THE CONVERGENCE OF ENTROPIC ESTIMATES FOR MOMENT PROBLEMS

A.S. Lewis

Abstract. We show that if x_n is optimal for the problem

$$\sup \left\{ \int_0^1 \log\ x(s)ds \ \Big|\ \int_0^1 (x(s)-\hat{x}(s))s^i ds = 0\ ,\ i = 0,\ldots,n\ ,\ 0 \le x \in L_1[0,1] \right\},$$

then $\frac{1}{x_n} \to \frac{1}{\hat{x}}$ weakly in L_1 (providing $\hat{x}$ is continuous and strictly positive). This result is a special case of a theorem for more general entropic objectives and underlying spaces.

Key Words: moment problem, entropy, semi-infinite program, duality, normal convex integrand.

AMS 1985 subject classification:

Primary: 41A46, 42A70

Secondary: 90C25, 49D39

§1. Introduction

The following problem, known as a 'moment problem' or 'underdetermined inverse problem', occurs frequently in physical and other applications (see for example [Mead and Papanicolaou, 1984]). We are given a finite number of 'moments' $\int_S \hat{x} a_i \, ds$, for $i = 1,...,n$, where (S,ds) is some measure space and $a_i \in L_\infty(S)$, $i = 1,...,n$ are given, and we wish to estimate the unknown non-negative density $\hat{x} \in L_1(S)$. One popular technique is to choose an estimate x to have the given moments and in order to minimize some objective function. Typically the objective function used is of the form $\int_S \phi(x(s))ds$, where $\phi : \mathbb{R} \to (-\infty,\infty]$ is convex, so the problem becomes

$$(MP_n) \quad \begin{cases} \inf & \int_S \phi(x(s))ds \\ \text{subject to} & \int_S (x-\hat{x})a_i ds = 0 \ , \quad i = 1,...,n \ , \\ & 0 \le x \in L_1(S) \ . \end{cases}$$

Various functions ϕ have been tried, including the classical 'maximum entropy' approach where $\phi(u) = u \log u$, (see [Mead and Papanicolaou, 1984] and the references therein), other measures of entropy such as $\phi(u) = -\log u$ (for example [Johnson and Shore, 1984]), and more recently norm objectives such as $\phi(u) = \frac{1}{2}u^2$ [Goodrich and Steinhardt, 1986].

A survey of objective functions, along with solution techniques based on duality, may be found in [Ben-Tal, Borwein and Teboulle, 1988], and these techniques, together with the question of the existence of optimal solutions for (MP_n), are studied further in [Borwein and Lewis, 1988(a)].

For this approach to the moment problem to be practically useful we would hope that as the number of known moments increases, our estimate converges in some sense to $\hat{x}$. Further conditions on the a_i's will be necessary to ensure this. Suppose therefore that S is a compact Hausdorff space, ds a regular Borel measure, and that the a_i's are densely spanning in $C(S)$. As essentially observed in [Mead and Papanicolaou, 1984], if x_n is feasible for (MP_n) then $x_n ds \to \hat{x}\, ds$ weak* in $M(S)$. However, it need not be the case that $x_n \to \hat{x}$ weakly in $L_1(S)$. Indeed, the following result appears in [Borwein and Lewis, 1988(b)].

<u>Theorem 1.1</u>. Suppose S is a compact metric space, ds a non-negative regular Borel measure, $cl(span(a_i)_1^\infty) = C(S)$, and for some $K, \delta > 0$, $\delta \le \hat{x}(s) \le K$ a.e. For a given $y \in L_\infty(S)$, a necessary and sufficient condition that $\int_S (x_n - \hat{x})\, y\, ds \to 0$ for every sequence (x_n) with x_n feasible for (MP_n) is that $y = z$ a.e. for some function z, continuous a.e.

