

## ON THE CONVERGENCE OF MOMENT PROBLEMS

J. M. BORWEIN AND A. S. LEWIS

**ABSTRACT.** We study the problem of estimating a nonnegative density, given a finite number of moments. Such problems arise in numerous practical applications. As the number of moments increases, the estimates will always converge weak\* as measures, but need not converge weakly in  $L_1$ . This is related to the existence of functions on a compact metric space which are not essentially Riemann integrable (in some suitable sense). We characterize the type of weak convergence we can expect in terms of Riemann integrability, and in some cases give error bounds. When the estimates are chosen to minimize an objective function with weakly compact level sets (such as the Boltzmann-Shannon entropy) they will converge weakly in  $L_1$ . When an  $L_p$  norm ( $1 < p < \infty$ ) is used as the objective, the estimates actually converge in norm. These results provide theoretical support to the growing popularity of such methods in practice.

### 0. INTRODUCTION

A very common problem in physics and engineering is known under the general title of "the moment problem". We are concerned with estimating some nonnegative measure, which may for example represent the power spectral density of some signal, or the density of states in some harmonic solid, or some other unknown physical property. Corresponding to some finite number of observations or calculations, we are given a set of "moments"—the integrals of various given functions with respect to the measure. Since these moments will not determine the measure uniquely, how should we best estimate the measure, and how good will our estimate be?

Such problems arise in a wide variety of settings. Moment problems occur frequently in spectral estimation, and in particular in speech processing, geophysics, radio astronomy, sonar and radar, and many other areas (see, for example, [Lang and McClellan, 1983; Kay and Marple, 1981] and the references therein). They are also common in numerous models from theoretical physics such as quantum spin systems, Ising models, and the summation of

---

Received by the editors April 20, 1989.

1980 *Mathematics Subject Classification* (1985 Revision). Primary 41A46, 42A70; Secondary 90C25, 49D39.

*Key words and phrases.* Moment problem, entropy, semi-infinite program, duality, Riemann integral, compact level sets, uniformly convex.

Research partially supported by the Natural Sciences and Engineering Research Council of Canada.

divergent series arising from Stieltjes transforms (see, for instance, [Mead and Papanicolaou, 1984] and the references therein). In applications moment problems are frequently also known as “underdetermined inverse problems.” The moment problem in pure mathematics dates back at least to [Hausdorff, 1921]. A recent survey of the wide range of approaches to the moment problem, and applications, is [Landau, 1987].

Given a finite number of moments, various approaches are possible in estimating the measure. The most popular traditional schemes are known collectively as “Padé approximation” (see, for instance, [Baker and Graves-Morris, 1980]). The idea here is to approximate the measure by atomic measures. In recent years, however, various authors have considered a rival approach where the measure is estimated by means of an objective function. Various entropy-like objectives have been tried (see, for example, [Burg, 1975; Ben-Tal, Borwein, and Teboulle, 1988(b); Borwein and Lewis, 1988(b)]), and advantages over more traditional approaches have been observed [Mead and Papanicolaou, 1984]. Norms have also been tried as objective functions [Goodrich and Steinhart, 1986].

A very important question arises in connection with this optimization approach to finite moment problems. How will our estimate converge to the underlying measure as the number of given moments increases? This question has been considered in the very special case of the Hausdorff moment problem (where the measures lie on  $[0, 1]$  and the moments are taken with respect to  $s^i$ ) in [Mead and Papanicolaou, 1984], where it was observed that, independent of the objective function used, the estimates converge weak\* as measures, and in [Forte, Hughes, and Pales, 1988], where some convergence results were proved when (minus) the Boltzmann-Shannon entropy ( $s \log s$ ) is used as the objective.

In this paper a rather more general approach is taken to the question of convergence. We begin by studying general sequences of optimization problems. It transpires that we can guarantee weak convergence if the level sets of the objective function are weakly compact. We therefore summarize some appropriate results of Rockafellar concerning the weak compactness of the level sets of various convex integral functionals on  $L_p$  spaces, and prove some related results. In particular these results apply to the Boltzmann-Shannon entropy.

Such weak convergence results are important because in general a sequence of feasible densities (in  $L_1$ ) for the finite moment problems need not converge weakly (in  $L_1$ ) to the underlying density. Using a duality theorem for semi-infinite linear programs we are able to give an elegant and surprising characterization of those integrands with respect to which the estimates converge weakly: loosely speaking, they are those functions which have good one-sided  $L_1$ -approximations by continuous functions. This property may be taken as a definition of Riemann integrability in a compact metric space. It agrees with the standard definition in the classical case, and as in that case is equivalent to almost everywhere continuity. The failure of estimates to converge weakly

in  $L_1$  arises because of the existence of functions which are not essentially Riemann integrable.

Despite this, when the underlying space is a subset of  $\mathbb{R}$ , estimates will converge in distribution, and when the integrand is continuous we can relate its degree of smoothness to error bounds for the rate of weak convergence.

The above observations indicate that the choice of objective function is rather important. In particular, using (minus) the Boltzmann-Shannon entropy ensures weak convergence in  $L_1$ . However, for certain objective functions we obtain even better convergence properties—convergence in norm. This happens in particular when an  $L_p$  norm ( $1 < p < \infty$ ) is used as the objective. Practical comparisons between this approach (using various objectives) and more traditional techniques would be of great interest, now that the optimization approach has been given a firm mathematical basis.

### 1. CONVERGENCE IN MOMENT PROBLEMS

We are interested in problems of the form

$$(MP_n) \quad \begin{cases} \text{inf} & f(x) \\ \text{subject to} & a_i(x) = b_i, \quad i = 1, \dots, n, \quad 0 \leq x \in X, \end{cases}$$

and the limiting problem (with infinitely many constraints)  $(MP_\infty)$ . Here,  $X$  is a partially-ordered, locally convex topological vector space,  $f: X \rightarrow (-\infty, +\infty]$  is a closed, proper, convex function, and the  $a_i$ 's are continuous linear functionals. In what sense can we expect the solutions of  $(MP_n)$  to converge to a solution of  $(MP_\infty)$ ?

Suppose  $S$  is a compact Hausdorff space and  $X = M(S)$ , the regular Borel measures, with the usual ordering and the weak\* topology (regarding  $M(S)$  as the dual of  $C(S)$ ). Suppose further that the  $a_i$ 's are densely spanning in  $C(S)$ . As we shall see later, if  $(MP_\infty)$  is consistent then it has a unique feasible solution  $\bar{\mu}$ , say, and if  $\mu_n$  is feasible for  $(MP_n)$  then  $\mu_n \rightarrow \bar{\mu}$  weak\*.

Now suppose that  $0 \leq \lambda \in M(S)$  and repose the problems in  $X = L_1(S, \lambda)$ , with the usual ordering and the weak topology, regarding the  $a_i$ 's as functions in  $L_\infty(S, \lambda)$ . It is easy to check that for  $x \in L_1(S, \lambda)$ ,  $x d\lambda \in M(S)$ . As before, if  $(MP_\infty)$  is consistent then it has a unique feasible solution,  $\bar{x}$  say (where  $\bar{x} = d\bar{\mu}/d\lambda$ ). However, we will show the remarkable fact that, under very mild assumptions, there will always exist a sequence of  $x_n$ 's feasible for  $(MP_n)$  such that  $x_n \not\rightarrow \bar{x}$  weakly in  $L_1(S, \lambda)$ . For this sequence we therefore have the pathological behaviour that

$$\forall y \in C(S), \quad \int_S (x_n - \bar{x})y d\lambda \rightarrow 0,$$

and yet

$$\exists y \in L_\infty(S, \lambda) \text{ s.t. } \int_S (x_n - \bar{x})y d\lambda \not\rightarrow 0.$$

The conclusion is that in order to ensure convergence, the objective function  $f$  must be carefully chosen.

We shall begin by considering a sequence of optimization problems posed in a topological space. With very little restriction on the underlying structure of the problems, we can give conditions ensuring that the optimal values of the problems converge to the optimal value of the limit problem, and furthermore that a sequence of optimal solutions converges to an optimal solution of the limit problem.

We shall consider a topological space  $X$ , with a nested sequence of closed subsets,  $X \supset F_1 \supset F_2 \supset \dots$ . The following simple result will be useful.

**Lemma 1.1.** *Suppose  $x_n \in F_n$  for each  $n$ , and  $\bar{x}$  is a cluster point of the sequence  $(x_n)$ . Then  $\bar{x} \in \bigcap_{n=1}^{\infty} F_n$ .*

*Proof.* Suppose  $\bar{x} \notin F_m$  for some  $m$ . Since  $F_m$  is closed,  $F_m^c$  is a neighbourhood of  $\bar{x}$ , so for some  $n \geq m$ ,  $x_n \in F_m^c$ . But as  $n \geq m$ ,  $x_n \in F_m$ , which is a contradiction.  $\square$

As usual in optimization, to ensure attainment we require some degree of compactness.

**Definition 1.2** (e.g. [Kelley, 1955]). A set  $C \subset X$  is *countably compact* if every sequence in  $C$  has a cluster point in  $C$ .

This property is weaker than either compactness or sequential compactness. It is easy to see that a closed subset of a countably compact set is countably compact.

We shall consider functions  $f: X \rightarrow (-\infty, +\infty]$ . The (lower) level sets of  $f$  are the sets  $\{x \in X | f(x) \leq \alpha\}$  for  $\alpha < +\infty$ . Thus  $f$  is lower semicontinuous if and only if it has closed level sets.

**Proposition 1.3.** *Suppose  $f: X \rightarrow (-\infty, \infty]$  has closed, countably compact level sets. If  $\inf f$  is finite, it is attained.*

*Proof.* Let  $\alpha := \inf f$  be finite, and define  $L_n := \{x | f(x) \leq \alpha + 1/n\}$ . For each  $n$ , pick  $x_n \in L_n$ . The sequence  $(x_n) \subset L_1$  and  $L_1$  is countably compact, so there exists a cluster point  $\bar{x}$ . Applying Lemma 1.1,  $\bar{x} \in \bigcap_{n=1}^{\infty} L_n$ , so  $f(\bar{x}) \leq \alpha$ . Thus  $\bar{x}$  attains the infimum.  $\square$

We now consider the sequence of problems

$$(P_n) \quad \inf\{f(x) | x \in F_n\},$$

and the limiting problem

$$(P_\infty) \quad \inf \left\{ f(x) | x \in \bigcap_{n=1}^{\infty} F_n \right\}.$$

We denote the value of a problem (P) by  $V(P)$ . The previous result shows that if  $f$  has closed, countably compact level sets and the values of  $(P_n)$  and  $(P_\infty)$  are finite, then they are attained.

