JOURNAL OF APPROXIMATION THEORY 79, 1–16 (1994)

# Entropy Minimization with Lattice Bounds

JONATHAN M. BORWEIN,* ADRIAN S. LEWIS,† AND MARK A. LIMBER*

*Centre for Experimental and Constructive Mathematics, Department of Mathematics and Statistics, Simon Fraser University, Burnaby, British Columbia, V5A 1S6, Canada; and †Department of Combinatorics and Optimization, University of Waterloo, Waterloo, N2L 3G1 Ontario, Canada

Communicated by Frank Deutsch

Received March 20, 1992; accepted September 24, 1993

We characterize solutions to the problem of minimizing a convex integral objective function subject to a finite number of linear constraints and requiring that the feasible functions lie in a strip $[\alpha, \beta]$ where $\alpha$ and $\beta$ are extended real valued measurable functions. We use the duality theory of J. M. Borwein and A. S. Lewis (*Math. Programming*, *Series B* 57 (1992), 15–48, 49–84) to show that the solutions are of the usual form, but truncated where they leave the strip. © 1994 Academic Press, Inc.

## 1. INTRODUCTION

In this paper we investigate the entropy minimization problem with a finite set of linear constraints and lattice bounds $\alpha$ and $\beta$. More specifically, suppose $(K, \mu)$ is a complete finite measure space, $\alpha$ and $\beta$ two extended real valued measurable functions, $\alpha(t) < \beta(t)$ for almost all $t \in K$, $\{\psi_i\}_{i=1}^n$ a collection of functions in $L^\infty(K, \mu)$ and suppose $\phi$: $\mathbb{R} \to (-\infty, +\infty]$ is a lower semicontinuous essentially smooth proper function strictly convex on its domain with (necessarily smooth) finite conjugate $\phi^*$. We shall use the terminology of Rockafellar [17] throughout. We seek to characterize solutions in $L^1(K, \mu)$ to

$$\inf\left\{\int_K \phi(x(t))\, d\mu(t) \colon \alpha \le x \le \beta,\ \int_K x(t)\psi_i(t)\, d\mu(t) = b_i,\ i = 1, \ldots, n\right\}, \tag{P}$$

where $b = (b_1, \ldots, b_n)$ is a given vector in $\mathbb{R}^n$.



0021-9045/94 $6.00

The general concept of entropy maximization (or minimization as we have formulated it) can be found in a great variety of applied fields. Applications that fit into the form of (P) can be found, for example, in convex interpolation [8], crystallography [15], digital signal processing [1], and information theory [11].

Examples of functions $\phi$ that fit into this framework include

- Boltzmann–Shannon Entropy.

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

- Cost Entropy.

$$\phi(u) = \cosh(u) - 1.$$

- Fermi–Dirac Entropy.

$$\phi(u) = \begin{cases} (1-u)\log(1-u) + u\log(u) & \text{if } 0 < u < 1, \\ 0 & \text{if } u = 0 \text{ or } u = 1, \\ +\infty & \text{else}. \end{cases}$$

- $L^p$ Spectral Estimation ($1 < p < \infty$).

$$\phi(u) = \frac{1}{p}|u|^p.$$

The additional constraint imposed when $\alpha \equiv 0$ and $\beta \equiv +\infty$ in the $L^p$ spectral estimation problems has been studied and the solutions have been characterized by Ben-Tal, Borwein, and Teboulle [2], Goodrich, Roberts, and Steinhardt [9, 10], Micchelli, Smith, Swetits, and Ward [13], and others. Cole and Goodrich [7] characterize the solutions to the $L^p$ problem when $\alpha = 0$ and $0 < \varepsilon \leq \beta \in L^\infty$ with a Slater condition. Dontchev [8] and Limber and Goodrich [12] have characterized solutions to the $L^p$ estimation problem with general $\alpha$ and $\beta$. Borwein and Lewis [3–5] have studied general entropies, again with $\alpha \equiv 0$ and $\beta \equiv +\infty$. This work generalizes these results to include the entropies mentioned above and general measurable $\alpha$ and $\beta$.

## 2. Preliminary Definitions

In this section let $X$ be a linear normed space, $X^*$ its topological dual, and let $F\colon X \to (-\infty, +\infty]$ and $f\colon \mathbb{R} \to (-\infty, +\infty]$ be convex functions. We define some of the basic concepts of convex analysis and some of the conventions used in this paper.

The function $F$ is *proper* if it is somewhere finite, and the *domain* of $F$, dom $F$, is the set of all points $x$ where $F$ is finite. $f$ is *essentially smooth* provided $f$ is differentiable on the interior of its domain and for $s_f =$ sup dom $f$ and $i_f =$ inf dom $f$

$$\lim_{y \downarrow i_f} f'(y) = -\infty \qquad \text{if } i_f > -\infty$$

and

$$\lim_{y \uparrow s_f} f'(y) = +\infty \qquad \text{if } s_f < +\infty.$$

It is known that a differentiable convex function on $\mathbb{R}$ is continuously differentiable; see Rockafellar [17].