□

It follows from this that in order to guarantee the weak convergence of optimal solutions of (MP_n) to $\hat{x}$ we will need further conditions on the objective function. One possibility is to require it to have weakly compact level sets. When (S,ds) is complete and totally σ-finite, and ϕ is a closed, convex, proper function with conjugate ϕ^* everywhere finite, a result of Rockafellar shows that the objective function in (MP_n) has weakly compact level sets. Under the further conditions on S, ds and the a_i's above, this will ensure that if x_n is optimal for (MP_n) then $x_n \to \hat{x}$ weakly in $L_1(S)$ (see [Borwein and Lewis, 1988(a)]). This will apply in particular to the Boltzmann-Shannon entropy defined by

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

For this objective function in the special case where $S = [0,1]$, ds is Lebesgue measure, and $a_i(s) = s^{i-1}$, the weak convergence of x_n to $\hat{x}$ was shown in [Forte, Hughes and Pales, 1988].

However, in the case of the logarithmic entropy,

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

ϕ^* is not everywhere finite, so the objective function typically will not have weakly compact level sets [Borwein and Lewis, 1988(b)], and this technique cannot be applied. The results presented in this paper will adopt a different approach to show, under suitable conditions, that if x_n is optimal for (MP_n) then $\phi'(x_n(\cdot)) \to \phi'(\hat{x}(\cdot))$ weakly in $L_1(S)$.

§2. Minimizing Sequences

Throughout this paper the finite-dimensional convex analytic terminology used will be that of [Rockafellar, 1970]. Suppose (S,ds) is a finite measure space. For a closed, convex, proper function $\theta : \mathbb{R} \to (-\infty,\infty]$, define $I_\theta : L_1(S) \to (-\infty,\infty]$ by $I_\theta(v) := \int_S \theta(v(s))ds$. Using the theory of normal convex integrands in [Rockafellar, 1974], I_θ is a well-defined convex functional with conjugate $(I_\theta)^* : L_\infty(S) \to (-\infty,\infty]$ given by $(I_\theta)^*(z) = I_{\theta^*}(z) = \int_S \theta^*(z(s))ds$.

For a given $y \in L_\infty(S)$ we shall be interested in the function $f : L_1(S) \to (-\infty,\infty]$ defined by $f(v) := I_\theta(v) - \langle v,y \rangle$. It is easy to check that the conjugate function $f^* : L_\infty(S) \to (-\infty,\infty]$ is given by $f^*(z) = I_{\theta^*}(z+y)$. We shall make the following assumptions about θ and y:

$$(2.1) \qquad \begin{cases} \theta^* \text{ is twice continuously differentiable on } \operatorname{int}(\operatorname{dom} \theta^*), \\ [\operatorname{ess\,inf} y, \operatorname{ess\,sup} y] \subset \operatorname{int}(\operatorname{dom} \theta^*). \end{cases}$$

<u>Proposition 2.2</u>. The infimum of f is attained uniquely by $\bar{v} \in L_\infty(S)$, where $\bar{v}(s) := (\theta^*)'(y(s))$ a.e., and $\inf f = -I_{\theta^*}(y)$.

Proof. By [Rockafellar, 1970, 23.5],

$$\theta(v(s)) - v(s)y(s) \geq -\theta^*(y(s)) \quad \text{a.e.} \ ,$$

with equality a.e. if and only if $v(s) = (\theta^*)'(y(s))$ a.e. The result follows by integrating over S. The fact that $\bar{v} \in L_\infty(S)$ follows from the continuity of $(\theta^*)'$ and the compactness of the essential range of y (2.1).

□

Lemma 2.3. For $w \in L_\infty(S)$ and $b \in \mathbb{R}$

$$\inf\{f(v) \mid \langle v,w\rangle \geq b\ ,\ v \in L_1(S)\} \ \geq$$

$$\sup\{b\lambda - f^*(\lambda w) \mid 0 \leq \lambda \in \mathbb{R}\}\ .$$

Proof. For $\langle v,w\rangle \geq b$ and $\lambda \geq 0$,

$$b\lambda - f^*(\lambda w) \leq \langle v,\lambda w\rangle - f^*(\lambda w) \leq f(v)\ ,$$

and the result follows, taking $\inf$ over v and $\sup$ over λ. □

Theorem 2.4. Suppose $(v_i)_1^\infty \subset L_1(S)$ and $f(v_i) \to \inf f$. Then $v_i(\cdot) \to (\theta^*)'(y(\cdot))$ (the unique minimizer for f) weakly in $L_1(S)$.