**Proposition 1.4.** *Suppose  $f$  has closed, countably compact level sets. Then  $V(P_n) \uparrow V(P_\infty)$  (finite or infinite).*

*Proof.* Clearly  $V(P_n)$  is nondecreasing in  $n$ , and bounded above by  $V(P_\infty)$ . Suppose therefore that  $V(P_n) \leq \gamma < +\infty$  for all  $n$ . By Proposition 1.3, we can pick for each  $n$ ,  $x_n \in L \cap F_n$ , where  $L := \{x | f(x) \leq \gamma\}$  is countably compact. Thus the sequence  $(x_n)$  has a cluster point  $\bar{x} \in L$ , and by Lemma 1.1,  $\bar{x} \in \bigcap_{n=1}^\infty F_n$ . Thus  $V(P_\infty) \leq \gamma$ , and the result follows.  $\square$

**Proposition 1.5.** *Suppose  $f$  has closed, countably compact level sets. Suppose further that  $x_n$  is optimal for  $(P_n)$ , and that  $x_\infty$  is the unique optimal solution of  $(P_\infty)$ , with  $f(x_\infty) < +\infty$ . Then  $x_n \rightarrow x_\infty$ .*

*Proof.* Suppose  $x_n \not\rightarrow x_\infty$ , so there exists an open neighbourhood  $N$  of  $x_\infty$ , and a subsequence  $(x_{n_i}) \subset N^C$ . Denoting the set  $\{x | f(x) \leq f(x_\infty)\}$  by  $L$ ,  $(x_{n_i}) \subset L \cap N^C$ , so by countable compactness there exists a cluster point  $\bar{x}$  of  $(x_{n_i})$  (and therefore of  $(x_n)$ ), with  $\bar{x} \in L \cap N^C$ . By Lemma 1.1,  $\bar{x} \in \bigcap_{n=1}^\infty F_n$ . It follows by uniqueness that  $\bar{x} = x_\infty$ , but this contradicts  $\bar{x} \in N^C$ .  $\square$

We are specifically interested in the case where  $X$  is a topological vector space and the sets  $F_n$  are given by  $F_n = \{x \in C | a_i(x) = b_i, i = 1, \dots, n\}$ , where  $C$  is a closed set (typically a positive cone) and the  $a_i$ 's are continuous linear functionals. In particular, we are concerned with the cases  $X = L_p[\alpha, \beta]$ ,  $1 \leq p < \infty$  (with the weak topology), and  $L_\infty[\alpha, \beta]$  and  $M[\alpha, \beta]$  (with the weak\* topology), where the  $a_i$ 's are determined by  $a_i(t) = t^{i-1}$ , or  $a_i(t) = e^{i\sqrt{-1}t}$ , and  $C = \{x \geq 0\}$ . Whether the set  $F_n$  is nonempty in these cases can be answered in quite a straightforward way (see, for example, [Karlin and Studden, 1966]). The question of when the limiting problem is feasible is a classical moment problem in these cases (see, for example, [Widder, 1941]).

Proposition 1.3 gives one condition for attainment. The question of attainment in the problems  $(P_n)$  when we do not necessarily know that  $f$  has compact level sets, is considered in [Borwein and Lewis, 1988(b)].

Assuming that  $V(P_\infty) < +\infty$  and is attained, there are two natural ways to ensure that the optimum is unique. This will be the case either if  $f$  is strictly convex on its domain  $\{x | f(x) < +\infty\}$ , or if  $\bigcap_{n=1}^\infty F_n$  is a singleton. The following simple result is useful (see [Schaefer, 1971] for the notation).

**Proposition 1.6.** *Let  $(X, Y)$  be paired vector spaces,  $\bar{x} \in X$ , and suppose the span of  $(a_i)_1^\infty$  is  $\sigma(Y, X)$ -dense in  $Y$ . Then*

$$\{x \in X | \langle x, a_i \rangle = \langle \bar{x}, a_i \rangle, i = 1, 2, \dots\} = \{\bar{x}\}.$$

*Proof.* If  $\langle x - \bar{x}, a_i \rangle = 0$  for each  $i$ , then

$$x - \bar{x} \in \{(a_i)_1^\infty\}^\perp = \{\sigma(Y, X) - \text{cl span}(a_i)_1^\infty\}^\perp = Y^\perp = \{0\}. \quad \square$$

For example, suppose  $S \subset \mathbb{R}^k$  is compact and  $\lambda$  is Lebesgue measure. Then  $C(S)$  is  $\|\cdot\|_q$ -dense in  $L_q(S, \lambda)$  for  $q < +\infty$  (see [Rudin, 1966, Theorem

3.14]). If the sequence  $(a_i)_1^\infty \subset C(S)$  spans the (multidimensional) polynomials, then, by the Stone-Weierstrass Theorem [Holmes, 1975, p. 209], the span of  $(a_i)_1^\infty$  is  $\|\cdot\|_\infty$ -dense in  $C(S)$ , and therefore  $\|\cdot\|_q$ -dense in  $L_q(S, \lambda)$ , and so  $\sigma(L_q, L_p)$ -dense. Obviously the span of  $(a_i)_1^\infty$  will also be  $\sigma(C(S), M(S))$ -dense in  $C(S)$ . On the other hand,  $C(S)$  is weak\*-dense in  $L_\infty(S)$  [Holmes, 1975, p. 68], so  $\text{span}(a_i)_1^\infty$  will be  $\sigma(L_\infty, L_1)$ -dense in  $L_\infty(S)$ .

For example, if the sequence  $(a_i(s_1, \dots, s_k))_{i=1}^\infty$  is given by  $1, s_1, s_2, \dots, s_k, s_1^2, s_1s_2, \dots, s_1s_k, s_2^2, \dots$  etc., then the above result will apply. In particular, it applies in the one-dimensional case where  $a_i(s) := s^{i-1}$  for each  $i$ . Similarly, denoting the unit circle by  $\mathbf{T}$  (so  $C(\mathbf{T})$  is the space of  $2\pi$ -periodic real-valued continuous functions on  $\mathbb{R}$ , and  $L_p(\mathbf{T})$  is defined in the usual way for real-valued functions), the trigonometric polynomials are dense in  $C(\mathbf{T})$  [Rudin, 1966, 4.25]. Thus if  $(X, Y) = (L_p(\mathbf{T}), L_q(\mathbf{T}))$  with  $1 \leq p \leq \infty$  or  $(C(\mathbf{T}), M(\mathbf{T}))$ , and  $(a_i)_1^\infty = \{1, \cos s, \sin s, \dots\}$  then Proposition 1.6 will apply (and analogous results will hold in several dimensions).

Throughout this paper we shall generally assume that the underlying space  $S$  is a compact metric space. This involves no essential loss of generality since the existence of a densely spanning sequence  $(a_i)_1^\infty$  in  $C(S)$  implies that  $C(S)$  is separable, and hence that  $S$  is metrizable [Jameson, 1974, 26.14].

A function  $g: S \rightarrow \mathbb{R}$  is Borel if  $g^{-1}(V)$  is a Borel set for every open set  $V \subset \mathbb{R}$ .

**Theorem 1.7.** *Let  $S$  be a compact metric space,  $0 \leq \bar{\mu} \in M(S)$ , and suppose  $(a_i)_1^\infty \subset C(S)$  with  $\text{clspan}(a_i)_1^\infty = C(S)$ . For  $n = 1, 2, \dots$ , define*

$$F_n := \left\{ 0 \leq \mu \in M(S) \mid \int_S a_i d(\mu - \bar{\mu}) = 0, \quad i = 1, \dots, n \right\}.$$

*Suppose  $\mu_n \in F_n$  for each  $n$ . Then  $\mu_n \rightarrow \bar{\mu}$  weak\* in  $M(S)$ . In fact*

$$(1.8) \quad \int_S g d(\mu_n - \bar{\mu}) \rightarrow 0 \quad \text{as } n \rightarrow +\infty$$

*for any bounded Borel function  $g: S \rightarrow \mathbb{R}$ , continuous  $\bar{\mu}$ -a.e.*

*Proof.* Proposition 1.6 shows that  $\bigcap_{n=1}^\infty F_n = \{\bar{\mu}\}$ . Since  $\text{span}(a_i)_1^\infty$  is dense in  $C(S)$ , for some  $m$  and some  $\lambda \in \mathbb{R}^m$ ,  $\|\sum_{i=1}^m \lambda_i a_i - 1\|_\infty < \frac{1}{2}$  (where 1 denotes the constant function), so in particular  $\sum_{i=1}^m \lambda_i a_i(s) \geq \frac{1}{2}$  for all  $s \in S$ . Now for any  $\mu \in F_m$  we have

$$\begin{aligned} \frac{1}{2}\|\mu\| &= \frac{1}{2}\mu(S) \quad (\text{since } \mu \geq 0) \\ &= \int_S \left(\frac{1}{2}\right) d\mu \leq \int_S \left(\sum_{i=1}^m \lambda_i a_i\right) d\mu = \sum_{i=1}^m \lambda_i b_i. \end{aligned}$$

Thus  $F_m$  is closed and bounded in norm, so is weak\* compact by the Alaoglu-Bourbaki theorem [Holmes, 1975].

Now apply Proposition 1.5 with  $f(\cdot) := \delta(\cdot|F_m)$  (where  $\delta(x|C) := 0$  if  $x \in C$  and  $+\infty$  otherwise) to deduce that  $\mu_n \rightarrow \bar{\mu}$  weak\*. The final statement follows by [Ash, 1972, 4.5.1].  $\square$

As a special case, suppose  $S = [\alpha, \beta] \subset \mathbb{R}$ , and let  $g := \chi_{[\alpha, s]}$  in (1.8) (where  $\chi_c(x) := 1$  if  $x \in C$  and 0 otherwise). It then follows that  $\mu_n[\alpha, s] \rightarrow \bar{\mu}[\alpha, s]$  as  $n \rightarrow \infty$  for all  $s \in [\alpha, \beta]$ , or  $\mu_n \rightarrow \bar{\mu}$  in distribution. We thus recover the result of [Mead and Papanicolaou, 1984, Theorem 2].

**Error bounds.** An alternative approach to Theorem 1.7 involves a direct approximation argument. This idea also furnishes error bounds for the rate of convergence in (1.8) when  $g$  is continuous. As in Theorem 1.7, let  $S$  be a compact metric space with  $(a_i)_1^\infty \subset C(S)$ .

