PARTIALLY-FINITE PROGRAMMING IN L_1 AND THE EXISTENCE OF MAXIMUM ENTROPY ESTIMATES*

J. M. BORWEIN† AND A. S. LEWIS†

Abstract. Best entropy estimation is a technique that has been widely applied in many areas of science. It consists of estimating an unknown density from some of its moments by maximizing some measure of the entropy of the estimate. This problem can be modelled as a partially-finite convex program, with an integrable function as the variable. A complete duality and existence theory is developed for this problem and for an associated extended problem which allows singular, measure-theoretic solutions. This theory explains the appearance of singular components observed in the literature when the Burg entropy is used. It also provides a unified treatment of existence conditions when the Burg, Boltzmann–Shannon, or some other entropy is used as the objective. Some examples are discussed.

Key words. convex analysis, duality, existence, generalized solution, image reconstruction, maximum entropy method, moment problem, partially finite program, spectral estimation

AMS subject classifications. primary 49A55, 90C25; secondary 65K05, 49B27

1. Introduction: Best entropy estimation. A very common problem in many areas of the physical sciences consists of trying to estimate an unknown density by measuring some of its moments. More precisely, given a number of integrals of an unknown function with respect to known weight functions, and a real interval in which the function is known to take its values, we seek to estimate the function. Typically, the weight functions are trigonometric polynomials, frequently multidimensional (so the given moments are Fourier coefficients), or algebraic polynomials (giving power moments), and the given interval is often (though not exclusively) the nonnegative reals.

Given only a finite number of moments this estimation problem is clearly under-determined. One extremely popular method for selecting an estimate from the family of all functions satisfying the prescribed moment constraints is to choose it to minimize some objective functional (subject to the given constraints). This objective is typically some measure of entropy—hence the term "best entropy estimation." This approach has been widely and successfully used in such diverse areas as astronomy, crystallography, speech processing, tomography, geophysics, and many others. For surveys, see [31] and [35] (containing in total almost 700 references), and the recent collections, [54], [53], [17], and [51].

Phrased mathematically, the best entropy estimation problem becomes, in its simplest form,

(1.1)
$$\text{minimize} \quad \int \phi(x(s))$$
$$\text{subject to} \quad \int a_i x = b_i \quad \text{for } i = 1, \ldots, n.$$

The variable density to be chosen is x, the a_i's are the known weight functions, and the b_i's are the measured moments. The function ϕ reflects our choice of entropy: it may take the value $+\infty$ to incorporate the known range constraint on x. For reasons

discussed in [6] we may as well restrict ourselves to closed, proper, convex functions ϕ. The two classical choices correspond to the Boltzmann–Shannon entropy, perhaps first suggested in this context in [27],

$$(1.2) \qquad \phi(u) := \begin{cases} u \log u & \text{if } u > 0, \\ 0 & \text{if } u = 0, \\ +\infty & \text{if } u < 0, \end{cases}$$

and the Burg entropy, first proposed in [12],

$$(1.3) \qquad \phi(u) := \begin{cases} -\log u & \text{if } u > 0, \\ 0 & \text{if } u \leqq 0, \end{cases}$$

although numerous other entropies have appeared in the literature, including L_2 and L_r entropies [25], [29], and [3], and the general families proposed in [40], [39], and [13].

The debate over the relative merits of the various entropies has been intense, as the above references will testify. The choice between (1.2) and (1.3) has been particularly controversial (see, for example, [28] and [52]). The issues in this debate can be grouped into three rather distinct areas. The first might be termed a priori reasons for selecting a particular entropy, generally involving a probabilistic, statistical, or information-theoretic discussion of the underlying phenomenon we seek to measure (see for example, [40], [52], [28], [39], and [13]). The second area of debate is empirical: the performance of the method is judged by its ability to reconstruct a known density from its moments (see, for example, [40], [28], [52], and [29]). Both of these areas lie outside our current scope.

The third area might be called a posteriori reasons: mathematical properties of the estimates arising from a particular choice of entropy are studied. Two particular properties have attracted attention: the existence of the optimal estimates, and their convergence to the underlying density as the number of given moments grows. For questions of convergence, see [52], [37], [22], [50], [18], [34], [19], [13], [7], [5], and [11]. In this paper we shall concentrate on the first property: the existence of an optimal solution for the estimation problem (1.1).

The basic idea for solving (1.1) has been explained widely in the applied literature, although for the most part without any degree of rigour: the form of the optimal solution is derived by attaching Lagrange multipliers $\lambda_1, \dots, \lambda_n$ to the constraints, and then differentiating (formally), giving

$$(1.4) \qquad \bar{x}(s) := (\phi')^{-1}\left(\sum_{1}^{n} \lambda_i a_i(s)\right),$$

where the λ_i's are chosen to ensure that $\bar{x}$ is feasible. Two existence questions need to be addressed to make this rigorous. First, when do the multipliers $\lambda_1, \dots, \lambda_n$ exist? Put differently, we require the existence of an optimal solution to the dual problem for (1.1). As usual in convex programming, the required condition is a primal constraint qualification for (1.1). This is straightforward to check: a general theory for "partially-finite programs" (convex programs with an infinite-dimensional variable subject to a finite number of linear constraints) is developed in [9] and [6].

The second question is more delicate: when does (1.4) give the optimal solution? Under mild conditions, it does so provided that we know a priori that an optimal solution exists. This is the case, for example, when the objective function has weakly compact level sets, as is the case with the Boltzmann–Shannon entropy [7], but the important case of the Burg entropy is not covered by this idea. Existence was shown

for important special cases in [15] and [56], and a general condition ensuring existence was introduced in [32] together with a demonstration that it may fail in general.

A fascinating concrete example of the nonexistence of a best Burg entropy estimate appeared in [40] (see also [52] and [14]). The problem was very simple: the unknown function was a probability density on the unit cube in $\mathbb{R}^3$, with three of its (multi-dimensional) Fourier coefficients given equal to a parameter α in [0, 1). It turns out that the Lagrange multipliers always exist, and, at least for small α, (1.4) gives the correct best Burg entropy estimate. However, as α increases to a certain critical value the solution becomes more and more concentrated, and beyond this value (1.4) fails to give even a feasible estimate.

