqq \Delta_1 = \mu(S)^{-1}$. Thus for some constant k (independent of n), $|\log x_n - \log \bar{x}| \leqq k\Delta_n E_n$ almost everywhere for all n; so by Lemma 4.4(b),

$$|x_n - \bar{x}| \leqq (k\Delta_n E_n)(1 + e^{(k\Delta_n E_n)}(k\Delta_n E_n)/2)\bar{x} \quad \text{a.e.}$$

Thus $\overline{\lim}_{n\to\infty} \|x_n - \bar{x}\|_\infty / \Delta_n E_n \leqq k\|\bar{x}\|_\infty$, and the result follows. □

So we see from the above that if $E_n = O(n^{-\alpha})$ and $\Delta_n = O(n^{\beta})$, where $\beta < \alpha$, then x_n converges to $\bar{x}$ in $\|\cdot\|_\infty$ with error no larger than $O(n^{\beta-\alpha})$. In general, E_n (and hence α) will depend on the smoothness of $\bar{x}$, whereas Δ_n (and β) depends only on the constraint functions $\{a_1, \cdots, a_n\}$.

Once we know that $\|x_n - \bar{x}\|_\infty \to 0$, we can replace the use of (5) in Theorem 4.7 by (6). Following through the above argument, and replacing $\|\cdot\|_1$ by $\|\cdot\|_p$ where appropriate ($p \geqq 2$), gives the slightly refined estimate $\|x_n - \bar{x}\|_\infty = O(\Delta_{n,p} E_n^{2/p})$, where

$$\Delta_{n,p} := \max\{\|f\|_\infty / \|f\|_p \mid 0 \neq f \in \operatorname{span}\{a_1, \cdots, a_n\}\}. \tag{11}$$

In particular,

$$\|x_n - \bar{x}\|_\infty = O(\Delta_{n,2} E_n). \tag{12}$$

In the final section we consider two classical cases where explicit bounds are known for E_n and Δ_n. This allows us to show that for algebraic and trigonometric moment problems on intervals, if the underlying density $\bar{x}$ is sufficiently smooth and strictly positive, the estimates x_n converge uniformly to $\bar{x}$.

5. The classical algebraic and trigonometric moment problems. We begin by summarizing Corollary 4.13. We consider the problem

$$(\mathrm{P}_n) \qquad \begin{cases} \text{minimize} & I_\phi(x) \\ \text{subject to} & \displaystyle\int_S (x - \bar{x}) a_i \, d\mu = 0, \qquad i = 1, \cdots, n, \\ & 0 \leqq x \in L_1(S, \mu), \end{cases}$$

where we suppose $\log \bar{x} \in L_\infty$, and $a_1 \equiv 1$, and we denote the unique optimal solution by x_n. Then Corollary 4.13 states that $\|x_n - \bar{x}\|_\infty \to 0$, providing $E_n \Delta_n \to 0$, where E_n and Δ_n are given by Definitions 4.3 and 4.9, respectively:

$$E_n = \min\{\|f - \log \bar{x}\|_\infty \mid f \in \operatorname{span}\{a_1, \cdots, a_n\}\},$$

$$\Delta_n = \max\{\|f\|_\infty / \|f\|_1 \mid 0 \neq f \in \operatorname{span}\{a_1, \cdots, a_n\}\}.$$

We consider two special cases.

Algebraic moment problems. In this case, $S = [0, 1]$, μ is Lebesgue measure, and $a_i(s) = s^{i-1}$, for $i = 1, \cdots, n$.

THEOREM 5.1. *Suppose, for the algebraic moment problem, that $\bar{x}$ is twice continuously differentiable and strictly positive. Then $n^2 E_n \to 0$ as $n \to \infty$.*

Proof. Since $0 < \bar{x} \in C^2[0, 1]$, it follows that $\log \bar{x} \in C^2[0, 1]$, so by Jackson's theorem [9], for some constant k, $E_n \leqq (k/n^2)\omega((\log \bar{x})'', 1/n)$, where

$$\omega(g, \delta) := \sup\{|g(s) - g(t)| \mid |s - t| \leqq \delta,\ s, t \in [0, 1]\},$$

is the modulus of continuity. Since $\omega(g, 0+) = 0$ for continuous g, the result follows. □

THEOREM 5.2. *For the algebraic moment problem,*

$$n^2 \geqq \Delta_n \geqq \begin{cases} (n+1)^2/4, & n \text{ odd}, \\ n(n+2)/4, & n \text{ even}. \end{cases} \qquad \square$$

Proof. For the proof, see [1].