Proof. Suppose not, so for some $w \in L_\infty(S)$,

$$\int_S [v_i(s) - (\theta^*)'(y(s))]w(s)ds \geq 1 \quad , \text{ each } i\ .$$

Applying Lemma 2.3 with $b := 1 + \int_S (\theta^*)'(y(s))w(s)ds$ it follows that for all $0 \leq \lambda \in \mathbb{R}$,

(2.5) $\qquad b\lambda - f^*(\lambda w) \leq \inf f\ .$

Now pick $\delta > 0$ such that

$$[(\text{ess inf } y)-\delta\ ,\ (\text{ess sup } y)+\delta] \subset \text{int}(\text{dom } \theta^*)\ .$$

By the continuity of $(\theta^*)''$, there exists M such that for all $u \in [(\text{ess inf } y)-\delta\ ,\ (\text{ess sup } y)+\delta]$, $0 \le (\theta^*)''(u) \le M$. Since $w \in L_\infty(S)$, for all λ sufficiently small

$$y(s) + \lambda w(s) \in [(\text{ess inf } y)-\delta\ ,\ (\text{ess sup } y)+\delta] \ \text{ a.e.}\ ,$$

so by the mean value theorem,

$$\theta^*(y(s)+\lambda w(s)) \le \theta^*(y(s)) + \lambda w(s)(\theta^*)'(y(s)) + \frac{1}{2} M(\lambda w(s))^2\ ,\ \text{a.e.}$$

Integrating over S gives

$$f^*(\lambda w) \le -\inf f + \lambda(b-1) + \lambda^2 \left(\frac{1}{2} M \int_S w(s)^2 ds \right)\ ,$$

for all λ sufficiently small. But then by (2.5), for all $\lambda \ge 0$ sufficiently small

$$-\lambda + \left(\frac{1}{2} M \int_S w(s)^2 ds \right) \lambda^2 \ge \inf f + f^*(\lambda w) - b\lambda \ge 0\ ,$$

which is a contradiction for small $\lambda > 0$. □

A similar, less direct approach to this result uses the results on minimizing sequences in [Rockafellar, 1974] .

§3. Weak Convergence

We are now ready to return to the original problem.

$$(P_n) \quad \begin{cases} \inf & \int_S \phi(x(s))ds \\ \text{subject to} & \int_S (x-\hat{x})a_i ds = 0 \ , \ i = 1,\dots,n \ , \\ & x \in L_1(S) \ . \end{cases}$$

Notice we have removed the constraint $x \geq 0$, assuming it to be implicit in the function ϕ . We make the following assumptions:

$$(3.1) \quad \begin{cases} S \text{ is a compact Hausdorff space,} \\ ds \text{ is a non-negative regular Borel measure on } S \ , \\ cl(span(a_i)_1^\infty) = C(S) \ , \\ \phi : \mathbb{R} \to (-\infty,\infty] \text{ is closed, convex, proper, essentially smooth} \\ \quad \text{and essentially strictly convex, and twice continuously} \\ \quad \text{differentiable on } int(dom\ \phi) \ , \\ \hat{x} \in C(S) \text{ with } [\min \hat{x} , \max \hat{x}] \subset int(dom\ \phi) \ . \end{cases}$$

A closed, convex, proper function $\phi : \mathbb{R} \to (-\infty,\infty]$ is essentially strictly convex if and only if it is strictly convex on dom ϕ (see [Borwein and Lewis, 1988(a)]), and is essentially smooth if it is differentiable on int(dom ϕ) and $|\phi'(u)| \to +\infty$ if u approaches a point in the boundary of dom ϕ . Functions which are both essentially smooth and essentially strictly convex are said to be 'of Legendre type', and have the following property.

Theorem 3.2. [Rockafellar, 1970, 26.5] The function ϕ is of Legendre type if and only if ϕ^* is. In this case the gradient map $\phi' : \mathrm{int}(\mathrm{dom}\,\phi) \to \mathrm{int}(\mathrm{dom}\,\phi^*)$ is 1-1 , onto, continuous, and with continuous inverse $(\phi^*)'$.