The explanation given in the above papers in a self-professed nonrigorous fashion is that part of the real solution has condensed to a point mass, a claim also supported by considering discretized versions of the problem. The initial motivation of this work is to give a rigorous explanation of this phenomenon. In the course of this explanation we will develop a rather general duality and existence theory for the problem (1.1).

If, as the above example suggests, we should accept the possibility of measure-theoretic solutions to (1.1), then the question arises of how to reformulate the objective function. The constraints give no difficulty providing the a_i's are continuous, and the case where ϕ is piecewise linear and continuous is also clear—there is a strong analogy with semi-infinite linear programming, where point-mass solutions are familiar (see, for example, [1]).

The correct approach in the general case turns out to be to replace the objective function in (1.1) by what is essentially its second conjugate, which becomes a functional defined on measures. This idea is not in itself particularly new: see, for example, the discussion of "generalized solutions" in [16] and [47]. What is more remarkable is the simplicity and tractability of the resulting problem. In the first three sections of this work, relying heavily on the work of Rockafellar [43]-[47], we derive this extended primal problem, and investigate its relationship with the original primal and dual problems.

The next section returns to the underlying question of the existence of an optimal solution for the original problem (1.1). Using the extended solutions, we provide a general theory linking the boundary behaviour of the entropy and the local geometry of the underlying measure space with the existence question. This provides a unified and illuminating explanation of previous results in the literature [15], [56], [32], [6]. The last section discusses how extended solutions can be computed, and ends with some examples including a resolution of the example described above.

Just prior to submitting this article for publication, the authors became aware of recent unpublished work [20], [21] on some similar questions. The approach therein is very different from the purely convex analytic attack employed here. It relies on discretization and a Bayesian statistical interpretation, which lead to the application of large deviation theory (building on results in [13]). This probabilistic method, while seemingly less constructive than the convex programming approach, suggests intriguing connections between the two.

Problem (1.1) is a very general partially-finite program. As such, it models very many problems other than best entropy estimation. In particular, as outlined in [9], it includes numerous examples from constrained approximation, interpolation, and smoothing (see, for example, [38], [26], and [10]); the duality theory developed here also applies to some of these problems. The theory in this paper also allows an arbitrary linear functional to be added to the objective function. There has been recent interest in log-barrier penalty methods for semi-infinite linear programming, in the context of

the asymptotic behaviour of Karmarkar's method [42], [55], and our results may be applied here.

In the interests of economy, many reasonably routine computations and proofs are omitted; they can be found in [8] and [33].

2. Preliminaries. The measures of entropy with which we shall be concerned are integral functionals of the form $\int \phi(x(s))$, where $\phi : \mathbb{R} \to (-\infty, +\infty]$ is a closed, proper, convex function. We shall use the notation and terminology of [45] throughout. The conjugate function is denoted by ϕ^*, and the recession function $\phi 0+ : \mathbb{R} \to (-\infty, +\infty]$ is given by $(\phi 0+)(u) = \lim_{\lambda \to +\infty} (1/\lambda)\phi(u_0 + \lambda u)$, where u_0 is arbitrary in the domain of ϕ (see [45, Thm. 8.5]). The following result defines the constants p and q, which will be crucial in this paper. (These are entirely unrelated to the notation for the spaces L_p and L_q.) The proof is standard (see [8, Lemma 2.2]).

LEMMA 2.1. *The following limits exist:*

(2.2)
$$q := \lim_{u \to +\infty} \phi(u)/u \in (-\infty, +\infty],$$
$$p := \lim_{u \to -\infty} \phi(u)/u \in [-\infty, +\infty).$$

Furthermore, $p \leqq q$,

(2.3)
$$(\phi 0+)(u) = \begin{cases} qu & \text{if } u > 0, \\ 0 & \text{if } u = 0, \\ pu & \text{if } u < 0, \end{cases}$$

and $\operatorname{int}(\operatorname{dom}(\phi^*)) = (p, q)$. *The function* ϕ *is affine if and only if* $p = q$ *(so* $\operatorname{dom}(\phi^*) = \{p\}$*).*

Lemma 2.1 characterizes $\operatorname{dom}(\phi^*)$. It will also be helpful to have some notation for $\operatorname{dom}(\phi)$, so define β in $(-\infty, +\infty]$ as $\sup(\operatorname{dom}\phi)$ and α in $[-\infty, +\infty)$ as $\inf(\operatorname{dom}\phi)$, so $\operatorname{int}(\operatorname{dom}(\phi)) = (\alpha, \beta)$. The ideas of *essential strict convexity* and *essential smoothness* [45] will be useful to us. These concepts are particularly simple for univariate functions. We have that ϕ is essentially strictly convex (or, equivalently, ϕ is strictly convex on $\operatorname{dom}(\phi)$) if and only if ϕ^* is essentially smooth. This in turn is equivalent to $p < q$ and ϕ^* differentiable on (p, q) with $\lim_{v \downarrow p} (\phi^*)'(v) = -\infty$ if $p > -\infty$, and $\lim_{v \uparrow q} (\phi^*)'(v) = +\infty$ if $q < +\infty$. In this case,

(2.4)
$$\partial \phi^*(v) = \begin{cases} \{(\phi^*)'(v)\} & \text{if } v \in (p, q), \\ \emptyset & \text{otherwise.} \end{cases}$$

One particularly well behaved class of convex functions is that of Legendre type [45].

DEFINITION 2.5. We say ϕ is of *Legendre type* if it is essentially smooth and essentially strictly convex.

LEMMA 2.6. *Suppose* ϕ *is of Legendre type. Then, so is* ϕ^*, *and* $\phi' : (\alpha, \beta) \to (p, q), (\phi^*)' : (p, q) \to (\alpha, \beta)$ *are continuous, strictly increasing, and mutually inverse maps between the interiors of the domains of* ϕ *and* ϕ^*. *Also,* $q < +\infty$ *if and only if* $\beta = +\infty$ *and* $p > -\infty$ *if and only if* $\alpha = -\infty$.