COROLLARY 5.3. *Suppose, for the algebraic moment problem, that $\bar{x}$ is twice continuously differentiable and strictly positive. Then the unique optimal solutions x_n converge uniformly to $\bar{x}$.*

Proof. For the proof, see Corollary 4.13 and Theorems 5.1 and 5.2. □

In fact, a rather more precise version of the above argument, using (12), shows that if $\bar{x}$ is k times continuously differentiable ($k \geqq 2$) and strictly positive, then $\|x_n - \bar{x}\|_\infty = o(n^{1-k})$: the relevant Jackson theorem states that in this case $E_n = o(n^{-k})$, while it is shown in [1] that $\Delta_{n,2} = n$. We also see in this case, from Theorem 4.7, that for any $2 \leqq p < +\infty$, $\|x_n - \bar{x}\|_p = o(n^{-2k/p})$. In particular, $\|x_n - \bar{x}\|_2 = o(n^{-k})$.

In general, the smoother $\bar{x}$ is, the more rapidly E_n tends to zero. If $\bar{x}$ is analytic on [0, 1], or in other words has an analytic extension to an open subset of the complex plane containing [0, 1], then $E_n \to 0$ linearly: $E_n = O(\rho^n)$ for some constant $0 \leqq \rho < 1$. If, in fact, $\bar{x}$ is an entire function, the convergence is superlinear (see [16]). It follows from Corollary 4.13 that for analytic, strictly positive $\bar{x}$ in the algebraic moment problem, $\|x_n - \bar{x}\|_\infty \to 0$ linearly with the same convergence ratio as E_n; and if $\bar{x}$ is entire, the convergence is superlinear.

Trigonometric moment problems. In this case, $S = [-\pi, \pi]$, $2\pi\mu$ is Lebesgue measure, and for $j = 1, 2, \cdots$, $a_{2j}(s) = \cos(js)$ and $a_{2j+1}(s) = \sin(js)$.

THEOREM 5.4. *Suppose, for the trigonometric moment problem, that $\bar{x}$ is strictly positive with both $\bar{x}$ and $\bar{x}'$ continuous and 2π-periodic. Then, $nE_{2n+1} \to 0$ as $n \to \infty$.*

Proof. By [9], $E_{2n+1} \leqq (578/n)\omega((\log \bar{x})', 1/n)$, where again $\omega(\cdot, \cdot)$ is the modulus of continuity; so the result follows. □

THEOREM 5.5. *For the trigonometric moment problem, $2n+1 \geqq \Delta_{2n+1} \geqq n$.*

Proof. For the proof, see [22]. □

COROLLARY 5.6. *Suppose, for the trigonometric moment problem, that $\bar{x}$ is strictly positive with both $\bar{x}$ and $\bar{x}'$ continuous and 2π-periodic. Then the unique optimal solutions x_n converge uniformly to $\bar{x}$.*

Proof. Theorems 5.4 and 5.5 show that

$$\Delta_{2n+1}E_{2n+1} \leqq (2n+1)E_{2n+1} \to 0,$$
$$\Delta_{2n+2}E_{2n+2} \leqq \Delta_{2n+3}E_{2n+1} \leqq (2n+3)E_{2n+1} \to 0.$$

Thus $\Delta_n E_n \to 0$, so the result follows by Corollary 4.13. □

In fact, just as in the algebraic case, a more precise version of the above argument (using the fact that $\Delta_{2n+1,2} = (2n+1)^{1/2}$ in this case [22]) shows that if $\bar{x}, \bar{x}', \cdots, \bar{x}^{(k)}$ are continuous and 2π-periodic, with $\bar{x}$ strictly positive, then $\|x_n - \bar{x}\|_\infty = o(n^{(1/2)-k})$. Furthermore, Theorem 4.7 shows that for any $2 \leqq p < +\infty$, $\|x_n - \bar{x}\|_p = o(n^{-2k/p})$. In particular, $\|x_n - \bar{x}\|_2 = o(n^{-k})$.

Our approach can be extended to prove similar results for multidimensional algebraic and trigonometric moment problems. Thus one can consider polynomials with maximum degree or sum of degrees not exceeding n, etc., on various domains. This becomes considerably more technical and we choose not to take the matter further herein.