The set of *subgradients* of $F$ at $x_0 \in \operatorname{dom} F$ is defined to be

$$\partial F(x_0) = \{x^* \in X^*: \langle x - x_0, x^* \rangle \leq F(x) - F(x_0)\ \forall x \in X\}.$$

The *subgradient inequality* states that for any $x^* \in \partial F(x_0)$,

$$\langle x - x_0, x^* \rangle \leq F(x) - F(x_0). \tag{2.1}$$

Clearly, $x_0$ is a minimum of $F$ if and only if $0 \in \partial F(x_0)$. If $f$ is everywhere finite then $f$ is everywhere continuous and $\partial f(x) \neq \varnothing$ for all $x$ [17]. If $F$ is proper then the *convex conjugate* of $F$ is defined as the function $F^*$: $X^* \to (-\infty, +\infty]$ given by

$$F^*(x^*) = \sup_{x \in X} \{\langle x, x^* \rangle - F(x)\}.$$

Clearly, if $F \leq G$ then $F^* \geq G^*$. For $x \in X$ and $x^* \in X^*$

$$F^*(x^*) + F(x) \geq \langle x, x^* \rangle; \tag{2.2}$$

this is called the Fenchel–Young inequality [17].

For $a$ and $b$ extended real numbers we use the notation

$$a \vee b = \max\{a, b\} \qquad \text{and} \qquad a \wedge b = \min\{a, b\}.$$

We use $f'$ for the derivative of $f$ and $\nabla F$ for the Fréchet derivative of $F$.

With the assumed conditions on $\phi$, it follows that $\phi^*$ is differentiable (Rockafellar [17, Theorem 26.3]), and that

$$(\phi')^{-1} = (\phi^*)'. \tag{2.3}$$

Also, since $\phi^*$ is differentiable, it is continuously differentiable.

Recall that if $g: \mathbb{R} \times K \to (-\infty, +\infty]$ is a *normal convex integrand* and $(K, \mu)$ is a complete finite measure space then

$$G(x) = \int_K g(x(t), t)\, d\mu(t)$$

takes on well defined values in $(-\infty, +\infty]$. In particular, if $g(u, t)$ is convex and lower semicontinuous in $u \in \mathbb{R}$ for almost all $t$, and measurable in $t$ for each $u$ with the set

$$\{u \in \mathbb{R}: g(u, t) < +\infty\}$$

having nonempty interior, then $g$ is a normal convex integrand. For example, see Rockafellar [16, Lemma 2].

For a convex set $C$ and a point $x \in C$ we define $\overline{\text{cone}}(C - x)$ to be the closed convex cone generated by the set $C - x$ with vertex at the origin

$$\overline{\text{cone}}(C - x) = \overline{\{r(c - x): r \geq 0, c \in C\}}.$$

The *quasi relative interior* [4] of a convex set $C \subset X$ is

$$\text{qri}(C) = \{x \in C: \overline{\text{cone}}(C - x) \text{ is a subspace}\}.$$

For example, if $C = \{x \in L^1(K, \mu): 0 \leq x \text{ a.e.}\}$ then it can be shown that $\text{qri}(C) = \{x \in X: 0 < x \text{ a.e.}\}$. In any locally convex space, $x_0 \in \text{qri}(C)$ whenever $x_0$ is not a proper support of $C$: in $L^1(K, \mu)$ this amounts to saying that if $y \in L^\infty(K, \mu)$ and $\int_K (x - x_0) y\, d\mu \geq 0$ for all $x \in C$ then equality holds for all $x \in C$. When the affine span of $C$ is dense this is equivalent to $x_0 \in \text{qi}(C)$ in the sense of Limber and Goodrich [12]. In [4] it is shown that for a convex set $C$ with nonempty quasi relative interior and a continuous linear map $A: X \to \mathbb{R}^n$,

$$\text{ri}(A(C)) = A(\text{qri}(C)). \tag{2.4}$$

This fact will be crucial in developing our theory.

Define the continuous linear map $A: L^1 \to \mathbb{R}^n$ by

$$(Ax)_k = \int_K x(t) \psi_k(t)\, d\mu(t).$$

The adjoint of $A$, $A^*: \mathbb{R}^n \to L^\infty$ is given by

$$A^*(\lambda) = \sum_{i=1}^{n} \lambda_i \psi_i.$$

Also, for $x \in L^1 \subset (L^\infty)^*$, $A^{**}x = Ax$ since for any $\lambda \in \mathbb{R}^n$, $\langle \lambda, A^{**}x \rangle = \langle A^*\lambda, x \rangle = \langle \lambda, Ax \rangle$.