Proof. See [45, Thm. 26.5]. The last part is immediate. □

We will use the notation, for u in $\mathbb{R}$, $u^+ := \max\{u, 0\}$, and $u^- := -\min\{u, 0\}$, so $u = u^+ - u^-$ and $|u| = u^+ + u^-$. If we adopt the convention that $(\pm\infty)\, 0 = 0$, we can rewrite (2.3) as $(\phi 0+)(u) = qu^+ - pu^-$.

The results in this paper will revolve around the computation of the conjugates and subdifferentials of various convex integral functions. We will rely heavily on the ideas and results of Rockafellar [45], [46]. For convenience, we will summarize the

notation to be used throughout the paper before proving the technical results that will be applied.

S is a compact Hausdorff space, with $z_0 \in C(S)$, the Banach space of continuous functions on S. Furthermore, $0 \leqq \rho \in M(S)$, the Banach space of regular Borel measures on S, and ρ has full support [48]. $I_\phi : L_1(S, \rho) \to (-\infty, +\infty]$ is defined by $I_\phi(x) := \int_S \phi(x(s)) \, d\rho$, and $I_{\phi^*} : L_\infty(S, \rho) \to (-\infty, +\infty]$ is defined by $I_{\phi^*}(z) := \int_S \phi^*(z(s)) \, d\rho$. $J_{\phi^*} : C(S) \to (-\infty, +\infty]$ is defined as the restriction of I_{ϕ^*} to $C(S)$. We have $b \in \mathbb{R}^n$, and $a = (a_1, \ldots, a_n) \in (C(S))^n$. The map $A : L_1(S, \rho) \to \mathbb{R}^n$ is defined by $(Ax)_i := \int_S a_i(s) x(s) \, d\rho$ for $i = 1, \ldots, n$. Finally, $B : \mathbb{R}^n \to C(S)$ is defined by $B\lambda := \lambda^T a$.

Some comments are in order concerning these definitions. We will often treat $C(S)$ with its usual supremum norm as a subspace of $L_\infty(S, \rho)$. We can regard $M(S)$, with its usual norm, as the dual of $C(S)$. The continuous linear map A has adjoint $A^* : \mathbb{R}^n \to L_\infty(S, \rho)$, which may be identified with the continuous linear map B, as is easily checked. Also, $B^* : M(S) \to \mathbb{R}^n$ is continuous and given by $(B^*\rho)_i := \int a_i \, d\rho$. For the relevant ideas, see, for example, [48] and [49].

The function ϕ is a normal convex integrand, so the integral functional I_ϕ is a well-defined, convex, lower semicontinuous function, with conjugate I_{ϕ^*} [46]. The function J_{ϕ^*} is also well defined and convex (see, for example, [47, Thm. 3]). Much of this section will be devoted to studying its conjugate.

We will write, for any μ in $M(S)$, $\mu = \mu^+ - \mu^-$ for the Jordan decomposition, $\mu = \mu_\alpha + \mu_\sigma$ for the Lebesgue decomposition with respect to ρ (so $\mu_\alpha \ll \rho$ and $\mu_\sigma \perp \rho$), and $(d\mu_\alpha / d\rho) \in L_1(S, \rho)$ for the Radon–Nikodym derivative [48].

THEOREM 2.7. *The function J_{ϕ^*} is well defined, lower semicontinuous, and convex. It is continuous on the set $\{z \in C(S) \mid z(s) \in (p, q) \text{ for all } s \in S\}$. The conjugate function $J_{\phi^*}^* : M(S) \to (-\infty, +\infty]$ is given by*

$$(2.8) \qquad J_{\phi^*}^*(\mu) = \int_S \phi\left(\frac{d\mu_\alpha}{d\rho}(s)\right) d\rho + q\mu_\sigma^+(S) - p\mu_\sigma^-(S).$$

The proof of this result in the affine case is a straightforward calculation, while the case $p < q$ is a direct application of [46, Thm. 5] (see also [8, Thm. 3.1]).

COROLLARY 2.9. *Suppose $x \in L_1(S, \rho)$ and $0 \leqq \nu, \xi \in M(S)$. If $d\mu = x \, d\rho + d\nu - d\xi$ then $J_{\phi^*}^*(\mu) \leqq I_\phi(x) + q\nu(S) - p\xi(S)$ with equality if ρ, ν, and ξ are mutually singular.*

Proof. Let $\gamma := \nu - \xi$. Then $\nu \geqq \gamma^+$ and $\xi \geqq \gamma^-$ (see [48, p. 127]). By Theorem 2.7 and the definition of $\phi 0+$,

$$J_{\phi^*}^*(\mu) = \int_S \phi\left(x(s) + \frac{d\gamma_\alpha}{d\rho}(s)\right) d\rho + q\gamma_\sigma^+(S) - p\gamma_\sigma^-(S)$$

$$\leqq \int_S \left[\phi(x(s)) + (\phi 0+)\left(\frac{d\gamma_\alpha}{d\rho}(s)\right)\right] d\rho + q\gamma_\sigma^+(S) - p\gamma_\sigma^-(S)$$

$$= I_\phi(x) + \int_S \left[q\frac{d\gamma_\alpha^+}{d\rho}(s) - p\frac{d\gamma_\alpha^-}{d\rho}(s)\right] d\rho + q\gamma_\sigma^+(S) - p\gamma_\sigma^-(S)$$

$$= I_\phi(x) + q\gamma_\alpha^+(S) - p\gamma_\alpha^-(S) + q\gamma_\sigma^+(S) - p\gamma_\sigma^-(S)$$

$$= I_\phi(x) + q\gamma^+(S) - p\gamma^-(S)$$

$$\leqq I_\phi(x) + p(\nu(S) - \xi(S)) + (q - p)\nu(S).$$

If ρ, ν, and ξ are mutually singular, then $\gamma_\alpha = 0$, so $d\gamma_\alpha / d\rho(s) = 0$ almost everywhere, $[\rho]$ on S, and $\nu = \gamma^+$ (by Hahn decomposition), so we have equality above. $\square$

We now compute the subdifferential of J_{ϕ^*}. This will be fundamental in deriving optimality conditions.