**Definition 1.9.** For any function  $g : S \rightarrow \mathbb{R}$  define the error function

$$E_n(g) := \inf \left\{ \left\| g - \sum_{i=1}^n \lambda_i a_i \right\|_\infty \mid \lambda \in \mathbb{R}^n \right\}.$$

Thus if  $\text{cl span}(a_i)_1^\infty = C(S)$  then for any  $g \in C(S)$ ,  $E_n(g) \rightarrow 0$  as  $n \rightarrow \infty$ . In fact, when  $S = [\alpha, \beta] \subset \mathbb{R}$  and  $a_i(s) := s^{i-1}$ , various theorems due to Jackson and others relate the rate of convergence of  $E_n(g)$  to zero with the smoothness of  $g$ . For example, if  $g$  is analytic on  $[\alpha, \beta]$  then there exists some  $\rho > 1$  and a constant  $K$  such that  $E_n(g) \leq K\rho^{-n}$  for all  $n$  (see [Lorentz, 1986]). Analogous results exist for approximation by trigonometric polynomials.

**Theorem 1.10.** *With the same notation as Theorem 1.7, suppose  $a_1 \equiv 1$ . Suppose  $\mu_n \in F_n$  for each  $n$ . Then for any bounded Borel function  $g : S \rightarrow \mathbb{R}$ ,*

$$\left| \int_S g d(\mu_n - \bar{\mu}) \right| \leq 2E_n(g)\bar{\mu}(S) \text{ for each } n.$$

*Proof.* By definition, given  $\delta > 0$  there exists  $\lambda \in \mathbb{R}^n$  with  $\|g - \sum_{i=1}^n \lambda_i a_i\|_\infty \leq E_n(g) + \delta$ . Therefore

$$\begin{aligned} \left| \int_S g d(\mu_n - \bar{\mu}) \right| &= \left| \int_S \left( g - \sum_{i=1}^n \lambda_i a_i \right) d(\mu_n - \bar{\mu}) \right| \\ &\leq \int_S \left| g - \sum_{i=1}^n \lambda_i a_i \right| d|\mu_n - \bar{\mu}| \\ &\leq \int_S (E_n(g) + \delta) d(\mu_n + \bar{\mu}) = 2(E_n(g) + \delta)\bar{\mu}(S). \end{aligned}$$

Since  $\delta$  was arbitrary the result follows.  $\square$

Note that since  $a_1 \equiv 1$ ,  $\bar{\mu}(S) = \int_S d\bar{\mu}$  is a known moment.

## 2. COMPACTNESS OF LEVEL SETS

The conditions we gave in the previous section for the convergence in value and optimal solution of the sequence of problems  $(P_n)$  depended on the objective function  $f$  having compact (in fact countably compact) level sets. We

are interested specifically in the case where the underlying space  $X$  is a Banach space with its associated weak or weak\* topology. The following concept is therefore useful.

**Definition 2.1.** Suppose  $X$  is normed and  $f: X \rightarrow (-\infty, +\infty]$ . Then  $f$  is coercive if  $\lim_{\gamma \rightarrow \infty} \inf_{\|x\| \geq \gamma} f(x) = +\infty$ .

**Proposition 2.2.** A function  $f$  is coercive if and only if all its level sets are bounded.

*Proof.* Suppose some level set is unbounded, so for some sequence  $(x_n)$  with  $\|x_n\| \rightarrow +\infty$ ,  $f(x_n) \leq \alpha < +\infty$  for all  $n$ . Then  $\inf_{\|x\| \geq \gamma} f(x) \leq \alpha$  for all  $\gamma$ , so  $\lim_{\gamma \rightarrow +\infty} \inf_{\|x\| \geq \gamma} f(x) \leq \alpha$  and  $f$  is not coercive.

Conversely, if  $f$  is not coercive then for some  $\alpha < +\infty$ ,  $\inf_{\|x\| \geq \gamma} f(x) < \alpha$  for all  $\gamma$ . Therefore  $\{x | f(x) \leq \alpha\}$  is unbounded.  $\square$

**Corollary 2.3.** Let  $X$  be the dual of a normed space, and suppose  $f: X \rightarrow (-\infty, +\infty]$  is coercive and weak\*-lower semicontinuous. Then  $f$  has weak\*-compact level sets.

*Proof.* The level sets of  $f$  are weak\*-closed by assumption, and bounded by Proposition 2.2, so they are weak\*-compact, by the Alaoglu-Bourbaki theorem [Holmes, 1975, p. 70].  $\square$

We are concerned with convex functions  $f$ . In this case there is a dual characterization, by the Moreau-Rockafellar Theorem. Given paired spaces  $(X, Y)$ , denote the Mackey topology on  $X$  by  $\tau(X, Y)$  (see, for example, [Schaefer, 1971]).

**Theorem 2.4** [Rockafellar, 1974, Theorem 10(a)]. Suppose  $(X, Y)$  are paired spaces, with  $f: X \rightarrow (-\infty, +\infty]$  a closed, proper, convex function with conjugate  $f^*: Y \rightarrow (-\infty, +\infty]$ . If  $f^*$  is bounded above on  $\tau(Y, X)$ -neighbourhood of 0 then the level sets,

$$(2.5) \quad \{x \in X | f(x) \leq \alpha\}, \quad \alpha < +\infty,$$

are  $\sigma(X, Y)$ -compact. Conversely, if one of the level sets (2.5) with  $\alpha > \inf f$  is  $\sigma(X, Y)$ -compact, then  $0 \in \text{core}(\text{dom } f^*)$  and  $f^*$  is  $\tau(Y, X)$ -continuous at 0 (so in particular,  $f^*$  is bounded above on a  $\tau(Y, X)$ -neighbourhood of 0).  $\square$

**Corollary 2.6.** Suppose, in Theorem 2.4,  $Y$  is a Banach space and  $X = Y^*$ . Then all the level sets (2.5) are weak\*-compact if and only if  $0 \in \|\cdot\| - \text{int}(\text{dom } f^*)$ .

*Proof.* This follows directly from the theorem, using the fact that, by [Rockafellar, 1974, Corollary 8B],  $f^*$  is continuous on  $\text{core}(\text{dom } f^*) = \text{int}(\text{dom } f^*)$ .  $\square$

We are particularly interested in objective functions  $f$  which are convex integral functionals:  $f: L_p(S) \rightarrow (-\infty, +\infty]$  defined by  $f(x) = \int_S \phi(x(s)) ds$ , where  $\phi$  is a convex function on  $\mathbb{R}$ . These integrals are defined as in [Rockafellar, 1974]. We will primarily be interested in the case where  $S$  is a finite

measure space. The following result is a specialization of a theorem [Rockafellar, 1974, Theorem 22] giving conditions for the level sets of such an  $f$  to be weakly compact.

**Theorem 2.7.** *Let  $(S, ds)$  be a complete, totally  $\sigma$ -finite measure space, and suppose  $\phi: \mathbb{R} \rightarrow (-\infty, +\infty]$  is closed, convex, proper.*

(A) *For  $1 \leq p \leq \infty$  and  $1/p + 1/q = 1$ , define  $f: L_p(S) \rightarrow (-\infty, +\infty]$  by  $f(x) = \int_S \phi(x(s)) ds$ . Suppose*

$$(2.8) \quad \phi^*(y(\cdot)) \in L_1(S) \quad \text{for all } y \in L_q(S).$$

*Then the level sets of  $f$  are  $\sigma(L_p(S), L_q(S))$ -compact.*

(B) *For  $1 < p \leq \infty$  and  $S$  a finite measure space, (2.8) is implied by the growth condition*

$$\phi^*(v) \leq \frac{a}{q}|v|^q + b \quad \text{for all } v,$$

*for some constants  $a > 0$  and  $b$ , or dually, for all  $u$ ,*

$$\phi(u) \geq \begin{cases} |u|^p/ap - b, & \text{if } p < +\infty, \\ \delta(u|[-a, a]) - b, & \text{if } p = +\infty. \end{cases}$$

*For  $p = 1$  and  $S$  finite, (2.8) is implied by  $\phi^*$  being everywhere finite.  $\square$*

**Example 2.9.** For applications to moment problems, the case where  $p = 1$  and  $\phi(u) = +\infty$  for  $u < 0$  is of particular interest. In this case, define  $d := \lim_{u \rightarrow +\infty} \phi(u)/u$  (which exists). It is easy to check that  $\phi^*$  is everywhere finite if and only if  $d = +\infty$  (see [Borwein and Lewis, 1988(b)]). As an example, consider (minus) the Boltzmann-Shannon entropy:

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

Then  $d = +\infty$  and  $\psi^*(v) = e^{v-1}$ . Applying the above results, the level sets

$$\left\{ 0 \leq x \in L_1[0, 1] \mid \int_0^1 x(s) \log x(s) ds \leq \alpha \right\}$$

are weakly compact ( $\alpha < +\infty$ ). This could be seen more directly from the Dunford-Pettis criterion for weak compactness in  $L_1$  (the approach taken in [Forte, Hughes, and Pales, 1988]).

The assumption that  $S$  is a finite measure space is important here. Writing  $I_\phi(x)$  for  $\int_S \phi(x(s)) ds$ , we know by [Rockafellar, 1974, Theorem 21] that with  $\psi$  as above and  $I_\psi: L_1(\mathbb{R}) \rightarrow (-\infty, +\infty]$ , the conjugate functional  $(I_\psi)^*: L_\infty(\mathbb{R}) \rightarrow (-\infty, +\infty]$  is given by  $(I_\psi)^* = I_{\psi^*}$ . Thus for  $y \in L_\infty(\mathbb{R})$ ,

$$(I_\psi)^*(y) = \int_{-\infty}^{\infty} e^{y(s)-1} ds,$$

from which it is clear that  $\text{dom}(I_\psi)^* = \emptyset$ . Thus by Theorem 2.4, the level sets of  $I_\psi$  are not weakly compact.

We can prove a converse to Theorem 2.7(A) (in the  $L_1$  case).

**Theorem 2.10.** *Let  $(S, \mu)$  be a complete, totally  $\sigma$ -finite measure space which is not purely atomic, and suppose  $\phi: \mathbb{R} \rightarrow (-\infty, +\infty]$  is closed, convex, proper. Suppose  $\phi^*$  is not everywhere finite. Then the level sets of  $I_\phi: L_1(S) \rightarrow (-\infty, +\infty]$  of the form  $\{x \in L_1(S) \mid \int_S \phi(x(s)) d\mu \leq \alpha\}$  are not weakly compact for  $\alpha > \inf I_\phi$ .*

*Proof.* Notice that the result holds vacuously if  $I_\phi$  is identically  $+\infty$ . By [Rockafellar, 1974, Theorem 21],  $(I_\phi)^*: L_\infty(S) \rightarrow (-\infty, +\infty]$  is given by  $(I_\phi)^* = I_{\phi^*}$ . Suppose  $\phi^*(v_0) = +\infty$ .