Finally, we shall use $x, y, z$ to denote functions in $L^1(K, \mu)$ or $L^\infty(K, \mu)$, $\lambda \in \mathbb{R}^n$ and $u$, $v$, and $w$ for real numbers.

## 3. The Integral Functional

In this section we introduce our integral objective function $F$, and show that its conjugate $F^*$ is Fréchet differentiable.

We begin by studying the restriction $\phi_I$ of $\phi$ to an interval $I = [\alpha_0, \beta_0]$ where $\alpha_0$ and $\beta_0$ are extended real numbers, $[\alpha_0, \beta_0] \cap \operatorname{dom} \phi \neq \varnothing$, and

$$\phi_I(u) = \phi(u) + i_I(u),$$

where

$$i_I(u) = \begin{cases} 0 & \text{if } u \in I \\ +\infty & \text{if } u \notin I \end{cases}$$

is the indicator function of $I$.

Lemma 3.1. *For all $u \in \mathbb{R}$, the derivative of $(\phi_I)^*$, $[(\phi_I)^*]'$ exists and is given by*

$$[(\phi_I)^*]'(u) = \alpha_0 \vee (\phi^*)'(u) \wedge \beta_0.$$

*Proof.* Since $\phi \leq \phi_I$, $(\phi_I)^* \leq \phi^*$, and so $(\phi_I)^*$ is everywhere finite and $\partial(\phi_I)^*(v) \neq \varnothing$ for all $v \in \mathbb{R}$. Now, since $\phi$ is lower semicontinuous, $w \in \partial(\phi_I)^*(v)$ if and only if $v \in \partial\phi_I(w)$ and since $\phi$ is essentially smooth, $w \in \operatorname{int}(\operatorname{dom} \phi) \cap [\alpha_0, \beta_0]$. Then we have

$$v \in \begin{cases} (-\infty, \phi'(\alpha_0)] & \text{if } w = \alpha_0 > -\infty, \\ \{\phi'(w)\} & \text{if } \alpha_0 < w < \beta_0, \\ [\phi'(\beta_0), +\infty) & \text{if } w = \beta_0 < +\infty. \end{cases}$$

By Eq. (2.3), $(\phi')^{-1} = (\phi^*)'$, and since $\phi'$ is strictly increasing we can solve this equation to get

$$w = \begin{cases} \alpha_0 & \text{if } v \leq \phi'(\alpha_0), \\ (\phi^*)'(v) & \text{if } \phi'(\alpha_0) < v < \phi'(\beta_0), \\ \beta_0 & \text{if } v \geq \phi'(\beta_0). \end{cases}$$

By [17, Theorem 25.1], this completes the proof. ∎

We now consider the parameterized interval $I(t) = [\alpha(t), \beta(t)]$. Throughout the rest of this paper we shall assume:

*Assumption* A.

$$\exists \hat{x} \in L^1(K, \mu) \text{ such that } \alpha \leq \hat{x} \leq \beta \text{ a.e. and } \int_K \phi(\hat{x}(t))\, d\mu(t) \in \mathbb{R}, \tag{A1}$$

$$\operatorname{int}(\operatorname{dom} \phi) \cap (\alpha(t), \beta(t)) \neq \varnothing \qquad \text{a.e.} \tag{A2}$$

LEMMA 3.2. *Let* $\hat{\phi}$: $\mathbb{R} \times K \to (-\infty, +\infty]$ *be defined by* $\hat{\phi}(u, t) = \phi(u) + i_{I(t)}(u)$. *Then* $\hat{\phi}$ *is a normal convex integrand.*

*Proof.* The conditions of normality are easily checked. To see that

$$\{u \in \mathbb{R}: \hat{\phi}(u, t) < +\infty\}$$

has nonempty interior a.e., we use the assumption (A2). ∎

We define our objective integral function $F$: $L^1(K, \mu) \to (-\infty, +\infty]$ by

$$F(x) \triangleq \int_K \phi_{I(t)}(x(t))\, d\mu(t) \tag{3.1}$$

which is equivalent to

$$F(x) = \begin{cases} \int_K \phi(x(t))\, d\mu(t) & \text{if } \alpha(t) \leq x(t) \leq \beta(t) \text{ a.e.} \\ +\infty & \text{else.} \end{cases}$$

Also,

$$F(x) = \int_K \hat{\phi}(x(t), t)\, d\mu(t)$$

and (A1) says dom $F \neq \varnothing$.

LEMMA 3.3. *For* $y \in L^\infty$ *the conjugate of* $F$ *at* $y$ *is*

$$F^*(y) = \int_K (\phi_{I(t)})^*(y(t))\, d\mu(t).$$

*Also*, $F^{**} = F$.