THEOREM 2.10. *Suppose $z \in C(S)$ and $\mu \in M(S)$. Then $\mu \in \partial J_{\phi^*}(z)$ (or equivalently, $J_{\phi^*}(z) + J_{\phi^*}^*(\mu) = \int_S z(s)\, d\mu$) if and only if*

$$z(s) \in [p, q] \quad \text{for all } s \in S,$$

$$\frac{d\mu_\alpha}{d\rho}(s) \in \partial\phi^*(z(s)) \quad a.e.\, [\rho]\, on\, S,$$

$$\text{support } (\mu_\sigma^+) \subset \{s \,|\, z(s) = q\},$$

and

$$\text{support } (\mu_\sigma^-) \subset \{s \,|\, z(s) = p\}.$$

Proof. We will assume that ϕ is not affine: the affine case, like Theorem 2.7, is a straightforward calculation, and we will not use this case in what follows. We assume, therefore, that $p < q$, and apply Corollary 5A of [46]. As in the proof of Theorem 2.7, we will apply Rockafellar's result with $D(s) := (p, q)$ for all s in S. Writing $\mathbb{R}_+$ for the nonnegative reals, the normal cone to $\text{cl}\,(D(s))$ is given by

$$(2.11) \qquad N_{[p, q]}(v) = \begin{cases} -\mathbb{R}_+ & \text{if } v = p, \\ \{0\} & \text{if } v \in (p, q), \\ \mathbb{R}_+ & \text{if } v = q, \end{cases}$$

for v in $[p, q]$. Applying Rockafellar's result shows that $\mu \in \partial J_{\phi^*}(z)$ is equivalent to the first two statements along with μ_σ being $N_{[p,q]}$-valued: in other words (using the fact that $\mu_\sigma \ll |\mu_\sigma|$),

$$(2.12) \qquad \frac{d\mu_\sigma}{d|\mu_\sigma|} \in N_{[p,q]}(z(s)) \quad a.e.\, [|\mu_\sigma|]\, on\, S.$$

The remainder of the proof is reasonably straightforward measure theory (see [8, Thm. 3.5]). $\quad\square$

3. Primal and dual constraint qualifications.

The optimization problem that we wish to consider is

$$\inf \qquad \int_S [\phi(x(s)) + z_0(s)x(s)]\, d\rho,$$

$$\text{subject to} \qquad \int_S a_i(s)x(s)\, d\rho = b_i \quad \text{for } i = 1, \ldots, n,$$

$$x \in L_1(S, \rho),$$

or in our previous notation,

$$(P) \qquad \inf \{I_\phi(x) + \langle z_0, x\rangle \,|\, Ax = b \text{ and } x \in L_1(S, \rho)\}.$$

The extra linear functional corresponding to z_0 in the objective is introduced to allow us to model some best entropy estimation problems where a prior estimate is given (see, for example, [28] and [29]), and to consider the log-barrier penalty function for semi-infinite linear programming [55]. We could consider the problem (P) posed in any of the spaces $L_r(S, \rho)$, for $1 \leq r \leq \infty$, or even in $C(S)$, but since $L_1(S, \rho)$ is the largest of these spaces, it is the natural choice if we wish to find an optimal solution.

Unfortunately, $L_1(S, \rho)$ is not typically a dual space, so we are unable to use weak-star compactness arguments to prove attainment. Furthermore, unless $p = -\infty$ and $q = +\infty$, the level sets of I_ϕ will not typically be weakly compact in L_1 (see [7]). In this case special arguments are needed to prove attainment, dependent on the underlying measure space (S, ρ) and the constraint map A (see, for example, [6]).

The idea of considering solutions to optimization problems in L_1 which may have singular components is not new. An example in optimal control appears in [4], and was extended in [41]. In this latter thesis the approach taken is to consider the problem in Fenchel form and then to solve the second dual. This gives a so-called "weak" solution (see [16, § III.6]).

For this reason we introduce the following "extended primal problem:"

(P_E) $\qquad\qquad$ $\inf\{J_{\phi*}^*(\mu) + \langle z_0, \mu \rangle \,|\, B^*\mu = b \text{ and } \mu \in M(S)\}.$

Using Corollary 2.9 we can rewrite this as

$$\inf \qquad \int_S [\phi(x(s)) + z_0(s)x(s)]\,d\rho + q\nu^+(S) - p\nu^-(S) + \int_S z_0(s)\,d\nu$$

(P_E^1) subject to $\quad \displaystyle\int_S a_i(s)x(s)\,d\rho + \int_S a_i(s)\,d\nu = b_i \quad \text{for } i = 1, \ldots, n,$

$$x \in L_1(S, \rho), \quad \nu \in M(S), \quad \nu \perp \rho.$$

Notice that (P_E^1) is exactly (P) if we require the singular component $\nu = 0$. Under reasonable conditions (P_E) will always have an optimal solution: as we shall see, the singular component corresponds with singularities observed in practice when (P) fails to have an optimal solution. In fact, Corollary 2.9 allows us to omit the constraint $\nu \perp \rho$ if so desired (see [8, Thm. 5.3]).

Our arguments are based on duality techniques. The dual problem for (P) (see [6]) is

(P^*) $\qquad\qquad$ $\sup\{b^T\lambda - I_{\phi*}(A^*\lambda - z_0) \,|\, \lambda \in \mathbb{R}^n\},$

which we may write as

(3.1) $\qquad\qquad$ $\sup\{b^T\lambda - J_{\phi*}(B\lambda - z_0) \,|\, \lambda \in \mathbb{R}^n\},$

or as

(3.2) $\qquad\qquad$ $\sup\left\{b^T\lambda - \int_S \phi^*(\lambda^T a(s) - z_0(s))\,d\rho \,|\, \lambda \in \mathbb{R}^n\right\}.$

We denote the value of an optimization problem (Q) by $V(Q) \in [-\infty, +\infty]$. We say (Q) is *consistent* if there is a choice of the variable that satisfies the constraints and has finite objective value.