Since  $S$  is not purely atomic [Holmes, 1975, p. 106], there exists a subset  $S_0$  contained in the complement of the atoms of  $S$ , with  $0 < \mu(S_0) < 1$ . Since  $S_0$  is not an atom, there exists  $S_1 \subset S_0$  with  $0 < \mu(S_1) < \frac{1}{2}$ . Continuing inductively, we can construct a nested sequence  $S_0 \supset S_1 \supset S_2 \supset \dots$  with  $0 < \mu(S_n) < 2^{-n}$  for each  $n$ .

Now define a sequence  $(y_n) \subset L_\infty(S)$  by  $y_n(\cdot) := v_0 \chi_{S_n}(\cdot)$ . Pick any weakly compact set  $\Omega$  in  $L_1(S)$ . By the Pettis criterion for weak compactness in  $L_1$ ,

$$\sup\{|\langle x, y_n \rangle| \mid x \in \Omega\} = |v_0| \sup\left\{\left|\int_{S_n} x d\mu\right| \mid x \in \Omega\right\} \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

(see [Dunford and Schwartz, 1958, IV.8.11]). Thus for large  $n$ ,  $y_n \in \Omega^0 := \{y \mid \langle x, y \rangle \leq 1 \text{ for all } x \in \Omega\}$ . Since these sets form a base of neighbourhoods of the origin for  $\tau(L_\infty, L_1)$ ,  $y_n \rightarrow 0$  in the Mackey topology (see, for example, [Schaefer, 1971]). However

$$I_{\phi^*}(y_n) = \int_S \phi^*(v_0 \chi_{S_n}(s)) d\mu(s) = \mu(S_n) \phi^*(v_0) + \mu(S_n^c) \phi^*(0) = +\infty,$$

so  $I_{\phi^*}$  is not finite on any Mackey neighbourhood of 0. The result now follows from Theorem 2.4.  $\square$

As an example, consider the logarithmic entropy

$$\phi(u) = \begin{cases} -\log u, & u > 0, \\ +\infty, & u \leq 0. \end{cases}$$

Then the conjugate function is

$$\phi^*(v) = \begin{cases} -1 - \log(-v), & v < 0, \\ +\infty, & v \geq 0, \end{cases}$$

so the above theorem shows for example that

$$\left\{x \in L_1[0, 1] \mid \int_0^1 \log x(s) ds \geq \alpha\right\}$$

is not weakly compact for any finite  $\alpha$ . Thus it is not possible to apply the results of §1 directly with objective functions of this form. Such problems are studied in [Lewis, 1989], where it is shown for example that if  $x_n$  is the (unique) optimal solution for the problem

$$\begin{cases} \text{maximize} & \int_0^1 \log x(s) ds \\ \text{subject to} & \int_0^1 (x(s) - \hat{x}(s))s^i ds = 0, \quad i = 0, \dots, n, \quad 0 \leq x \in L_1[0, 1], \end{cases}$$

where  $\hat{x}$  is continuous and strictly positive, then  $1/x_n \rightarrow 1/\hat{x}$  weakly in  $L_1$ .

### 3. RIEMANN INTEGRATION

Consider the set

$$G_n := \left\{ 0 \leq x \in L_1[0, 1] \mid \int_0^1 (x(s) - \hat{x}(s))s^i ds = 0, \quad i = 0, \dots, n \right\}.$$

Thus  $G_n$  is the set of nonnegative densities agreeing with  $\hat{x}$  up to the first  $n$  moments (where  $0 \leq \hat{x} \in L_1$ ). If  $x_n \in G_n$  for  $n = 1, 2, \dots$  then Theorem 1.7 shows that

$$(3.1) \quad \int_0^1 (x_n(s) - \hat{x}(s))g(s) ds \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

for any bounded Borel function  $g: [0, 1] \rightarrow \mathbb{R}$  which is continuous a.e. (or in other words for Riemann integrable Borel functions). What we shall show in the next two sections is that, loosely speaking, this condition on  $g$  is also necessary for (3.1) to hold in general. In particular this shows that  $x_n$  does not necessarily converge to  $\hat{x}$  weakly in  $L_1$ . It is precisely this fact that motivates our interest in objective functions with weakly compact level sets.

In order to demonstrate this result in a general setting we begin in this section by developing a theory of Riemann integration in compact metric spaces. Throughout this section,  $S$  is a compact metric space with  $0 \leq \mu \in M(S)$ .

**Definition 3.2.** A function  $g: S \rightarrow \mathbb{R}$  is *Riemann integrable (with respect to  $\mu$ )* if given any  $\varepsilon > 0$  there exist  $p, q \in C(S)$  with  $p \leq g \leq q$  and  $\int_S (q-p) d\mu < \varepsilon$ . We say  $g$  is *essentially Riemann integrable (with respect to  $\mu$ )* if there exists a Riemann integrable Borel function  $h$  with  $h(s) = g(s)$   $\mu$ -a.e.

Thus a function is Riemann integrable when it can be  $L_1$ -approximated from above and below by continuous functions (cf. [Jacobs, 1978, 1.9.1]). When  $S = [\alpha, \beta]$  and  $\mu$  is Lebesgue measure, step functions and continuous functions  $L_1$ -approximate each other arbitrarily well from above and below. It follows by a standard result on the Riemann integral [Stromberg, 1981, 6.28], that Definition 3.2 agrees with the standard notion in this case.

Denote the bounded functions on  $S$  by  $B(S)$  and the Riemann integrable functions by  $RI(S, \mu)$ . Then  $C(S) \subset RI(S, \mu) \subset B(S)$ , and  $RI(S, \mu)$  and is a closed subspace of  $B(S)$  (with the uniform norm), by [Jacobs, 1978, 9.5.1]. It is easy to check that in general  $RI(S, \mu)$  will not be separable. We can also regard the (equivalence classes of) essentially Riemann integrable functions as a closed subspace of  $L_\infty(S, \mu)$ , which again will in general not be separable.

**Proposition 3.3.** *Suppose  $\text{support}(\mu) = S$ . Then a Borel function  $g$  is essentially Riemann integrable if and only if given any  $\varepsilon > 0$  there exists  $p, q \in C(S)$  with  $p(s) \leq g(s) \leq q(s)$  a.e. and  $\int_S(q - p) d\mu < \varepsilon$ .*

*Proof.* Suppose  $g$  has the given property. For each  $n = 1, 2, \dots$  there exist  $p_n, q_n \in C(S)$  with  $p_n(s) \leq g(s) \leq q_n(s)$  a.e. and  $\int_S(q_n - p_n) d\mu < 1/n$ . By replacing  $p_n$  with  $\bigvee_{i=1}^n p_i$  and  $q_n$  with  $\bigwedge_{i=1}^n q_i$  for each  $n$  (where  $\bigvee$  and  $\bigwedge$  denote pointwise maxima and minima respectively), we can assume  $p_1 \leq p_2 \leq \dots$ , and  $q_1 \geq q_2 \geq \dots$ . Suppose for some  $n$  and  $s_0 \in S$ ,  $p_n(s_0) > q_n(s_0)$ . By continuity there exists an open neighbourhood  $U$  of  $s_0$  with  $p_n(s) > q_n(s)$  for all  $s \in U$ . Since  $\text{support}(\mu) = S$ ,  $\mu(U) > 0$ , which contradicts  $p_n(s) \leq q_n(s)$  a.e. Thus  $p_n \leq q_n$  for all  $n$ .

Now define Borel functions  $u, v: S \rightarrow \mathbb{R}$  by  $u := \bigvee_{i=1}^\infty p_i$  and  $v := \bigwedge_{i=1}^\infty q_i$ . Then  $p_n \leq u \leq v \leq q_n$  for each  $n$ , and  $u(s) \leq g(s) \leq v(s)$  a.e. Define null sets

$$S_1 := \{s | g(s) < u(s)\} \quad \text{and} \quad S_2 := \{s | g(s) > v(s)\},$$

and then define a Borel function  $h: S \rightarrow \mathbb{R}$  by

$$h(s) := \begin{cases} u(s), & \text{if } s \in S_1 \cup S_2, \\ g(s), & \text{otherwise.} \end{cases}$$

Then  $h(s) = g(s)$  a.e., and  $p_n \leq u \leq h \leq v \leq q_n$  for each  $n$ , so  $h$  is Borel and Riemann integrable. Thus  $g$  is essentially Riemann integrable.

The converse is immediate.  $\square$

We are interested in the existence of functions which are not essentially Riemann integrable. The following construction is useful.

**Proposition 3.4.** *Suppose  $\text{support}(\mu) = S$  and  $\mu$  is nonatomic. Given any  $\varepsilon > 0$  there exists an open set  $A \subset S$  with  $0 < \mu(A) < \varepsilon$ , and such that  $\mu(A \cap U) > 0$  for every nonempty open set  $U \subset S$ .*

*Proof.* By [Jameson, 1974, 11.3],  $S$  is separable, so pick a dense sequence  $(s_i)_{i=1}^\infty$ . Write  $B_\varepsilon(\bar{s})$  for the open ball, centre  $\bar{s}$  with radius  $\varepsilon$ . For each  $i$ ,  $\bigcap_{k=1}^\infty B_{1/k}(s_i) = \{s_i\}$ , so since  $\mu$  is nonatomic,  $\mu(B_{1/k}(s_i)) \rightarrow 0$  as  $k \rightarrow \infty$ . Pick  $\delta_i > 0$  such that  $\mu(B_{\delta_i}(s_i)) < \varepsilon 2^{-i}$ , and define  $A := \bigcup_{i=1}^\infty B_{\delta_i}(s_i)$ . Then  $A$  is open with  $0 < \mu(A) < \sum_{i=1}^\infty \varepsilon 2^{-i} = \varepsilon$ . Furthermore, if  $\emptyset \neq U$  then  $s_i \in U$  for some  $i$ , so  $\emptyset \neq U \cap B_{\delta_i}(s_i)$  is open. Since  $\text{support}(\mu) = S$ ,  $\mu(A \cap U) \geq \mu(U \cap B_{\delta_i}(s_i)) > 0$  as required.  $\square$

**Theorem 3.5.** *Suppose  $\mu$  is not purely atomic. Then there exists a set  $A$ , open in  $S$ , such that the function  $\chi_A$  is not essentially Riemann integrable.*

*Proof.* Since  $\mu$  is not purely atomic there exists a subset  $S_1$  contained in the complement of the atoms of  $\mu$  with  $\mu(S_1) > 0$ . Since  $\mu$  is regular, there exists compact  $S_2 \subset S_1$  with  $\mu(S_2) > 0$ . Let  $S_0 := S_2 \cap \text{support}(\mu)$ . Then  $S_0$  is a compact metric space,  $\mu|_{S_0}$  is nonatomic, and  $\text{support}(\mu|_{S_0}) = S_0$ .