*Proof.* We can apply Theorem 2 in Rockafella [16] once we verify that $\hat{\phi}$ is a normal convex integrand and that $\int_K \phi_I^*(y)\, d\mu$ is finite for some $y \in L^\infty$. We known $\hat{\phi}$ is a normal convex integrand from Lemma 3.2. Now

$\phi_I \geq \phi$ implies $(\phi_I)^* \leq \phi^*$, so for any $y \in L^\infty(K, \mu)$,

$$\int_K (\phi_I)^*(y) \leq \int_K \phi^*(y) \in \mathbb{R}$$

since $\phi^*$ is continuous and $\phi^*(y) \in L^\infty \subset L^1$. Therefore,

$$\int_K (\phi_I)^*(y) < +\infty.$$

By the Fenchel–Young inequality given in Eq. (2.2),

$$(\phi_I)^*(y(t)) + \phi_I(\hat{x}(t)) \geq \hat{x}(t)y(t) \qquad \text{a.e.}$$

for $\hat{x}$ as in (A), and thus

$$\int_K (\phi_I)^*(y)\,d\mu \geq \int_K \hat{x}y\,d\mu - \int_K \phi_I(\hat{x})\,d\mu \in \mathbb{R}$$

so $\int_K (\phi_I)^*(y) \in \mathbb{R}$. In fact, $\int_K (\phi_I)^*(y)$ is finite for any $y \in L^\infty$, a little more than we needed. The last statement follows from Theorem 2 in [16]. ∎

Lemma 3.4. *Let $x \in L^\infty(K, \mu)$ and define*

$$y_1(t) \triangleq \alpha(t) \vee (\phi^*)'(x(t)) \wedge \beta(t).$$

*Then $y_1 \in L^1$.*

*Proof.* Let $y(\cdot) \triangleq (\phi^*)'(x(\cdot))$. Then $y \in L^\infty$ since $x \in L^\infty$ and $(\phi^*)'$ is continuous. Define the measurable sets

$$T_\alpha = \{t \in K: y(t) \leq \alpha(t)\},$$
$$T_\beta = \{t \in K: y(t) \geq \beta(t)\},$$

and

$$T_\gamma = \{t \in K: \alpha(t) \leq y(t) \leq \beta(t)\}.$$

Then for $\hat{x}$ as in assumption (A1),

$$y(t) \leq \alpha(t) = y_1(t) \leq \hat{x}(t) \qquad \text{on } T_\alpha,$$
$$\hat{x}(t) \leq \beta(t) = y_1(t) \leq y(t) \qquad \text{on } T_\beta,$$

and

$$y_1(t) = y(t) \qquad \text{on } T_\gamma.$$

Since $\hat{x}, y \in L^1$, it follows that $y_1 \in L^1$. ∎

We can now verify that $F^*$ is Fréchet differentiable.

THEOREM 3.5. *The conjugate $F^*$ of $F$ is Fréchet differentiable at every $x \in L^\infty$. In fact, for $x, h \in L^\infty$,*

$$\langle \nabla F^*(x), h \rangle = \int_K [\alpha(t) \vee (\phi^*)'(x(t)) \wedge \beta(t)] h(t)\, d\mu(t).$$

*Proof.* Fix $h$ and $x \in L^\infty(K, \mu)$. Let $\|h\|$ denote $\|h\|_\infty$ and

$$T(x)(t) = \alpha(t) \vee (\phi^*)'(x(t)) \wedge \beta(t),$$

then $T(x) \in L^1$ by Lemma 3.4. Consider the Fréchet quotient

$$\begin{aligned}\frac{1}{\|h\|}&|F^*(x+h) - F^*(x) - \langle T(x), h\rangle| \\ &\leq \frac{1}{\|h\|}\int_K |(\phi_I)^*(x+h) - (\phi_I)^*(x) - T(x)h|\, d\mu. \qquad (3.2)\end{aligned}$$

We fix $t$ and study the integrand. By Lemma 3.4,

$$\alpha(t) \vee (\phi^*)'(x(t) + \|h\|) \wedge \beta(t) \in L^1,$$

and

$$\alpha(t) \vee (\phi^*)'(x(t) - \|h\|) \wedge \beta(t) \in L^1.$$

Since $(\phi_I)^*$ is differentiable, by the mean value theorem for functions of a single variable there is a $\theta_t \in (0,1)$ such that

$$\begin{aligned}&|(\phi_{I(t)})^*(x(t) + h(t)) - (\phi_{I(t)})^*(x(t)) - T(x)(t)h(t)| \\ &\quad = |[(\phi^*_{I(t)})'(x(t) + \theta_t h(t)) - T(x)(t)]h(t)| \\ &\quad \leq \|h\|((\alpha(t) \vee (\phi^*)'(x(t) + \|h\|) \wedge \beta(t)) \\ &\qquad\qquad - (\alpha(t) \vee (\phi^*)'(x(t) - \|h\|) \wedge \beta(t)))\end{aligned}$$