As usual, we have an easy weak duality result. The problems (P_E) and (P^*) (written in the form (3.1)) are Fenchel duals of each other, so a simple dual constraint qualification ensures that $V(P_E) = V(P^*)$ and $V(P_E)$ is attained (the motivation for its introduction). We will henceforth ignore the case where ϕ is affine, which is trivial.

Dual Constraint Qualification. The function ϕ is not affine, and there exists a $\hat{\lambda}$ in $\mathbb{R}^n$ with $\hat{\lambda}^T a(s) - z_0(s) \in (p, q)$ for all s in S.

Note that the assumption that ϕ is not affine ensures that $p < q$. If, as frequently occurs in practice, one of the a_i's is a nonzero constant function, $z_0 = 0$, and ϕ is not affine then the Dual Constraint Qualification will hold.

In order to ensure attainment in the dual problem (P^*) we need a primal constraint qualification. We recall from [9] that if x lies in a convex subset C of a topological vector space X, then x is a *quasi-relative interior* point of C ($x \in \text{qri}(C)$) if cl (cone $(C - x)$) is a subspace.

We will write $[\alpha, \beta]_{L_1}$ for the order interval $\{x \in L_1 | \alpha \leq x(s) \leq \beta$ almost everywhere$\}$. The usual constraint qualification for (P) is written

(PCQ_1) $\qquad\qquad\qquad\qquad b \in \text{ri}(A\,\text{dom}(I_\phi))$

(see, for example, [43]). Since this condition may be difficult to check, we will rewrite it in a more familiar Slater-type form.

(PCQ_2) $\qquad$ There exists $\tilde{x} \in \text{qri}(\text{dom}(I_\phi))$, which is feasible for (P).

This in turn can be stated in the following equivalent but more applicable form.

Primal Constraint Qualification. There exists a function $\hat{x}$ in $L_1(S, \rho)$ such that $\hat{x}(s) \in \text{ri}(\text{dom}(\phi))$ almost everywhere, and $A\hat{x} = b$.

(Of course, $\text{ri}(\text{dom}(\phi)) = (\alpha, \beta)$ unless ϕ is the indicator function of a point.) The following result may be found in [33].

LEMMA 3.3. *The Primal Constraint Qualification, (PCQ_1), and (PCQ_2) are equivalent. Furthermore,* $\text{ri}(A\,\text{dom}(I_\phi)) = \text{ri}(A[\alpha, \beta]_{L_1})$, *and providing* $\alpha < \beta$,

$$\text{aff}(A\,\text{dom}(I_\phi)) = \text{aff}(A[\alpha, \beta]_{L_1}) = \text{Range}(A) = \text{Range}(B^*).$$

If the constraint functions $a_1, \ldots, a_n$ are *pseudo-Haar*, or, in other words, linearly independent on every subset of S with positive measure (see [6]), then the Primal Constraint Qualification can be weakened to:

(PCQ_3) $\qquad$ There exists an $\hat{x}$ in $L_1(S, \rho)$ with $A\hat{x} = b$, and

$$\rho\{s \in S \,|\, \alpha < \hat{x}(s) < \beta\} > 0.$$

For a proof, see [33]. In summary, the Primal Constraint Qualification is easy to check in practice.

THEOREM 3.4 (duality). $V(P) \geqq V(P_E) \geqq V(P^*)$. *If the Dual Constraint Qualification holds, then* $V(P_E) = V(P^*)$, *and if, furthermore, (P_E) is consistent then $V(P_E)$ is attained. If, on the other hand, the Primal Constraint Qualification holds then* $V(P) = V(P_E) = V(P^*)$, *and if, furthermore, (P^*) is consistent then $V(P^*)$ is attained.*

Proof. The first claim (weak duality) is straightforward (see [8, Prop. 4.3]). Suppose the Dual Constraint Qualification holds. By Theorem 2.7, J_{ϕ^*} is finite and continuous at $B\hat{\lambda} - z_0$, where $\hat{\lambda}$ is the point in the Dual Constraint Qualification. Thus by [46, Thm. 3],

$$\min\{J_{\phi^*}^*(\mu) + \langle z_0, \mu\rangle \,|\, B^*\mu = b, \text{ and } \mu \in M(S)\}$$

$$= \sup\{b^T\lambda - J_{\phi^*}(B\lambda - z_0) \,|\, \lambda \in \mathbb{R}^n\},$$

which is exactly the required result. If, on the other hand, the Primal Constraint Qualification holds, then $V(P) = V(P^*)$ by Corollary 2.6 of [6]. It follows by weak duality that $V(P) = V(P_E) = V(P^*)$. $\qquad\square$

Our next step is to derive the optimality conditions. The proof is an easy application of weak duality and Theorem 2.10 (see [8, Thm. 4.10]).

(OCP_E) $\qquad$ $\begin{cases} \bar{\mu} \text{ is feasible for } (P_E), \text{ and} \\ \\ \bar{\mu} \in \partial J_{\phi^*}(B\bar{\lambda} - z_0), \end{cases}$

(OCP_E^1) $\begin{cases} (\bar{x}, \bar{\nu}) \text{ is feasible for } (P_E^1), \\[2mm] \bar{x}(s) \in \partial\phi^*(\bar{\lambda}^T a(s) - z_0(s)) \quad \text{a.e. on } S, \\[2mm] \text{support } (\bar{\nu}^+) \subset \{s \in S \,|\, \bar{\lambda}^T a(s) - z_0(s) = q\}, \text{ and} \\[2mm] \text{support } (\bar{\nu}^-) \subset \{s \in S \,|\, \bar{\lambda}^T a(s) - z_0(s) = p\}, \end{cases}$