Note added in proof. Error bounds for the trigonometric case under certain conditions on $\bar{x}$ (and numerical results) may be found in [23], and bounds for problems involving some entropies other than the Boltzmann-Shannon entropy appear in [24].

REFERENCES

[1] D. AMIR AND Z. ZIEGLER, *Polynomials of extremal L_p-norm on the L_∞-unit sphere*, J. Approx. Theory, 18 (1976), pp. 86–98.

[2] J. M. BORWEIN AND A. S. LEWIS, *Duality relationships for entropy-like minimization problems*, SIAM J. Control Optim., 29 (1991), pp. 325–338.

[3] ———, *On the convergence of moment problems*, Trans. Amer. Math. Soc., 1991, to appear.

[4] J. M. BORWEIN, A. S. LEWIS, AND R. NUSSBAUM, *Entropy minimization*, DAD *problems and doubly-stochastic kernels*, to appear.

[5] W. BRITTON, *Conjugate duality and the exponential Fourier spectrum*, Lecture Notes in Statistics 18, Springer-Verlag, New York, 1983.

[6] I. CSISZÁR, *Information-type measures of difference of probability distributions and indirect observations*, Studia Sci. Math. Hungar., 2 (1967), pp. 299–318.

[7] A. DECARREAU, D. HILHORST, C. LEMARÉCHAL, AND J. NAVAZA, *Dual methods in entropy maximization: Application to some problems in crystallography*, SIAM J. Optimization, submitted.

[8] J. DIESTEL, *Sequences and Series in Banach Spaces*, Springer-Verlag, New York, 1984.

[9] R. P. FEINERMAN AND D. J. NEWMAN, *Polynomial Approximation*, Williams and Wilkins, Baltimore, MD, 1974.

[10] B. FORTE, W. HUGHES, AND Z. PALES, Maximum entropy estimators and the problem of moments, Rend. Mat. Ser. VII, 9 (1989), pp. 689–699.

[11] S. M. KAY AND S. L. MARPLE, *Spectrum analysis—a modern perspective*. Proc. IEE-F, 69 (1981), pp. 1380–1419.

[12] J. H. B. KEMPERMAN, *On the optimum rate of transmitting information*, in Probability and Information Theory, Proceedings of an International Symposium, McMaster University, Hamilton, Ontario, pp. 126–169; Lecture Notes in Mathematics 89, Springer-Verlag, Berlin, 1969.

[13] S. KULLBACK, A lower bound for discrimination information in terms of variation, IEEE Trans. Inform. Theory, IT-13 (1967), pp. 126–127.

[14] S. W. LANG AND J. H. MCCLELLAN, *Spectral estimation for sensor arrays*, IEEE Trans. Acoust. Speech Signal Process., ASSP-31 (1983), pp. 349–358.

[15] A. S. LEWIS, *The convergence of entropic estimates for moment problems*, in Workshop/Miniconference on Functional Analysis/Optimization, S. Fitzpatrick and J. Giles, eds., Centre for Mathematical Analysis, Australian National University, Canberra, Australia, 1988, pp. 100–115.

[16] G. G. LORENTZ, *Approximation of Functions*, Second Edition, Chelsea, New York, 1986.

[17] L. R. MEAD AND N. PAPANICOLAOU, *Maximum entropy in the problem of moments*, J. Math. Phys., 25 (1984), pp. 2404–2417.

[18] R. T. ROCKAFELLAR, *Integrals which are convex functionals*, Pacific J. Math., 24 (1968), pp. 525–539.

[19] ———, *Integrals which are convex functionals*, II, Pacific J. Math., 39 (1971), pp. 439–469.

[20] J. SKILLING AND S. F. GULL, *The entropy of an image*, SIAM-AMS Proc., Applied Mathematics, 14 (1984), pp. 167–189.

[21] K. R. STROMBERG, *An Introduction to Classical Real Analysis*, Wadsworth, Belmont, CA, 1981.

[22] Z. ZIEGLER, *Minimizing the $L_{p,\infty}$-distortion of trigonometric polynomials.*, J. Math. Anal. Appl., 61 (1977), pp. 426–431.

[23] E. GASSIAT, *Problème sommatoire par maximum d'entropie*, C. R. Acad. Sci. Paris Sér. I, 303 (1986), pp. 675–680.

[24] D. DACUNHA-CASTELLE AND F. GAMBOA, *Maximum d'entropie et problème des moments*, Ann. Inst. H. Poincaré Probab. Statist., 26 (1990), pp. 567–596.