since $(\phi^*)'$ is increasing, and

$$\begin{aligned}&(\alpha(t) \vee (\phi^*)'(x(t) + \|h\|) \wedge \beta(t)) \\ &\qquad - (\alpha(t) \vee (\phi^*)'(x(t) - \|h\|) \wedge \beta(t)) \downarrow 0 \qquad \text{a.e.}\end{aligned}$$

as $\|h\| \downarrow 0$ since $(\phi^*)'$ is continuous and increasing. By the monotone convergence theorem applied to the above equation

$$\lim_{\|h\|\to 0}\int_K (\alpha(t) \vee (\phi^*)'(x(t) + \|h\|) \wedge \beta(t))$$
$$-(\alpha(t) \vee (\phi^*)'(x(t) - \|h\|) \wedge \beta(t))\, d\mu(t)$$
$$= 0,$$

and therefore from Eq. (3.2) we see that the Fréchet quotient goes to zero as desired. ∎

## 4. The General Form of the Solution

We consider the following primal problem where $F$, $A$, and $b$ are as defined above and we assume Assumption (A). The value of the problem is defined as

$$V(P) \triangleq \inf\{F(x)\colon Ax = b\}. \tag{$P$}$$

In this section we characterize solutions to $(P)$.

From Borwein and Lewis [4] we extract the following Fenchel duality theorem.

Theorem 4.1. *Suppose $V(P)$ is finite in problem $(P)$, and $b \in \operatorname{ri}(A(\operatorname{dom} F))$. If*

$$V(D) \triangleq \sup\{\langle b, \lambda\rangle - F^*(A^*\lambda)\colon \lambda \in \mathbb{R}^n\} \tag{$D$}$$

*then $V(P) = V(D)$ where the supremum is attained at some $\bar{\lambda} \in \mathbb{R}^n$.*

To interpret the constraint qualification $b \in \operatorname{ri}(A(\operatorname{dom} F))$ in our context, we use Eq. (2.4) to see that we need only check that there is a feasible point in the quasi relative interior of the domain of $F$, i.e., that there is a $\hat{x} \in \operatorname{qri}(\operatorname{dom} F)$ such that $A\hat{x} = b$. (See [4].) Thus we characterize $\operatorname{qri}(\operatorname{dom} F)$.

Theorem 4.2. *Define $\iota \triangleq \inf \operatorname{dom} \phi$ and $\sigma \triangleq \sup \operatorname{dom} \phi$. Then*

$$\operatorname{qri}(\operatorname{dom} F) = \left\{x \in L^1\colon \int_K \phi(x) \in \mathbb{R}, (\alpha \vee \iota) < x < (\beta \wedge \sigma) \text{ a.e.}\right\}. \tag{4.1}$$

*Proof.* Define $M$ to be the right hand side of Eq. (4.1) and let $x_0 \in M$. We assume $x_0$ is a support point of $M$ and generate a contradiction.

Suppose $y \in L^\infty$ is nonzero and satisfies

$$\int_K (x - x_0) y \, d\mu \geq 0 \qquad \forall x \in \operatorname{dom} F, \tag{4.2}$$

that is, $y$ is a support functional at $x_0$. Without loss of generality, let $T$ be a measurable subset of $K$ with $y(t) < 0$ a.e. on $T$. Since $x_0 \in M$, there is a compact interval in $\mathbb{R}$, $[u, v] \subseteq (\iota, \sigma)$ such that

$$T' \triangleq \{t \in T \colon x_0(t) \in [u, v] \text{ a.e.}\}$$

has positive measure. Pick $\varepsilon > 0$ so that $v + \varepsilon < \sigma$, and define $z \in L^1$ by

$$z(t) = \begin{cases} \varepsilon \wedge (\beta(t) - x_0(t)) & \text{if } t \in T', \\ 0 & \text{otherwise}. \end{cases}$$

Then $x_0(t) + z(t) \leq \beta(t)$ a.e. on $K$.

Since $x_0 + z \in L^\infty(T', \mu)$ and $\phi$ is continuous, it then follows that $x_0 + z \in \operatorname{dom} F$ and

$$\int_K [(x_0(t) + z(t)) - x_0(t)] y(t) \, d\mu(t)$$

$$= \int_{T'} [\varepsilon \wedge (\beta(t) - x_0(t))] y(t) \, d\mu(t) < 0$$

since $\varepsilon \wedge (\beta(t) - x_0(t)) > 0$ and $y(t) < 0$ a.e. on $T'$. But this contradicts Eq. (4.2).