(OCP) $\begin{cases} \bar{x} \text{ is feasible for } (P), \text{ and} \\[2mm] \bar{x}(s) \in \partial\phi^*(\bar{\lambda}^T a(s) - z_0(s)) \quad \text{a.e. on } S. \end{cases}$

THEOREM 3.5. (i) (OCP_E) *holds if and only if* $\bar{\mu}$ *is optimal for* (P_E) *and* $\bar{\lambda}$ *is optimal for* (P^*), *with equal objective value.*

(ii) (OCP_E^1) *holds if and only if* $(\bar{x}, \bar{\mu})$ *is optimal for* (P_E^1) *and* $\bar{\lambda}$ *is optimal for* (P^*), *with equal objective value.*

(iii) (OCP) *holds if and only if* $\bar{x}$ *is optimal for* (P) *and* $\bar{\lambda}$ *is optimal for* (P^*), *with equal objective value.*

COROLLARY 3.6 (strong duality). *Suppose that the Primal and Dual Constraint Qualifications hold. Then the two primal problems* (P) *and* (P_E) *(and* (P_E^1)*)) and the dual problem* (P^*) *all have equal, finite value, and there exist optimal solutions* $\bar{\mu}$ *for* (P_E) *(and* $(\bar{x}, \bar{\nu})$ *for* (P_E^1)*)), and* $\bar{\lambda}$ *for* (P^*), *satisfying* (OCP_E) *(or* (OCP_E^1), *respectively).*

Part (iii) of Theorem 3.5 is extremely instructive. In practice ϕ^* is usually differentiable, so the last condition of (OCP) becomes

$$(3.7) \qquad\qquad \bar{x}(s) = (\phi^*)'(\bar{\lambda}^T a(s) - z_0(s)).$$

It has been a frequent error in the more practical literature to assume that if $\bar{\lambda}$ is dual optimal then (3.7) gives the optimal solution of the primal problem (P). The feasibility of this $\bar{x}$ is justified by differentiating under the integral in (3.2) with respect to λ. Unfortunately, as we shall see, in quite simple examples (satisfying the Primal and Dual Constraint Qualifications) the $\bar{x}$ given by (3.7) can lie in L_1 and yet *fail to be feasible*.

Theorem 3.6 shows that, under reasonable conditions, the $\bar{x}$ given by (3.7) corresponds to the *absolutely continuous part* of an optimal solution of the *extended* primal problem (P_E). It will be optimal for the original primal problem (P) if and only if it is *feasible*. If it fails to be feasible this is due to singular components of the optimal solution, supported on the set where $\bar{\lambda}^T a(s) - z_0(s)$ hits the boundary of the domain of ϕ^*. In principle, if this set is large, these singular components could be very unpleasant, making any practical application or interpretation impossible. In fact, we can generally restrict our attention to singular components consisting of finitely many point masses (see [8]).

For the time being we confine ourselves to interpreting the singular components in terms of primal optimizing sequences (cf. [16, Prop. III.6.1]). A standard argument (see [8, Thm. 4.13]) gives the following result.

THEOREM 3.8. *Suppose the sequence* $(x_r)_1^\infty$ *in* $L_1(S, \rho)$ *is an* optimizing sequence *for the primal problem* (P): $Ax_r \to b$ *and* $I_\phi(x_r) + \langle z_0, x_r \rangle \to V(P)$ *as* $r \to \infty$. *Suppose also that the Primal Constraint Qualification holds. Then the limit of any weak-star convergent subsequence of* $(x_r \, d\rho)_1^\infty$ *in* $M(S)$ *is optimal for the extended primal problem* (P_E).

Standard compactness arguments show that there will exist weak-star convergent subsequences in the above result if, for example, S is metrizable, $\phi(u) = +\infty$ for $u < 0$, and for some j, $a_j(s) > 0$ on S (see [8, Cor. 4.14]).

In [33] these results are applied to progressively refined discretizations of the primal problem: it is shown that the corresponding optimal solutions typically have weak-star convergent subsequences, any of which converge to an optimal solution of the extended problem. This provides another more concrete justification for considering this extension of the primal problem.

4. Primal attainment. As we saw in § 3, the existence of an optimal solution of the extended primal problem (P_E) (or any of its equivalent formulations) is a straightforward consequence of the Dual Constraint Qualification. By contrast, attainment in the original primal problem (P) is a much more delicate matter: as we shall see, there may fail to be an optimal solution in even very simple examples. The existence question depends not only on the function ϕ in the objective but also on the smoothness of the constraint functions $a_1, \ldots, a_n$, on z_0, and on geometric and measure-theoretic properties of the underlying space (S, ρ). This question was addressed in [6], where the existence of an optimal solution was demonstrated in particular for classical (algebraic and trigonometric) moment problems with the Burg entropy as objective, when (S, ρ) is a one-dimensional interval with Lebesgue measure. This had been known previously for the trigonometric case (where the interval is $[-\pi, +\pi]$ and the moment conditions consist of the first n Fourier coefficients of x) using very special contour integral techniques [15], and for the two-dimensional trigonometric case in [56], and more generally in [32]. The approach of the latter two papers is a direct investigation of the map that takes a polynomial to the moments of its reciprocal. A contrasting, duality-based approach is taken in [36]: some technical difficulties remain, as discussed after Corollary 3.6.

In this section we will extend and clarify the results in [6] by using the results in § 3 on the existence of *extended* primal solutions. In particular, our new results will give an entirely rigorous proof that the Burg entropy also entails the existence of an optimal solution in the two-dimensional trigonometric case. By contrast, as we shall see, simple three-dimensional problems fail to have optimal solutions. The idea is very simple: given an extended primal solution $(\bar{x}, \bar{\nu})$, we need a condition to ensure, via Theorem 3.5, that the singular part $\bar{\nu}$ vanishes.

To summarize, the approach here has three substantial advantages over [6]. First, it is extremely natural, unlike the techniques in [6]. Second, it generalizes the results in [6] to other important practical cases. Third, it reveals exactly the sense in which existence can fail.