By Proposition 3.4, there exists a set  $A_0$ , open in  $S_0$ , with  $0 < \mu(A_0) < \frac{1}{2}\mu(S_0)$  and such that  $\mu(A_0 \cap U_0) > 0$  for every nonempty set  $U_0$ , open in  $S_0$ . In other words, there exists a set  $A$ , open in  $S$ , with  $0 < \mu(A \cap S_0) < \frac{1}{2}\mu(S_0)$  and such that  $\mu(A \cap U) > 0$  for every set  $U$ , open in  $S$ , with  $U \cap S_0 \neq \emptyset$ .

Now suppose  $p \in C(S)$  with  $p(s) \geq \chi_A(s)$  a.e. Suppose  $p(\bar{s}) < 1$  for some  $\bar{s} \in S_0$ . By continuity, there exists  $V$ , open in  $S$ , with  $\bar{s} \in V$  such that  $p(s) < 1$  for all  $s \in V$ . However  $V \cap S_0 \neq \emptyset$  implies that  $\mu(A \cap V) > 0$ , and this contradicts  $p(s) \geq 1$  a.e. on  $A$ . Thus  $p(s) \geq 1$  for all  $s \in S_0$ . Hence

$$\begin{aligned} \int_S (p - \chi_A) d\mu &= \int_{S_0^c} (p - \chi_A) d\mu + \int_{S_0} (p - \chi_A) d\mu \\ &\geq \int_{S_0} (p - \chi_A) d\mu = \int_{S_0} p d\mu - \mu(A \cap S_0) \\ &\geq \mu(S_0) - \mu(A \cap S_0) > \mu(S_0) - \frac{1}{2}\mu(S_0) = \frac{1}{2}\mu(S_0). \end{aligned}$$

It follows immediately from the definition that  $\chi_A$  is not essentially Riemann integrable.  $\square$

The last result of this section shows that the classical characterization of Riemann integrability on  $\mathbb{R}$  as continuity a.e. extends to this setting. (Note that we do not require  $h$  to be Borel.)

**Theorem 3.6.** *Let  $h: S \rightarrow \mathbb{R}$  be bounded. Then  $h$  is Riemann integrable if and only if it is continuous a.e.*

*Proof.* Define the upper and lower envelopes of  $h$ ,  $\bar{h}, \underline{h}: S \rightarrow \mathbb{R}$  respectively, by

$$\bar{h}(s) := \overline{\lim}_{t \rightarrow s} h(t), \quad \underline{h}(s) := \underline{\lim}_{t \rightarrow s} h(t), \quad \text{for } s \in S.$$

Then  $\underline{h} \leq h \leq \bar{h}$ ,  $\underline{h}$  is lsc,  $\bar{h}$  is usc, and

$$\bar{h} = \bigwedge \{q \geq h \mid q \text{ is usc}\}, \quad \underline{h} = \bigvee \{p \leq h \mid p \text{ is lsc}\}.$$

See [Ash, 1972, A6] for these ideas. Clearly  $h$  is continuous at  $s$  if and only if  $\underline{h}(s) = \bar{h}(s) = h(s)$ .

Now suppose  $h$  is continuous a.e., so  $\underline{h}(s) = \bar{h}(s)$  a.e. By [Ash, 1972, A6.6] there exist sequences  $(p_n)_1^\infty, (q_n)_1^\infty \subset C(S)$  with  $p_n \uparrow \underline{h}$  and  $q_n \downarrow \bar{h}$  as  $n \rightarrow \infty$ . By the monotone convergence theorem,

$$\lim_{n \rightarrow \infty} \int_S p_n d\mu = \int_S \underline{h} d\mu = \int_S \bar{h} d\mu = \lim_{n \rightarrow \infty} \int_S q_n d\mu.$$

Thus  $p_n \leq h \leq q_n$  for each  $n$  and  $\int_S (q_n - p_n) d\mu \downarrow 0$  as  $n \rightarrow \infty$ , so  $h$  is Riemann integrable.

Conversely, suppose  $h$  is not continuous a.e. so for some measurable  $S_0 \subset S$  with  $\mu(S_0) > 0$ ,  $\underline{h}(s) < \bar{h}(s)$  for all  $s \in S_0$ . Then for some  $\varepsilon > 0$  and some measurable  $S_1 \subset S$  with  $\mu(S_1) > 0$ ,  $\underline{h}(s) \leq \bar{h}(s) - \varepsilon$  for all  $s \in S_1$ .

Now for any  $p, q \in C(S)$  with  $p \leq h \leq q$ ,  $p \leq \underline{h}$  and  $q \geq \bar{h}$ , and so

$$\int_S (q - p) d\mu \geq \int_S (\bar{h} - \underline{h}) d\mu \geq \int_{S_1} (\bar{h} - \underline{h}) d\mu \geq \varepsilon \mu(S_1).$$

Thus  $h$  is not Riemann integrable.  $\square$

#### 4. DUALITY AND WEAK CONVERGENCE

We make the following assumptions:

$$(4.1) \quad \left\{ \begin{array}{l} S \text{ is a compact metric space,} \\ 0 \leq \mu \in M(S), \\ 0 \leq \hat{x} \in L_1(S, \mu), \\ g: S \rightarrow \mathbb{R}, \text{ Borel, bounded } \mu\text{-a.e.}, \\ \text{cl span}(a_i)_1^\infty = C(S). \end{array} \right.$$

Let  $G_n \subset L_1(S, \mu)$  be defined by

$$(4.2) \quad G_n := \left\{ 0 \leq x \in L_1(S, \mu) \mid \int_S (x - \hat{x}) a_i d\mu = 0, i = 1, \dots, n \right\}.$$

We consider sequences  $(x_n)_1^\infty$ , with  $x_n \in G_n$  for each  $n$ , and we are interested in conditions ensuring

$$(4.3) \quad \int_S (x_n - \hat{x}) g d\mu \rightarrow 0 \text{ as } n \rightarrow \infty.$$

From Theorem 1.7 we immediately obtain the following result.

**Theorem 4.4.** *Suppose assumptions (4.1) hold and define  $G_n$  as in (4.2). If  $g(s) = h(s)$   $\mu$ -a.e., where  $h$  is a bounded Borel function which is continuous a.e. with respect to the measure  $\hat{x} d\mu$ , then (4.3) holds for every sequence  $x_n \in G_n$ . In particular, this holds if  $g$  is essentially Riemann integrable.*

*Proof.* Since  $h$  is continuous a.e. with respect to the measure  $\hat{x} d\mu$ , we can apply Theorem 1.7 with  $d\bar{\mu} := \hat{x} d\mu$  and  $d\mu_n := x_n d\mu$ . For the last part, if  $g$  is essentially Riemann integrable then  $g(s) = h(s)$   $\mu$ -a.e., where  $h$  is Borel and Riemann integrable, so bounded, and continuous  $\mu$ -a.e. (and so  $\hat{x} d\mu$ -a.e.), by Theorem 3.6.  $\square$

In order to prove the main result we shall use a duality result from [Borwein and Lewis, 1988(a)].

**Definition 4.5.** Let  $X$  be a topological vector space, with convex  $C \subset X$ . Then the *quasirelative interior* of  $C$  (qri  $C$ ) is the set of those  $x \in C$  for which  $\text{cl cone}(C - x)$  is a subspace.

**Example 4.6.** For a  $\sigma$ -finite measure space  $(T, \mu)$  and  $X = L_1(T, \mu)$ ,

$$\text{qri}\{x \mid x(t) \geq 0 \text{ a.e.}\} = \{x \mid x(t) > 0 \text{ a.e.}\}$$

[Borwein and Lewis, 1988(a), Example 3.11].

**Theorem 4.7.** *Let  $X$  be a locally convex topological vector space;  $f: X \rightarrow (-\infty, +\infty]$  convex,  $A: X \rightarrow \mathbb{R}^n$  continuous and linear, and  $b \in \mathbb{R}^n$ . Suppose there exists  $\hat{x} \in \text{qri}(\text{dom } f)$  with  $A\hat{x} = b$ . Then*

$$\inf\{f(x) | Ax = b, x \in X\} = \max\{b^T \lambda - f^*(A^T \lambda) | \lambda \in \mathbb{R}^n\}$$

(where the right-hand side is attained).

*Proof.* See [Borwein and Lewis, 1988(a), Corollary 4.10].  $\square$

**Theorem 4.8.** *Suppose assumptions (4.1) hold and define  $G_n$  as in (4.2). Suppose there exists  $\delta > 0$  and a closed set  $S_0 \subset \text{support}(\mu)$  with  $\hat{x}(s) \geq \delta$   $\mu$ -a.e. on  $S_0$ , such that  $g|_{S_0}$  is not essentially Riemann integrable with respect to  $\mu|_{S_0}$ . Then there exists a sequence  $x_n \in G_n$ ,  $n = 1, 2, \dots$ , such that (4.3) fails.*

*Proof.* Without loss of generality (if necessary replace  $g$  with  $-g$ ), by Proposition 3.3 there exists  $\varepsilon > 0$  such that for any  $q \in C(S_0)$  with  $q(s) \geq g(s)$   $\mu|_{S_0}$ -a.e.,  $\int_{S_0} (q - g) d\mu \geq \varepsilon$ . (Note that  $g|_{S_0}$  is Borel.)

Consider the semi-infinite linear program,

$$(SIP_n) \quad \begin{cases} \sup & \int_{S_0} u g d\mu \\ \text{subject to} & \int_{S_0} u a_i d\mu = \int_{S_0} \hat{x} a_i d\mu, \quad i = 1, \dots, n, \\ & 0 \leq u \in L_1(S_0, \mu|_{S_0}). \end{cases}$$

Applying Theorem 4.7, the dual problem is

$$(DISP_n) \quad \begin{cases} \text{minimize} & \int_{S_0} \left( \sum_{i=1}^n \lambda_i a_i \right) \hat{x} d\mu \\ \text{subject to} & \sum_{i=1}^n \lambda_i a_i(s) \geq g(s) \quad \mu|_{S_0}\text{-a.e., } \lambda \in \mathbb{R}^n. \end{cases}$$

Now  $u := \hat{x}|_{S_0} \in \text{qri}(L_1(S_0, \mu|_{S_0})_+)$  by Example 4.6, and is feasible for  $(SIP_n)$ . Thus by Theorem 4.7,  $V(SIP_n) = V(DSIP_n)$ .