□

The dual problem for (MP_n) , from [Borwein and Lewis, 1988(a)], is

$$(DP_n) \qquad \begin{cases} \text{maximize} & \left\langle \hat{x}, \sum_{i=1}^{n} \lambda_i a_i \right\rangle - I_{\phi^*}\left(\sum_{i=1}^{n} \lambda_i a_i \right) \\ \text{subject to} & \lambda \in \mathbb{R}^n . \end{cases}$$

Theorem 3.3. The values of (P_n) and (DP_n) are equal, with attainment in (DP_n).

Proof. This follows from the duality theorem [Borwein and Lewis, 1988(a), 2.4], since $\hat{x} \in \mathrm{qri}(\mathrm{dom}\, I_\phi)$ (or in other words $\mathrm{cl\ cone}(\mathrm{dom}\, I_\phi - \hat{x})$ is a subspace) so the required constraint qualification is satisfied. To see this, observe that from 3.1, $\hat{x} \in \|\cdot\|_\infty - \mathrm{int}(\mathrm{dom}(I_\phi|_{L_\infty(S)}))$ (restricting I_ϕ to $L_\infty(S) \subset L_1(S)$) , so certainly $\mathrm{cone}(\mathrm{dom}\, I_\phi - \hat{x}) \supset L_\infty(S)$. Since $L_\infty(S)$ is dense in $L_1(S)$ [Rudin, 1966, 3.13], the result follows. □

The question of attainment in the primal problem (P_n) is harder. We have the following result from [Borwein and Lewis, 1988(a)].

Theorem 3.4. Suppose assumptions (3.1) hold for the problem (P_n). Suppose further that $S = [\alpha,\beta] \subset \mathbb{R}$ with ds Lebesgue measure, that the a_i's are locally Lipschitz (or in particular continuously differentiable), and that $\phi(u) = +\infty$ for $u < 0$. Define two numbers,

$$d := \lim_{u \to +\infty} \frac{\phi(u)}{u} \text{ , and if } d < +\infty \text{ ,}$$

$$c := \underline{\lim}_{u \to +\infty} (d - \phi'(u))u \text{ .}$$

Suppose either $d = +\infty$, or $c > 0$, and there exists $\mu \in \mathbb{R}^n$ with $\sum_{i=1}^{n} \mu_i a_i(s) < d$ for all $s \in [\alpha,\beta]$ (which holds in particular for all sufficiently large n, or if $a_1 \equiv 1$). Then there exists an optimal solution of (P_n). □

It is easy to check for example that the conditions on ϕ are satisfied in particular for the two entropies in the introduction.

When we know the existence of an optimal solution, it is easy to identify it.

Theorem 3.5. Suppose x_n is optimal for (P_n) and λ^n is optimal for (DP_n). Then

$$\sum_{i=1}^{n} \lambda_i^n a_i(s) = \phi'(x_n(s)) \quad , \text{ a.e., and}$$

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

Proof. If x_n and λ^n are both optimal then

$$I_\phi(x_n) = \left\langle \hat{x} , \sum_{i=1}^{n} \lambda_i^n a_i \right\rangle - I_{\phi^*}\left(\sum_{i=1}^{n} \lambda_i^n a_i\right)$$

$$= \left\langle x_n , \sum_{i=1}^{n} \lambda_i^n a_i \right\rangle - I_{\phi^*}\left(\sum_{i=1}^{n} \lambda_i^n a_i\right) ,$$

so it follows that

$$\int_S \left[\phi(x_n(s)) + \phi^*\left(\sum_{i=1}^{n} \lambda_i^n a_i(s)\right) - x_n(s) \sum_{i=1}^{n} \lambda_i^n a_i(s) \right] ds = 0 \quad .$$

Thus by [Rockafellar, 1970, 23.5],

$$\sum_{i=1}^{n} \lambda_i^n a_i(s) \in \partial\phi(x_n(s)) \quad , \text{ a.e.,}$$

and the result follows by Theorem 3.2. □

This result shows in particular that primal optimal solutions, if they exist, are unique. This is clear alternatively from the strict convexity of I_ϕ .

Let us denote the value of a problem by $V(\cdot)$.

<u>Theorem 3.6</u>. $V(DP_n) \uparrow I_\phi(\hat{x})$ as $n \to \infty$.