We begin with an informal discussion. Let us denote by $\Psi : \mathbb{R}^n \to (-\infty, +\infty]$ the function $I_{\phi^*}(A^*(\cdot) - z_0)$, so the dual problem ($P^*$) consists of minimizing the convex function $\Psi(\lambda) - b^T\lambda$. Suppose for simplicity that $a(s)$ is nonzero for every s in S. Then it is easily checked that the interior of the domain of the dual objective function is equal to

$$\text{int}(\text{dom}(\Psi)) = \{\lambda \in \mathbb{R}^n \mid \lambda^T a(s) - z_0(s) \in (p, q) \text{ for all } s \in S\}.$$

Suppose the Primal Constraint Qualification holds so there exists a dual optimum, say $\bar{\lambda}$, with $b \in \partial\Psi(\bar{\lambda})$. As is usual in convex analysis, the difficulties, if any, occur at the boundary of dom(Ψ), while if $\bar{\lambda} \in \text{int}(\text{dom}(\Psi))$ easy arguments identical to those that follow show the existence of a solution to the primal problem (P). Of course, this must be the case if dom(Ψ) is open. However, it will be true more generally, provided there are no boundary points in dom Ψ at which subgradients exist. The difficulty is in checking this, since boundary subgradients may exist even when ϕ^* is essentially smooth.

That is the origin of the following condition; we will work with it directly, but similar arguments show it implies that $\partial \Psi(\lambda) = \emptyset$ whenever $\lambda \notin \text{int}(\text{dom}(\Psi))$. This also has important computational consequences. In practice, (P) is generally solved via the dual, so we seek to minimize Ψ. When the condition below holds any minimizer must lie in $\text{int}(\text{dom}(\Psi))$. Thus we can apply unconstrained search techniques (appropriately safeguarded).

INTEGRABILITY CONDITION. *For any function* $z := \lambda^T a - z_0$ *(with λ in $\mathbb{R}^n$), if* $z(s) \in (p, q)$ *almost everywhere on S and* $(\phi^*)'(z(\cdot)) \in L_1(S, \rho)$, *then it follows that* $z(s) \in (p, q)$ *for all s in S where $a(s)$ is nonzero.*

THEOREM 4.1. (i) *Suppose ϕ is essentially strictly convex. Then if $(\bar{x}_1, \bar{\nu})$ and $(\bar{x}_2, \bar{\nu})$ are both optimal for the extended primal problem (P_E^1) then $\bar{x}_1 = \bar{x}_2$, so in particular the original primal problem (P) has at most one solution.*

(ii) *Let us suppose furthermore that the Primal Constraint Qualification holds. Then the dual problem (P^*) has an optimal solution, and if $(\bar{x}.\bar{\nu})$ is optimal for (P_E^1) and $\bar{\lambda}$ is optimal for (P^*), then*

(4.2) $$\bar{x}(s) = (\phi^*)'(\bar{\lambda}^T a(s) - z_0(s)) \quad a.e. \text{ on } S;$$

so, in particular, if (P) has an optimal solution it is given uniquely by (4.2).

(iii) *Moreover, suppose also that the Dual Constraint Qualification holds. Then (P_E^1) has an optimal solution $(\bar{x}, \bar{\nu})$ with the absolutely continuous part given uniquely by (4.2).*

(iv) *If, in addition, the Integrability Condition holds, then the singular part $\bar{\nu}$ vanishes, so (4.2) gives the unique optimal solution of (P).*

Proof. Part (i) follows by strict convexity.

Parts (ii) and (iii) follow by Theorem 3.5 and (2.4). Assume finally that $(\bar{x}, \bar{\nu})$ is optimal for (P_E^1) with $\bar{x}$ given by (4.2), and suppose the Integrability Condition holds. If we write $S_0 := \{s \in S \mid a(s) = 0\}$ then, from (OCP_E^1), $\bar{\nu}$ is supported on S_0, and by the Dual Constraint Qualification, $-z_0(s) \in (p, q)$ for all s in S_0.

But now $(\bar{x}, 0)$ is also feasible for (P_E^1), with a corresponding drop in the objective value of

$$q\bar{\nu}^+(S) - p\bar{\nu}^-(S) + \int_S z_0(s)\,d\bar{\nu} = \int_{S_0}(q + z_0(s))\,d\bar{\nu}^+ - \int_{S_0}(p + z_0(s))\,d\bar{\nu}^- > 0,$$

unless $\bar{\nu} = 0$. Hence the result. $\square$

The Integrability Condition actually turns out to be necessary, as well as sufficient, for the existence of a primal solution in general. That is the substance of the next result.

THEOREM 4.3. *Suppose ϕ is of Legendre type and the Integrability Condition fails. Then there exists a right-hand side b in $\mathbb{R}^n$ such that the primal problem (P) satisfies the Primal Constraint Qualification, but has no optimal solution.*

Proof. Since the Integrability Condition fails, there exists a function $\bar{z} := \bar{\lambda}^T a - z_0$ satisfying $\bar{z}(s) \in (p, q)$ almost everywhere and with $\bar{x}(\cdot) := (\phi^*)'(\bar{z}(\cdot))$ in $L_1(S, \rho)$, but with $S_1 := \{s \in S \mid a(s) \neq 0, \bar{z}(s) = p \text{ or } q\}$ nonempty. Define $\bar{b} := A\bar{x}$. Note that $\alpha < \bar{x}(s) < \beta$ almost everywhere by Lemma 2.6. Thus $\bar{b} \in \text{ri}(A \text{ dom}(I_\phi))$ by Lemma 3.3.

Now choose any ν in $M(S)$, with

$$\text{support}(\nu^+) \subset \{s \in S \mid a(s) \neq 0, \bar{z}(s) = q\},$$
$$\text{support}(\nu^-) \subset \{s \in S \mid a(s) \neq 0, \bar{z}(s) = p\},$$

and $B^*\nu = \int_S a\,d\nu \neq 0$. For example, a point mass at any point of S (with the appropriate sign) will do. It follows by Lemma 3.3 that

$$b := \bar{b} + \varepsilon B^*\nu \in \text{ri}(A \text{ dom}(I_\phi)),$$

provided that $\varepsilon > 0$ is sufficiently small.