However, for any feasible  $\lambda$  for  $(DSIP_n)$ ,  $\sum \lambda_i a_i$  is continuous and thus  $\int_{S_0} (\sum_{i=1}^n \lambda_i a_i - g) d\mu \geq \varepsilon$ , so

$$\int_{S_0} \left( \sum_{i=1}^n \lambda_i a_i \right) \hat{x} d\mu \geq \int_{S_0} g \hat{x} d\mu + \varepsilon \delta \mu(S_0).$$

Note that  $\mu(S_0) > 0$  by assumption (otherwise  $g|_{S_0}(s) = 0$  a.e., so is essentially Riemann integrable). It follows that

$$V(SIP_n) = V(DSIP_n) \geq \int_{S_0} g \hat{x} d\mu + \varepsilon \delta \mu(S_0).$$

Thus for each  $n$  there exists  $u_n$ , feasible for  $(SIP_n)$ , with

$$\int_{S_0} u_n g d\mu \geq \int_{S_0} \hat{x} g d\mu + \frac{1}{2} \varepsilon \delta \mu(S_0).$$

Finally, define  $x_n \in L_1(S, \mu)$  by

$$x_n(s) = \begin{cases} u_n(s), & s \in S_0, \\ \hat{x}(s), & s \in S_0^c. \end{cases}$$

Then  $x_n \in G_n$  for each  $n$ , and

$$\int_S (x_n - \hat{x})g \, d\mu = \int_{S_0} (u_n - \hat{x})g \, d\mu \geq \frac{1}{2}\varepsilon\delta\mu(S_0) > 0.$$

Thus (4.3) fails.  $\square$

The duality argument above could be circumvented by a direct separation argument if so desired.

If, in Theorem 4.8,  $\hat{x} \in L_p(S, \mu)$ , where  $1 \leq p \leq \infty$ , then we can in fact find a sequence  $x_n \in G_n$ ,  $n = 1, 2, \dots$ , such that (4.3) fails, with each  $x_n \in L_p(S, \mu)$ . To see this, simply follow the same proof, with  $(SIP_n)$  reposed in  $L_p$ . The dual problem remains unchanged, and Theorem 4.7 still applies because  $\text{qri}((L_p)_+) = \{x|x(t) > 0 \text{ a.e.}\}$  by [Borwein and Lewis, 1988(a), Example 3.11] (where in the case  $p = \infty$  we use the weak\* topology).

Furthermore, if actually  $\hat{x} \in C(S)$  and  $S_0 = S$  in Theorem 4.8, then we can even require each  $x_n \in C(S)$ . Again follow the same proof, this time with  $(SIP_n)$  reposed in  $C(S)$ . The dual problem is again unchanged, and Theorem 4.7 still applies because  $\hat{x} \in \text{int}(C(S)_+) = \text{qri}(C(S)_+)$ .

Putting Theorems 4.4 and 4.8 together we obtain an exact characterization of those functions  $g$  for which (4.3) holds.

**Corollary 4.9.** *Suppose assumptions (4.1) hold and define  $G_n$  as in (4.2). Suppose further that  $\text{support}(\mu) = S$ , and for some  $\delta > 0$ ,  $\hat{x}(s) \geq \delta$   $\mu$ -a.e. Then (4.3) holds for all sequences  $x_n \in G_n$  if and only if  $g$  is essentially Riemann integrable.*

*Proof.* One direction is contained in Theorem 4.4, while the converse follows by taking  $S_0 := S$  in Theorem 4.8.  $\square$

It is interesting to observe that in fact both directions in the above proof could be accomplished using the duality argument: since the  $a_i$ 's are densely spanning in  $C(S)$ , when  $g$  is essentially Riemann integrable  $V(\text{DSIP}_n) \rightarrow \int_S g\hat{x} \, d\mu$  as  $n \rightarrow \infty$  (with  $S_0 := S$ ). The result then follows by weak duality.

The principal motivation of the last two sections was to show that  $x_n \in G_n$  for each  $n$  very rarely implies  $x_n \rightarrow \hat{x}$  weakly in  $L_1(S, \mu)$ .

**Corollary 4.10.** *Suppose  $\mu$  is nonatomic and  $\hat{x} \neq 0$ . Then there exists a sequence  $x_n \in G_n$  for each  $n$ , and an open set  $A \subset S$ , such that  $\int_A (x_n - \hat{x}) \, d\mu \rightarrow 0$  as  $n \rightarrow \infty$ . In particular,  $x_n \rightarrow \hat{x}$  weakly in  $L_1(S, \mu)$ .*

*Proof.* Since  $\hat{x} \neq 0$ , for some  $\delta > 0$ ,  $\mu\{s|\hat{x}(s) \geq \delta\} > 0$ . By the regularity of  $\mu$  there exists compact  $S_1 \subset \{s|\hat{x}(s) \geq \delta\}$  with  $\mu(S_1) > 0$ . Set  $S_0 := S_1 \cap \text{support}(\mu)$ , so  $S_0$  is compact,  $\mu(S_0) = \mu(S_1) > 0$ , and  $\hat{x}(s) \geq \delta$   $\mu$ -a.e. on  $S_0$ . Clearly  $\mu|_{S_0}$  is nonatomic and has support  $S_0$ . By Theorem 3.5 we can

find a set  $A_0$ , open in  $S_0$ , such that  $\chi_{A_0}$  is not essentially Riemann integrable with respect to  $\mu|_{S_0}$ . Now  $A_0 = A \cap S_0$  for some set  $A$ , open in  $S$ . Define  $g := \chi_A$  and apply Theorem 4.8.  $\square$

5. EXAMPLES

The following model gives some interesting examples. As before, we denote by  $L_2(\mathbb{T})$  the space of real-valued Lebesgue measurable functions on  $[-\pi, \pi]$  with the norm  $\|x\|_2 := (\frac{1}{2\pi} \int_{-\pi}^{\pi} x(s)^2 ds)^{1/2}$ , and the pointwise ordering.

$$(FS_n) \quad \begin{cases} \inf & \|x\|_2 \\ \text{subject to} & \int_{-\pi}^{\pi} (x(s) - \hat{x}(s)) \cos(js) ds = 0, \quad j = 0, \dots, n, \\ & \int_{-\pi}^{\pi} (x(s) - \hat{x}(s)) \sin(js) ds = 0, \quad j = 1, \dots, n, \\ & 0 \leq x \in L_2(\mathbb{T}). \end{cases}$$

Here,  $0 \leq \hat{x} \in L_2(\mathbb{T})$ . The problem therefore is to find the function in  $L_2(\mathbb{T})$  with minimal norm, given its nonnegativity and first  $(2n + 1)$  Fourier coefficients. By the strict convexity of  $\|\cdot\|_2$  and the weak compactness of the unit ball in  $L_2$ , if  $x_n$  is the unique optimal solution of  $(FS_n)$  then  $x_n \rightarrow \hat{x}$  weakly in  $L_2(\mathbb{T})$  (by Proposition 1.5). In fact we shall see in the next section that  $\|x_n - \hat{x}\|_2 \rightarrow 0$ .

For  $h \in L_2(\mathbb{T})$ , denote by  $S_n(h) \in C(\mathbb{T})$  the  $n$ th partial sum of the Fourier series for  $h$ . Thus

$$S_n(h)(s) = \frac{1}{2}d_0 + \sum_{j=1}^n (d_j \cos(js) + c_j \sin(js)),$$

where  $d_j := \frac{1}{\pi} \int_{-\pi}^{\pi} h(s) \cos(js) ds$  and  $c_j := \frac{1}{\pi} \int_{-\pi}^{\pi} h(s) \sin(js) ds$ .

**Lemma 5.1.** *If  $S_n(\hat{x}) \geq 0$  then it is the unique optimal solution of  $(FS_n)$ .*

*Proof.* Since  $\{1, \cos(s), \sin(s), \cos(2s), \dots\}$  forms an orthogonal basis for  $L_2(\mathbb{T})$ ,  $S_n(\hat{x})$  is the unique optimal solution of  $(FS_n)$  with the constraint  $x \geq 0$  omitted. The result follows.  $\square$

**Theorem 5.2.** *Suppose  $\hat{x} \in C(\mathbb{T})$  is of bounded variation and strictly positive. Then for all  $n$  sufficiently large,  $S_n(\hat{x})$  is the unique optimal solution of  $(FS_n)$ , and it approaches  $\hat{x}$  uniformly as  $n \rightarrow \infty$ .*

*Proof.*  $S_n(\hat{x}) \rightarrow \hat{x}$  uniformly by [Katznelson, 1968, II.2.2]. Since  $\hat{x}$  is strictly positive and continuous, by compactness  $S_n(\hat{x}) \geq 0$  for all  $n$  sufficiently large, and the result now follows by Lemma 5.1.  $\square$

**Example 5.3.** There exists a function  $h \in C(\mathbb{T})$  such that the partial Fourier sums  $S_n(h)$  are uniformly bounded, but  $S_n(h)(0) \not\rightarrow h(0)$  as  $n \rightarrow \infty$  (the Fourier series of  $h$  diverges at 0). See [Stromberg, 1981, p. 530] for this example. By adding a suitable constant to  $h$  we can assume  $h$  and  $S_n(h)$  are nonnegative for each  $n$ .

Set  $\hat{x} := h$  in  $(FS_n)$ , and repose the problem in  $C(\mathbb{T})$ . By Lemma 5.1,  $S_n(h)$  is the unique optimal solution of  $(FS_n)$ . Of course,  $\|S_n(h) - h\|_2 \rightarrow 0$  as

$n \rightarrow \infty$ , and in fact  $S_n(h)(s) \rightarrow h(s)$  a.e. [Stromberg, 1981, 8.26]. However,  $S_n(h) \not\rightarrow h$  pointwise, so  $S_n(h) \not\rightarrow h$  weakly in  $C(\mathbf{T})$ .