The other inclusion is easier. If $x = \beta \wedge \sigma$ on $E$ a set of positive measure, then $\chi_E$, the characteristic function of $E$, is a support functional in $L^\infty$ at $x$ of dom $F$, thus $x \notin \operatorname{qri}(\operatorname{dom} F)$. Similar results are obtained if $x = \iota \vee \alpha$ on a set $F$ of positive measure. ∎

Using Eq. (2.4) and Theorems 4.1 and 4.2 we have the following constraint qualification for problem $(P)$:

$$\exists \hat{x} \text{ such that } A\hat{x} = b, \qquad \int_K \phi(\hat{x}) \, d\mu \in \mathbb{R}, \ (\alpha \vee \iota) < \hat{x} < (\beta \wedge \sigma) \text{ a.e.} \tag{CQ}$$

In the special case where $\alpha$ and $\beta$ are constants, this constraint qualification is equivalent to

$$\exists \hat{x} \text{ such that } A\hat{x} = b, \qquad \alpha \vee \iota < \hat{x}(t) < \beta \wedge \sigma \text{ a.e.} \tag{CQ'}$$

See Borwein and Lewis [5, Lemma 4.5]. Also note that (CQ) includes Assumption (A).

If $\bar{\lambda}$ solves the dual problem $(D)$, then since $F^*$ is differentiable we can differentiate the maximum problem $(D)$ with respect to $\lambda$ and apply Theorem 19 in [18] to get

$$b = A^{**}\big(\nabla F^*(A^*\bar{\lambda})\big) = A\big(\nabla F^*(A^*\bar{\lambda})\big) \tag{4.3}$$

since $\nabla F^*(A^*\bar{\lambda}) \in L^1$ by Lemma 3.4. If we let $\bar{x} = \nabla F^*(A^*\bar{\lambda})$ then we see from (4.9) that $A\bar{x} = b$, so $\bar{x}$ is feasible. Also, from Theorem 3.5 we have

$$\bar{x} = \alpha \vee (\phi^*)'(A^*\bar{\lambda}) \wedge \beta \qquad \text{for some } \bar{\lambda} \in \mathbb{R}^n \tag{4.4}$$

and so $\bar{x}$ lies in the strip $[\alpha, \beta]$. Finally, let $x$ be any other function such that $Ax = b$. Since $F^{**} = F$, by Rockafella [18, Corollary 12A], $A^*\bar{\lambda} \in \partial F(\bar{x})$, so $\bar{x} \in \operatorname{dom} F$ and by the subgradient inequality (2.1),

$$\langle x - \bar{x}, A^*\bar{\lambda}\rangle \le F(x) - F(\bar{x}).$$

Since $Ax = b$ and $A\bar{x} = b$, $\bar{x}$ solves the minimization problem $(P)$. Thus, the form of the solution to $(P)$ is given in Eq. (4.4). Compare [4, Part I, Corollary 4.10].

We have proven the following theorem.

THEOREM 4.3. *Let $\phi$: $\mathbb{R} \to (-\infty, +\infty]$ be a proper, lower semicontinuous, essentially smooth function strictly convex on its domain with finite conjugate $\phi^*$. Let $(K, \mu)$ be a complete finite measure space, let $\alpha$ and $\beta$ be extended real valued measurable functions on $K$, $\alpha(t) < \beta(t)$ a.e., and let $A$: $L^1 \to \mathbb{R}^n$ be the continuous linear map given by $(Ax)_j = \int_K x(t)\psi_j(t)\,d\mu(t)$ for the $L^\infty$ functions $\{\psi_j\}_{j=1}^n$. If $b \in \mathbb{R}^n$ is a fixed vector and*

$$p = \inf\left\{\int_K \phi(x(t))\,d\mu(t) \colon \alpha \le x \le \beta,\ Ax = b,\ x \in L^1(K, \mu)\right\}$$

*is finite, then provided* (CQ) *holds the unique solution to $(P)$ is of the form*

$$\bar{x} = \alpha \vee (\phi^*)'(A^*\bar{\lambda}) \wedge \beta,$$

*where $\bar{\lambda} \in \mathbb{R}^n$ is any solution of*

$$\sup\left\{\langle b, \lambda\rangle - \int_K (\phi_t)^*(A^*\lambda(t))\,d\mu(t) \colon \lambda \in \mathbb{R}^n\right\}$$

*and the supremum is attained.*

## 5. Examples

We revisit the examples in the Introduction and characterize the solutions to $(P)$ for each $\phi$.

1. *Boltzmann–Shannon.* In this case the conjugate of $\phi$ is

$$\phi^*(v) = \exp(v - 1),$$

clearly a differentiable finite function on $\mathbb{R}$. The solution to the corresponding entropy problem is then

$$\bar{x}(t) = \alpha(t) \vee \exp\big(A^*\bar{\lambda}(t) - 1\big) \wedge \beta(t).$$

2. Cosh *Entropy.* The conjugate is

$$\phi^*(v) = v \operatorname{arcsinh}(v) - \sqrt{v^2 + 1} + 1,$$

and the solution to the corresponding entropy problem is

$$\bar{x}(t) = \alpha(t) \vee \operatorname{arcsinh}\big(A^*\bar{\lambda}(t)\big) \wedge \beta(t).$$

In this example if we start with

$$\phi(u) = u \operatorname{arcsinh}(u) - \sqrt{u^2 + 1} + 1,$$

then $\phi^*(v) = \cosh(v) - 1$, so we could solve this entropy problem with solution

$$\bar{x}(t) = \alpha(t) \vee \sinh\big(A^*\bar{\lambda}(t)\big) \wedge \beta(t).$$

This example is something of a two-sided Boltzmann–Shannon type entropy.