<u>Proof</u>. Clearly $V(DP_n)$ is increasing in n. Since $[\min \hat{x}, \max \hat{x}] \subset \text{int}(\text{dom}\,\phi)$, and ϕ is continuously differentiable on $\text{int}(\text{dom}\,\phi)$, $\phi' \circ \hat{x} \in C(S)$, and by Theorem 3.2, $[\min \phi'\circ\hat{x}, \max \phi'\circ\hat{x}] \subset \text{int}(\text{dom}\,\phi^*)$. Pick $\varepsilon > 0$ such that

$$[\min(\phi'\circ\hat{x}) - \varepsilon, \max(\phi'\circ\hat{x}) + \varepsilon] \subset \text{int}(\text{dom}\,\phi^*),$$

so ϕ^* is uniformly continuous on $[\min(\phi'\circ\hat{x}) - \varepsilon, \max(\phi'\circ\hat{x}) + \varepsilon]$. Since $\text{span}\,(a_i)_1^\infty$ is dense in $C(S)$ it follows that, given $\delta > 0$, there exists N and $\lambda \in \mathbb{R}^N$ such that

$$\|(\phi'\circ\hat{x}) - \sum_{i=1}^{N} \lambda_i a_i\|_\infty < \delta,$$

and (for $\delta < \varepsilon$) by the uniform continuity of ϕ^* we can also ensure that

$$\|\phi^*\circ\phi'\circ\hat{x} - \phi^*\circ\sum_{i=1}^{N} \lambda_i a_i\|_\infty < \delta .$$

We then have

$$<\hat{x}, \sum_{i=1}^{N} \lambda_i a_i> - I_{\phi^*}(\sum_{i=1}^{N} \lambda_i a_i)$$

$$= \int_S \left[\hat{x}(s) \sum_{i=1}^{N} \lambda_i a_i(s) - \phi^*(\sum_{i=1}^{N} \lambda_i a_i(s)) \right] ds$$

$$\geq \int_S \left[\hat{x}(s)\phi'(\hat{x}(s)) - \delta|\hat{x}(s)| - \phi^*(\phi'(\hat{x}(s))) - \delta \right] ds$$

$$= \int_S \phi(\hat{x}(s)) - \delta(1 + |\hat{x}(s)|)ds$$

$$= I_\phi(\hat{x}) - \delta\int_S [1 + |\hat{x}(s)|]ds ,$$

by [Rockafellar, 1970, 23.5], so $V(DP_N) \geq I_\phi(\hat{x}) - \delta\int_S [1 + |\hat{x}(s)|]ds$. However $V(DP_N) = V(P_N) \leq I_\phi(\hat{x})$, since $\hat{x}$ is feasible for (P_N) . Since δ was arbitrary, the result now follows. □

Notice that the strong duality theorem (3.3) is not in fact necessary to prove this result: it is sufficient to observe by weak duality that $V(DP_N) \leq V(P_N)$.

We are finally ready to deduce our main result. We include in the statement a summary of the above results.

<u>Theorem 3.7.</u> Suppose assumptions (3.1) hold for the problem (P_n) . Then $V(P_n) = V(DP_n) \uparrow I_\phi(\hat{x})$ as $n \to \infty$, with attainment in (DP_n) . If λ^n is is optimal for (DP_n) then $\sum_{i=1}^n \lambda_i^n a_i \to \phi'(\hat{x}(\cdot))$ weakly in $L_1(S)$. If there exists an optimal solution of (P_n) then it is given uniquely by $x_n(s) := (\phi^*)'\left(\sum_{i=1}^n \lambda_i^n a_i(s)\right)$, a.e. , and $\phi'(x_n(\cdot)) \to \phi'(\hat{x}(\cdot))$ weakly in $L_1(S)$.