**Example 5.4.** Denote the function with constant value 2 by 2, and the ball of radius  $1/\sqrt{2}$ , center 2, by

$$B(2, 1/\sqrt{2}) := \{x \in L_2(\mathbf{T}) \mid \|x - 2\|_2 \leq 1/\sqrt{2}\}.$$

Define  $g \in L_2(\mathbf{T})$  by  $g(s) := \sum_{j=1}^\infty 2^{-j} \cos(js)$ . Consider the problem

$$(\mathbf{R}_n) \quad \begin{cases} \text{inf} & \delta(x|B(2, 1/\sqrt{2})) - \int_{-\pi}^\pi g(s)x \, ds \\ \text{subject to} & \int_{-\pi}^\pi x(s) \, ds = 4\pi, \\ & \int_{-\pi}^\pi \cos(js)x(s) \, ds = 0, \quad j = 1, \dots, n, \\ & \int_{-\pi}^\pi \sin(js)x(s) \, ds = 0, \quad j = 1, \dots, n, \\ & 0 \leq x \in L_2(\mathbf{T}). \end{cases}$$

Since  $\{1, \cos(s), \sin(s), \dots\}$  is an orthogonal basis of  $L_2(\mathbf{T})$ , any feasible solution of  $(\mathbf{R}_n)$  can be written

$$x(s) = 2 + \sum_{j>n} (d_j \cos(js) + c_j \sin(js)),$$

and for  $x \in B(2, 1/\sqrt{2})$  we require  $\sum_{j>n} (d_j^2 + c_j^2) \leq 1$ . For this  $x$ ,

$$\int_{-\pi}^\pi g(s)x(s) \, ds = \pi \sum_{j>n} 2^{-j} d_j.$$

The optimal solution of  $(\mathbf{R}_n)$  will thus be determined by the optimal solution of

$$\begin{cases} \text{sup} & \sum_{j>n} 2^{-j} d_j \\ \text{subject to} & \sum_{j>n} (d_j^2 + c_j^2) \leq 1, \quad c, d \in l_2, \end{cases}$$

which is easily seen to be given by  $c := 0$ ,  $d_j := k2^{-j}$  for  $j > n$ , and 0 for  $j \leq n$  for some positive constant  $k$ , using the inner product structure of  $l_2$ . The constant  $k$  is determined by

$$\sum_{j>n} k^2 (2^{-j})^2 = k^2 \sum_{j>n} 4^{-j} = k^2 \left( \frac{1}{3} 4^{-n} \right) = 1,$$

so  $k = 2^n \sqrt{3}$ . Thus the unique optimal solution of  $(\mathbf{R}_n)$  is

$$x_n(s) := 2 + \sqrt{3} \sum_{j>n} 2^{n-j} \cos(js).$$

(Notice that  $x_n(s) \geq 2 - \sqrt{3} \sum_{j>n} 2^{n-j} = 2 - \sqrt{3} > 0$ , so  $x_n$  is indeed feasible.) On the other hand, the unique feasible solution of the limiting problem  $(\mathbf{R}_\infty)$  is clearly  $\hat{x} := 2$ . But  $\|x_n - 2\|_2 = 1/\sqrt{2}$  for all  $n$ .

Thus in this example the objective function has weakly compact level sets (and so by Propositions 1.4 and 1.5, or direct computation,  $x_n \rightarrow 2$  weakly and  $V(\mathbf{R}_n) \uparrow V(\mathbf{R}_\infty)$ ) but  $x_n \not\rightarrow 2$  in norm.

Furthermore, as before,  $|x_n(s) - 2| \leq \sqrt{3}$  a.e., so no subsequence of  $(x_n)$  can converge to 2 a.e. (otherwise the dominated convergence theorem would give a contradiction to  $\|x_n - 2\|_2 = 1/\sqrt{2}$  for all  $n$ ).

**Example 5.5.** As in Theorem 1.7, let  $S$  be a compact metric space,  $0 \leq \bar{\mu} \in M(S)$ , and suppose  $\text{clspan}(a_i)_1^\infty = C(S)$ . Consider the semi-infinite linear program

$$(SILP_n) \quad \begin{cases} \inf & \int_S g \, d\mu \\ \text{subject to} & \int_S a_i \, d(\mu - \bar{\mu}) = 0, \quad i = 1, \dots, n, \\ & 0 \leq \mu \in M(S), \end{cases}$$

where  $g \in C(S)$ . By Theorem 1.7, if  $\mu_n$  is feasible for  $(SILP_n)$  then  $\mu_n \rightarrow \bar{\mu}$  weak\* in  $M(S)$ .

However, it is well known that for such problems we can restrict attention to measures supported on a finite number of points (see, for example, [Borwein, 1983]). Suppose  $\bar{\mu}$  is nonatomic. Then for any purely atomic measure  $\mu$ , supported on, say,  $\{s_1, \dots, s_m\}$ , given any  $\delta > 0$  there exists compact  $K \subset S \setminus \{s_1, \dots, s_m\}$  with  $\bar{\mu}(K) > \bar{\mu}(S) - \delta$  (since  $\bar{\mu}$  is regular). By Urysohn's lemma [Jameson, 1974, 12.2] there exists continuous  $h: S \rightarrow [0, 1]$  with  $h(s_i) = 0$ ,  $i = 1, \dots, m$ , and  $h|_K \equiv 1$ . Thus  $\int_S h \, d(\bar{\mu} - \mu) > \bar{\mu}(S) - \delta$ . It follows that  $\|\bar{\mu} - \mu\| \geq \bar{\mu}(S)$  (since  $\delta$  was arbitrary).

Thus in this case, no sequence of finitely supported optimal solutions  $\mu_n$  to  $(SILP_n)$  can possibly approach the limiting solution  $\bar{\mu}$  in norm.

### 6. NORM AND FINITE CONVERGENCE

Until now we have been concerned with weak convergence of solutions to truncated moment problems with respect to various spaces of functions. In particular, we saw that under reasonable conditions the solutions converged weakly with respect to essentially Riemann integrable functions. In this section we shall consider stronger types of convergence.

**Definition 6.1.** A normed space  $X$  is *locally uniformly convex* (LUC) if for any  $x, x_n \in U(X)$ , the unit ball of  $X$ ,  $n = 1, 2, \dots$ , with  $\|x_n + x\| \rightarrow 2$ ,  $\|x_n - x\| \rightarrow 0$ .

This property is called "localized uniform rotundity" in [Day, 1962], and is a weaker condition than uniform convexity. Since any  $L_p$  space with  $1 < p < \infty$  is uniformly convex (Clarkson's Theorem), these spaces are also LUC.

**Proposition 6.2.** Suppose  $X$  is LUC,  $x_n \rightarrow x$ , weakly, and  $\|x_n\| \rightarrow \|x\|$ . Then  $x_n \rightarrow x$  in norm.

*Proof.* The case  $x = 0$  is clear, so suppose without loss of generality that  $x_n, x \neq 0$ ,  $n = 1, 2, \dots$ , and define  $y := x/\|x\|$ ,  $y_n := x_n/\|x_n\|$ . Then  $\|y_n\| = \|y\| = 1$  for each  $n$ , and  $y_n \rightarrow y$  weakly.

By the Hahn-Banach theorem, there exists  $\phi \in U(X^*)$  with  $\phi(y) = 1$ . Now  $\phi(y_n + y) \rightarrow 2\phi(y) = 2$  as  $n \rightarrow \infty$ , so  $\|y_n + y\| \rightarrow 2$ . Thus by the definition of LUC,  $y_n \rightarrow y$  in norm, and the result follows.  $\square$

Motivated by this, we make the following definition.

**Definition 6.3.** Let  $X$  be a normed space. A function  $f: X \rightarrow (-\infty, +\infty]$  is *Kadec* if  $f(x_n) \rightarrow f(x) < +\infty$ , and  $x_n \rightarrow x$  weakly as  $n \rightarrow \infty$  implies  $x_n \rightarrow x$  in norm.

**Example 6.4.** (i) If  $X$  is finite-dimensional then any  $f: X \rightarrow (-\infty, +\infty]$  is Kadec.

(ii) If  $X = l_1$  then any  $f$  is Kadec, by Schur's lemma [Holmes, 1975].

(iii) If  $X$  is LUC then  $\|\cdot\|$  is Kadec.

Suppose that  $X$  is a normed space, and  $f$  has weakly compact level sets. Under the conditions of Proposition 1.5 we know that the solutions  $x_n$  of the finite problems  $(P_n)$  approach the solution  $\bar{x}$  of the limit problem  $(P_\infty)$  weakly. Furthermore, by Proposition 1.4,  $f(x_n) \rightarrow f(\bar{x})$ . If  $f$  is Kadec we can then deduce that  $x_n \rightarrow \bar{x}$  in norm.

For example, consider the following moment problems, where  $(S, \mu)$  is an arbitrary measure space,  $1 < p < \infty$ ,  $1/p + 1/q = 1$ , and  $a_i \in L_q(S, \mu)$ ,  $i = 1, 2, \dots$ , are densely spanning:

$$(LMP_n) \quad \begin{cases} \inf & \|x\|_p \\ \text{subject to} & \int_S a_i x \, d\mu = b_i, \quad i = 1, \dots, n, \\ & 0 \leq x \in L_p(S, \mu), \end{cases}$$

and the corresponding limiting problem  $(LMP_\infty)$ . Suppose that  $(LMP_\infty)$  is consistent. Since  $\|\cdot\|_p$  is strictly convex,  $(LMP_\infty)$  has a unique optimal solution,  $\bar{x}$ . Furthermore, since  $L_p(S, \mu)$  is reflexive [Holmes, 1975, p. 129], the unit ball is weakly compact. Thus, as above, the optimal solutions of  $(LMP_n)$ ,  $x_n$ , satisfy  $\|x_n\|_p \rightarrow \|\bar{x}\|_p$  and  $x_n \rightarrow \bar{x}$  weakly, so since  $L_p(S, \mu)$  is LUC,  $x_n \rightarrow \bar{x}$  in norm.

This raises the question of what happens when  $p = \infty$ . Providing  $(S, \mu)$  is  $\sigma$ -finite,  $L_\infty(S, \mu) = L_1(S, \mu)^*$  [Holmes, 1975], so the unit ball in  $L_\infty(S, \mu)$  is weak\* compact. Thus, as before,  $\|x_n\|_\infty \rightarrow \|\bar{x}\|_\infty$ , and  $x_n \rightarrow \bar{x}$  weak\*. However, the optimal solutions of such problems in general are multiples of characteristic functions of subsets of  $S$  (see [Borwein and Lewis, 1988(a)]), so clearly we will not in general have  $\|x_n - \bar{x}\|_\infty \rightarrow 0$ .

**Theorem 6.5.** Let  $X$  be a normed space, and suppose  $f, g: X \rightarrow (-\infty, +\infty]$  are weakly lsc, with  $f$  Kadec. Then  $f + g$  is Kadec.