3. *Fermi–Dirac.* The conjugate is

$$\phi^*(v) = \ln(1 + \exp(v)),$$

and the solution to the corresponding entropy problem is

$$\bar{x}(t) = \alpha(t) \vee \frac{\exp\big(A^*\bar{\lambda}(t)\big)}{1 + \exp\big(A^*\bar{\lambda}(t)\big)} \wedge \beta(t).$$

4. *$L^p$ Spectral Estimation.* The conjugate is

$$\phi^*(v) = \frac{1}{q}|v|^q,$$

where $(1-q)(1-p) = 1$. The solution to the corresponding entropy problem is

$$\bar{x}(t) = \alpha(t) \vee |A^*\bar{\lambda}(t)|^{q-1} \operatorname{sign}(A^*\bar{\lambda}(t)) \wedge \beta(t).$$

Now, for particular choices of $\alpha$ and $\beta$ we have some results from the literature.

5. *Positive $L^2$ Spectral Estimation.* Consider the problem in $L^2(K, \mu)$,

$$\inf\left\{\tfrac{1}{2}\|x\|_2^2 \colon Ax = b, 0 \le x\right\},$$

as studied in [9, 10, 13, 14] among others. With the choices $\alpha \equiv 0$ and $\beta \equiv +\infty$ and $\phi(u) = |u|^2/2$ we have the same problem, and the solution is of the form

$$\bar{x}(t) = \max\{0, A^*\bar{\lambda}\} = (A^*\bar{\lambda})^+$$

for some $\bar{\lambda} \in \mathbb{R}^n$. That is, the solution is a truncated linear function in the $\psi_i$'s.

6. *$L^p$ Spectral Estimation with an Upper Bound.* Cole and Goodrich [7] have characterized the solution to the following problem in $L^p(K, \mu)$. Let $\beta \in L^\infty$ and suppose there is a feasible function $\hat{x}$ and an $\varepsilon > 0$ such that $\varepsilon \le \hat{x} \le \beta - \varepsilon$. Then they characterize the solution to

$$\inf\left\{\frac{1}{p}\|x\|_p^p \colon Ax = b, 0 \le x \le \beta\right\}$$

via a standard Lagrange multiplier theorem in $L^\infty$. If we let $\alpha \equiv 0$ then our theory shows the solution is

$$\bar{x}(t) = \max\left\{0, \min\left\{(A^*\bar{\lambda}(t))^{q-1}, \beta(t)\right\}\right\}. \tag{5.1}$$

In fact, we may strengthen this result to only require an $\hat{x} \in L^p$ such that $0 < \hat{x} < \beta$ where $\beta$ is now any $L^p$ function strictly positive on $K$. The solution is again of the form in (5.1). Limber and Goodrich [12] have shown that a generalization of the Lagrange multiplier theorem employed in [7] is valid and can be used to derive this same result.

7. *$L^p$ Spectral Estimation with Lattice Bounds.* More generally, Dontchev [8] and Limber and Goodrich [12] have shown that in the case $\alpha$ and $\beta$ are any extended valued measurable functions, the solution to the $L^p$ problem is of the form

$$\bar{x}(t) = \max\left\{\alpha(t), \min\left\{\left(A^*\bar{\lambda}(t)\right)^{q-1} \operatorname{sign}\left(A^*\bar{\lambda}(t)\right), B(t)\right\}\right\};$$

a power of a linear function in the $\psi_i$'s truncated to fit in the strip $[\alpha, \beta]$. This includes the following special case: let $S_1$ and $S_2$ be two subsets of $K$ and define

$$\alpha(t) = \begin{cases} -\infty & \text{if } t \in S_1 \\ 0 & \text{if } t \notin S_1, \end{cases} \qquad \beta(t) = \begin{cases} +\infty & \text{if } t \in S_2 \\ 0 & \text{if } t \notin S_2. \end{cases}$$

This effectively restricts the feasible functions to be nonnegative on $S_1^c$, nonpositive on $S_2^c$, and unrestricted on $S_1 \in S_2$. In one case the original problem is from convex interpolation and we are minimizing $\|y''\|_2$ (where we make the change of variable $x \triangleq y''$) and we are forcing $y$ to be convex, concave, and unrestricted on $S_1^c$, $S_2^c$, and $S_1 \cap S_2$, respectively.