Proof. Consider the function $g : L_1(S) \to (-\infty,\infty]$ defined by

$g(v) := I_{\phi^*}(v) - \langle v,\hat{x}\rangle$. By Proposition 2.2, $\inf g = -I_\phi(\hat{x})$, and is attained uniquely by $\bar{v} \in L_\infty(S)$, where $\bar{v}(s) := \phi'(\hat{x}(s))$ a.e. By Theorem 3.6, $\sum_{i=1}^{n} \lambda_i^n a_i$ is a minimizing sequence for g , so by Theorem 2.4 $\sum_{i=1}^{n} \lambda_i^n a_i \to \bar{v}$ weakly in $L_1(S)$. The remaining assertions follow from Theorem 3.5. □

Examples

Consider the special case

$$(E_n) \begin{cases} \inf & \int_0^1 \phi(x(s))ds \\ \text{subject to} & \int_0^1 s^i(x(s)-\hat{x}(s))ds = 0 , \quad i = 0,\ldots,n , \\ & 0 \le x \in L_1[0,1] , \end{cases}$$

where ds is Lebesgue measure, for three different measures of entropy:

$$\text{(i)} \qquad \phi(u) = \begin{cases} u\log u - u, & u > 0 , \\ 0 , & u = 0 , \\ +\infty , & u < 0 , \end{cases}$$

$$\text{(ii)} \qquad \phi(u) = \begin{cases} -\log u , & u > 0 , \\ +\infty , & u \le 0 , \end{cases}$$

$$\text{(iii)} \qquad \phi(u) = \begin{cases} u\log u - (1+u)\log(1+u) , & u > 0 , \\ 0 , & u = 0 , \\ +\infty , & u < 0 . \end{cases}$$

In all three cases (assuming $\hat{x}$ is continuous and strictly positive) Theorems 3.4 and 3.7 apply. Suppose in each case λ^n is dual optimal, and let x_n denote the unique optimal solution of (E_n) .

(i) $$x_n(s) = \exp\left(\sum_{i=0}^{n} \lambda_i^n s^i\right) , \text{ a.e.} ,$$

and $\log x_n(\cdot) \to \log \hat{x}(\cdot)$ weakly in L_1 .

(ii) $$x_n(s) = -\left(\sum_{i=0}^{n} \lambda_i^n s^i\right)^{-1} , \text{ a.e.} ,$$

and $\frac{1}{x_n(\cdot)} \to \frac{1}{\hat{x}(\cdot)}$ weakly in L_1 .

(iii) $$x_n(s) = \left[\exp\left(-\sum_{i=0}^{n} \lambda_i^n s^i\right) - 1\right]^{-1} , \text{ a.e.} ,$$

and $\log(1 + \frac{1}{x_n(\cdot)}) \to \log(1 + \frac{1}{\hat{x}(\cdot)})$ weakly in L_1 .

Further examples may be found in [Borwein and Lewis, 1988(a)]. Clearly we could replace $a_i(s) = s^{i-1}$ in the above with trigonometric polynomials, cos $i\theta$ and sin $i\theta$ (alternating).

In case (i) the theory in [Borwein and Lewis, 1988(b)] shows that in fact $x_n \to \hat{x}$ weakly in L_1 . Whether or not this is necessarily the case in the other examples remains unclear.

§4. References

Ben-Tal, A., J.M. Borwein and M. Teboulle, 1988: Spectral estimation via convex programming, to appear in the Proceedings of the conference in honour of A. Charnes' 70th Birthday.

Borwein, J.M. and A.S. Lewis, 1988(a): Duality relationships for entropy-like minimization problems, submitted to SIAM J. Appl. Math.

Borwein, J.M. and A.S. Lewis, 1988(b): On the convergence of moment problems, to appear.

Forte, B., W. Hughes and Z. Pales, 1988: Maximum entropy estimators and the problem of moments, to appear.

Goodrich, B.K. and A. Steinhardt, 1986: L_2 spectral estimation, SIAM J. Appl. Math. 46, 417-428.

Johnson, R.W. and J.E. Shore, 1984: Which is the better entropy for speech processing: slogs or logs? IEEE Trans. on acoustics, speech and signal processing, ASSP 32, 129-136.

Mead, L.R. and N. Papanicolaou, 1984: Maximum entropy in the problem of moments, J. Math. Phys. 25, 2404-2417.

Rockafellar, R.T., 1970: Convex analysis, Princeton University Press, Princeton, N.J.

Rockafellar, R.T., 1974: Conjugate duality and optimization, SIAM Publications, Philadelphia.

Rudin, W., 1966: Real and complex analysis, McGraw-Hill, New York.

Department of Mathematics, Statistics and Computing Science
Dalhousie University, Halifax, N.S., Canada B3H 3J5