*Proof.* Suppose  $x_n \rightarrow x$  weakly and  $f(x_n) + g(x_n) \rightarrow f(x) + g(x)$ . Since  $f$  and  $g$  are lsc,  $\underline{\lim} f(x_n) \geq f(x)$  and  $\underline{\lim} g(x_n) \geq g(x)$ . Furthermore

$$\begin{aligned} \overline{\lim} f(x_n) &= \overline{\lim}(f(x) + g(x) - g(x_n)) = f(x) + g(x) - \underline{\lim} g(x_n) \\ &\leq f(x) + g(x) - g(x) = f(x). \end{aligned}$$

Thus  $f(x_n) \rightarrow f(x)$ , and so since  $f$  is Kadec,  $x_n \rightarrow x$  in norm as required.  $\square$

**Example 6.6.** Suppose  $1 \leq m \in \mathbb{N}$  and  $(S, \mu)$  is a finite measure space. Consider the function  $f: L_m(S, \mu) \rightarrow (-\infty, +\infty]$  defined by

$$f(x) = \int_S e^{x(s)} d\mu = \frac{1}{m!} \|x\|_m^m + \int_S \sum_{0 \leq i \neq m} \frac{1}{i!} x(s)^i d\mu.$$

By Clarkson's theorem,  $\|\cdot\|_m$  is Kadec, so clearly  $\|\cdot\|_m^m/m!$  is also Kadec (and weakly lsc). The second function in the expression above is weakly lsc since the integrand is a continuous, proper convex function (see [Rockafellar, 1968]). Thus by Theorem 6.5,  $f$  is Kadec. Furthermore, since for  $x \geq 0$ ,  $f(x) \geq \|x\|_m^m/m!$ , the intersection of level sets of  $f$  with the positive cone are weakly compact. Note also,  $f$  is strictly convex.

Now suppose that  $a_i \in L_\infty(S, \mu)$ , and that  $\bar{x} \in L_\infty(S, \mu)$ . Consider the problem

$$(EX_n) \quad \begin{cases} \inf & \int_S e^{x(s)} d\mu \\ \text{subject to} & \int_S a_i(s)(x(s) - \bar{x}(s)) d\mu = 0, \quad i = 1, \dots, n, \\ & 0 \leq x \in L_1(S, \mu). \end{cases}$$

Since  $\bar{x}$  is feasible, with finite value, the unique optimal solution  $x_n$  of  $(EX_n)$  has finite value, so  $x_n \in L_m(S, \mu)$  for every  $1 < m \in \mathbb{N}$ . Reposing the problem in  $L_m$ , the fact that  $f$  has weakly compact level sets implies  $f(x_n) \uparrow f(\bar{x})$  and  $x_n \rightarrow \bar{x}$  weakly by Propositions 1.4 and 1.5. Since  $f$  is Kadec,  $x_n \rightarrow \bar{x}$  in  $\|\cdot\|_m$ . Finally, since  $m$  was an arbitrary integer, we obtain  $\|x_n - \bar{x}\|_p \rightarrow 0$  in every  $L_p$  space with  $p < +\infty$ . Thus for this choice of objective function we obtain very strong convergence with few restricting assumptions.

**Finite convergence.** In some cases we will have even stronger convergence than norm convergence:  $x_n$  converges to  $\bar{x}$  in a finite number of steps. The following example suffices to demonstrate this, although stronger results are possible using the results of [Borwein and Lewis, 1988(b)].

Let  $(S, \mu)$  be a finite measure space, and  $a_i \in L_\infty(S, \mu)$  for  $i = 1, 2, \dots$ . Suppose  $\phi: \mathbb{R} \rightarrow (-\infty, +\infty]$  is closed, and strictly convex on its domain, with  $\phi(x) = +\infty$  for  $x < 0$  and  $\phi(x) < +\infty$  for  $x > 0$ , and suppose  $\lim_{x \rightarrow \infty} \phi(x)/x = +\infty$ . Consider the following sequence of problems:

$$(EP_n) \quad \begin{cases} \inf & \int_S \phi(x(s)) d\mu(s) \\ \text{subject to} & \int_S a_i(s)x(s) d\mu(s) = b_i, \quad i = 1, \dots, n, \\ & 0 \leq x \in L_1(S, \mu). \end{cases}$$

Suppose finally that  $\bar{x}$  is defined by

$$\bar{x}(s) := (\phi^*)' \left( \sum_{i=1}^N \bar{\lambda}_i a_i(s) \right) > 0 \quad \text{a.e. on } S,$$

and is feasible for each  $(EP_n)$ .

Theorem 4.7 and [Borwein and Lewis, 1988(b), Theorem 4.8] imply that the value of  $(EP_n)$  is equal to the value of the dual problem

$$(DEP_n) \quad \begin{cases} \text{maximize} & g_n(\lambda) := \sum_{i=1}^n b_i \lambda_i - \int_S \phi^* \left( \sum_{i=1}^n \lambda_i a_i(s) \right) d\mu(s), \\ \text{subject to} & \lambda \in \mathbb{R}^n, \end{cases}$$

and that if  $\lambda^n$  is optimal for  $(\text{DEP}_n)$  then the unique optimal solution of  $(\text{EP}_n)$  is  $x_n$ , where

$$x_n(s) := (\phi^*)' \left( \sum_{i=1}^n \lambda_i^n a_i(s) \right), \quad s \in S.$$

However, essentially following the proof of [Borwein and Lewis, 1988(b), Theorem 4.8],

$$(\nabla g_n(\lambda))_i = b_i - \int_S a_i(s) (\phi^*)' \left( \sum_{i=1}^n \lambda_i a_i(s) \right) d\mu(s),$$

from which it follows that if we define  $\bar{\lambda}^n \in \mathbb{R}^n$  by

$$\bar{\lambda}_i^n := \begin{cases} \bar{\lambda}_i, & i \leq N, \\ 0, & i > N, \end{cases}$$

then  $\nabla g_n(\bar{\lambda}^n) = 0$ , so  $\bar{\lambda}^n$  is optimal for  $(\text{DEP}_n)$ . Thus  $(\phi^*)'(\sum_{i=1}^n \bar{\lambda}_i^n a_i(\cdot)) = \bar{x}$  is uniquely optimal for  $(\text{EP}_n)$  for all  $n \geq N$ . So we see that the optimal solution of  $(\text{EP}_n)$  converges to the optimal solution of  $(\text{EP}_\infty)$ , the limiting problem, in at most  $N$  steps.

**Example 6.7.** Consider the case of Example 3.4, where  $S = [0, 1]$  with Lebesgue measure,  $a_i(s) = s^{i-1}$  for each  $i$ , and  $\phi$  is (minus) the Boltzmann-Shannon entropy:

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

Then  $\phi^*(v) = e^{v-1}$ . Thus whenever the limiting solution  $\bar{x}(s) = e^{P(s)}$ , where  $P(s)$  is a polynomial in  $s$ , the solutions of the corresponding truncated moment problems will converge to  $\bar{x}$  in a finite number of steps.

#### REFERENCES

- R. B. Ash, 1972: *Measure, integration, and functional analysis*, Academic Press, New York.
- G. A. Baker and P. Graves-Morris, 1980: *Pade approximants*, Addison-Wesley, Reading, Mass.
- A. Bén-Tal, J. M. Borwein, and M. Teboulle, 1988(a): *A dual approach to multidimensional  $L_p$  spectral estimation problems*, SIAM J. Control Optim. **26**, 985–996.
- , 1988(b): *Spectral estimation via convex programming* (to appear).
- J. M. Borwein, 1983: *Semi-infinite programming duality: how special is it?*, in *Semi-Infinite Programming and Applications* (A. V. Fiacco and K. O. Kortanek, eds.), Springer-Verlag, Berlin.
- J. M. Borwein and A. S. Lewis, 1988(a): *Partially finite convex programming*, Parts I and II, Math. Programming (to appear).
- , 1988(b): *Duality relationships for entropy-like minimization problems*, SIAM J. Control. Optim. (to appear).
- J. P. Burg, 1975: *Maximum entropy spectral analysis*, Ph. D. dissertation, Stanford University, Stanford, Calif.
- E. W. Cheney, 1966: *Introduction to approximation theory*, McGraw-Hill, New York.

- M. M. Day, 1962: *Normed linear spaces*, Academic Press, New York.
- N. Dunford and J. T. Schwartz, 1958: *Linear operators*, Part I, Interscience, New York.
- B. Forte, W. Hughes and Z. Pales, 1989: *Maximum entropy estimators and the problem of moments*, *Rend. Mat. (7)* **9**, 689–699.
- B. K. Goodrich and A. Steinhardt, 1986:  *$L_2$  spectral estimation*, *SIAM J. Appl. Math.* **46**, 417–428.
- F. Hausdorff, 1921: *Summationsmethoden und momentfolgen*. I, *Math. Z.* **9**, 74–109.
- R. B. Holmes, 1975: *Geometric functional analysis and its applications*, Springer-Verlag, New York.
- L. D. Irvine, S. P. Marin, and P. W. Smith, 1986: *Constrained interpolation and smoothing*, *Constr. Approx.* **2**, 128–151.
- K. Jacobs, 1978: *Measure and integral*, Academic Press, New York.
- G. J. O. Jameson, 1974: *Topology and normed spaces*, Chapman and Hall, London.
- S. Karlin and W. Studden, 1966: *Tchebycheff systems and applications in analysis and statistics*, Interscience.
- Y. Katznelson 1968: *An introduction to harmonic analysis*, Wiley, New York.
- S. M. Kay and S. L. Marple, 1981: *Spectrum analysis—a modern perspective*, *Proc. IEEE* **69**, 1380–1419.
- J. L. Kelley, 1955: *General topology*, Van Nostrand, Princeton, N. J.
- H. J. Landau, (ed.), 1987: *Moments in mathematics*, Short Course Lecture Notes, vol. 37, Amer. Math. Soc., Providence, R.I.
- S. W. Lang and J. H. McClellan, 1983: *Spectral estimation for sensor arrays*, *IEEE Trans. Acoust. Speech Signal Process.* **31**, 349–358.
- A. S. Lewis, 1989: *The convergence of entropic estimates for moment problems*, in *Functional Analysis/Optimization* (J. R. Giles and S. Fitzpatrick, eds.), *Proceedings of the Centre for Mathematical Analysis*, vol. 20, Canberra, Australia, pp. 100–115.
- G. G. Lorentz, 1986: *Approximation of functions*, Chelsea, New York.
- L. R. Mead and N. Papanicolaou, 1984: *Maximum entropy in the problem of moments*, *J. Math. Phys.* **25**, 2404–2417.
- R. T. Rockafellar 1968: *Integrals which are convex functionals*, *Pacific J. Math.* **24**, 525–539.
- , 1974: *Conjugate duality and optimization*, SIAM, Philadelphia, Pa.
- W. Rudin, 1966: *Real and complex analysis*, McGraw-Hill, New York.
- H. H. Schaefer, 1971: *Topological vector spaces*, Springer-Verlag, New York.
- K. R. Stromberg, 1981: *An introduction to classical real analysis*, Wadsworth, Belmont, Calif.
- D. V. Widder, 1941: *The Laplace transform*, Princeton Univ. Press, Princeton, N.J.

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