If $S_0 \equiv S_1^c \cap S_2^c \neq \varnothing$ then $\alpha \not< \beta$, but on $S_0$ all feasible functions $x$ must be zero. Thus we can redefine our measure space to be $K \setminus S_0$ and adjust our linear constraints accordingly. This is possible in general whenever $\alpha = \beta$ on a set of positive measure, so if the assumption (A2) fails, either the problem is infeasible, or we may adjust our constraints.

These last two examples illustrate the utility in allowing the functions $\alpha$ and $\beta$ to be extended valued.

8. *Burg Entropy.* The Burg [6] entropy is another popular entropy functional. In this case

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

and the corresponding conjugate is

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

which does not satisfy our hypothesis that $\phi^*$ is everywhere finite. Following the results in Borwein and Lewis [5], if a solution exists in $L^1$, it should be of the form

$$\bar{x}(t) = \alpha(t) \vee \frac{1}{A^*\bar{\lambda}(t)} \wedge \beta(t). \tag{5.2}$$

However, it is possible that a solution will not exist in $L^1$, but rather in $(L^\infty)^*$. In this case the piece of the solution that is absolutely continuous with respect to $\mu$ will be of the form in Eq. (5.2).

## 6. Conclusions

We have shown that the general solution to the minimum entropy problem with lattice bounds and a finite number of constraints can be characterized with the duality theory of Borwein and Lewis [4]. The interval constraints include a great many applications, for example, shape preserving interpolation and bandwidth limited spectral estimation.

It should also be mentioned that for most choices of $\alpha$ and $\beta$ the solutions can be easily computed numerically by solving the unconstrained finite dimensional dual problem with an unconstrained minimization technique such as Newton's method.

## References

1. C. Auyeung and R. M. Mersereau, A dual approach to signal restoration, *in* "IEEE International Conference on Acoustics, Speech, and Signal Processing," Glasgow, 1989, pp. 1326–1329.
2. A. Ben-Tal, J. M. Borwein, and M. Teboulle, A dual approach to multidimensional $L^p$ spectral estimation problems, *SIAM J. Control Optim.* **26**, No. 4 (1988), 985–996.
3. J. M. Borwein and A. S. Lewis, Duality relationships for entropy-like minimization problems, *SIAM J. Control Optim.* **29**, No. 2 (1991), 325–338.
4. J. M. Borwein and A. S. Lewis, Partially-finite convex programming, Parts I and II, *Math. Programming, Series B* **57** (1992), 15–48, 49–84.
5. J. M. Borwein and A. S. Lewis, Partially-finite programming in $L_1$: Entropy maximization, *SIAM J. Optimization*, **3** (1993), 248–267.
6. J. P. Burg, Maximum entropy spectral analysis, paper presented at the 37th meeting of the Societv of Exploration Geophysicists, Oklahoma City, 1967.
7. R. Cole and R. K. Goodrich, $L_p$ spectral analysis with an $L_\infty$ upper bound, *J. Optim. Theory Appl.*, **76**, No. 2 (1993) 321–355.
8. A. L. Dontchev, Duality methods for constrained best interpolation, *Math. Balkanica* **1** (1987).
9. R. K. Goodrich and A. Steinhardt, $L_2$ spectral estimation, *SIAM J. Appl. Math.* **46** (1986), 417–426.
10. R. K. Goodrich, A. Steinhardt, and R. Roberts, Spectral estimation via minimum energy correlation extension, *IEEE Trans. Acoust. Speech Signal Process.* **33**, No. 6 (1985).
11. E. T. Jaynes, Information theory and statistical mechanics, *Phys. Rev.* **106** (1957), 620–630; *Phys. Rev.* **108** (1957), 171–190.
12. M. A. Limber and R. K. Goodrich, Quasi interiors, Lagrange multipliers and $L^p$ spectral estimation with lattice bounds, *J. Optim. Theory Appl.*, **78**, No. 1 (1993).

13. C. A. MICCHELLI, P. W. SMITH, J. SWETITS, AND J. D. WARD, Constrained $L_p$ approximation, *Constr. Approx.* (1985), 93–102.
14. C. A. MICCHELLI AND F. I. UTRERAS, Smoothing and interpolation in a convex subset of a Hilbert space, *SIAM J. Sci. Statist. Comput.* **9**, No. 4 (1988), 728–746.
15. J. NAVAZA, The use of non-local constraints in maximum-entropy electron density reconstruction, *Acta Cryst.* **42** (1986), 212–223.
16. R. T. ROCKAFELLAR, Integrals which are convex functionals, *Pacific J. Math.* **24**, No. 3 (1968).
17. R. T. ROCKAFELLAR, "Convex Analysis," Princeton Univ. Press, Princeton, NJ, 1970.
18. R. T. ROCKAFELLAR, "Conjugate Duality and Optimization," SIAM, Philadelphia, PA, 1974.