Clearly now, the Primal Constraint Qualification holds (Lemma 3.3). Furthermore, if we write $\bar{\nu} := \varepsilon\nu$, then $(\bar{x}, \bar{\nu})$ and $\bar{\lambda}$ satisfy (OCP_E^1) and thus are optimal for (P_E^1) and (P^*), respectively, so $\bar{x}$ is the only possible optimal solution of (P), by Theorem 3.5(ii) and (iii). However, $\bar{x}$ is not feasible for (P), since $b \neq \bar{b}$. $\square$

We now pursue a slight digression, to discuss the approach of [56] and [32]. We will show that their key supporting result, which is of some independent interest, can be subsumed by this approach. The idea of Woods, and Lang and McClellan (working in the special case where the a_i's are multidimensional trigonometric polynomials and ϕ is the Burg entropy) is to consider the nonlinear system of equations in $\lambda \in \mathbb{R}^n$ derived (formally in these references but rigorously above) from the optimality conditions (OCP):

$$(NLE) \qquad \int_S a_i(s)(\phi^*)'(\lambda^T a(s) - z_0(s)) \, d\rho = b_i \quad \text{for } i = 1, \ldots, n.$$

Assuming the existence of a dual optimal λ (a difficulty not addressed in the above papers), the primal optimal solution $\bar{x}$ (if it exists) must have the form $(\phi^*)' (\lambda^T a(s) - z_0(s))$, so it may be obtained by solving (NLE) for λ.

Assuming ϕ is of Legendre type it is clear, as in the proof of Theorem 4.3, that (NLE) is certainly not solvable unless $b \in \mathrm{ri}(A[\alpha, \beta]_{L_1})$. The point (obvious from an optimization viewpoint but surprising ab initio) is that the Integrability Condition gives a complete characterization.

COROLLARY 4.4. *Suppose ϕ is of Legendre type and the Dual Constraint Qualification holds. Then (NLE) is solvable for every b in $\mathrm{ri}(A[\alpha, \beta]_{L_1})$ if and only if the Integrability Condition holds.*

Proof. The first direction follows from Theorem 4.3 and the comments above. The converse follows from Theorem 4.1. $\square$

Taking ϕ to be the Burg entropy and the a_i's as (multidimensional) trigonometric polynomials, we obtain the result in the Appendix of [32].

These results demonstrate the importance of the Integrability Condition for the question of attainment in the original primal problem. The remainder of this section will be devoted to investigating for what spaces (S, ρ), objectives ϕ and z_0, and constraints a it holds. We shall see that the important features are the local geometry of the set S, and the growth rate of $(\phi^*)'$ near p and q. We adopt an approach which gives unified conditions for the cases of common interest, namely, $S \subset \mathbb{R}^m$ for $m = 1, 2, 3$. We shall suppose for the remainder of this section that S is a compact metric space with metric $d(\cdot, \cdot)$, and we write $B(s, r)$ for the open ball, centre s, radius r. For any s in S we define $\chi_s(r) := \rho(S \cap B(s, r))$. The following result is derived from an elementary estimate of the integral in the Integrability Condition (see [8, Thm. 6.6]).

THEOREM 4.5. *Suppose S is a compact metric space, $a_1, \ldots, a_n$ and z_0 are Lipschitz on S, ϕ is essentially strictly convex, and the following two conditions hold for any s_0 in S and $k > 0$:*

$$\lim_{\delta \downarrow 0} \inf_{0 < \varepsilon \leq r \leq \delta} r[(1/\varepsilon)(\chi_{s_0}(r + \varepsilon) - \chi_{s_0}(r))][(\phi^*)'(q - kr)] > 0, \quad \text{if } q < +\infty;$$

$$\lim_{\delta \downarrow 0} \inf_{0 < \varepsilon \leq r \leq \delta} (-r)[(1/\varepsilon)(\chi_{s_0}(r + \varepsilon) - \chi_{s_0}(r))][(\phi^*)'(p + kr)] > 0, \quad \text{if } p > -\infty.$$

Then the Integrability Condition holds.

In practice S is often a compact subset of $\mathbb{R}^m$ with Lebesgue measure, and in this case we can often simplify the required conditions. The *Dubovitskij–Miljutin* (DM) *cone* will be useful in what follows (see, for example, [2]).

DEFINITION 4.6. For a subset K of a normed space V, and for s in cl K, we define $D_K(s) := \{v \in V \mid s + (0, \varepsilon]B(v, \varepsilon) \subset K$ for some $\varepsilon > 0\}$. We say that K is DM-*regular* if $D_K(s)$ is nonempty for all s in K.

The condition of DM-regularity ensures that the sets in which we are interested have no cusps. In a normed space it is easily checked that any convex set with nonempty interior is DM-regular. A subset of a normed space defined by inequalities will be DM-regular providing a suitable constraint qualification holds everywhere (see [2, p. 126], for example). Obviously, an arbitrary union of DM-regular sets is DM-regular.

Let us denote m-dimensional Lebesgue measure by τ_m. Using the fact that an open convex cone must intersect the surface of the unit sphere with positive area, we obtain the following (see [8, Lemma 6.11]).

LEMMA 4.7. *Suppose* $S \in \mathbb{R}^m$ *(with Euclidean distance) is compact and DM-regular, and for some* $k > 0$, $\rho \geqq k\tau_m$ *on* S. *Then for any* s_0 *in* S,

$$(4.8) \qquad \lim_{\delta \downarrow 0} \inf_{0 < \varepsilon \leqq r \leqq \delta} r^{1-m}[(1/\varepsilon)(\chi_{s_0}(r + \varepsilon) - \chi_{s_0}(r))] > 0.$$

We can now derive a more useful version of Theorem 4.5.

THEOREM 4.9. *Suppose that* S *is a compact, DM-regular subset of* $\mathbb{R}^m$, *and* ρ *dominates a positive multiple of Lebesgue measure on* S. *Suppose that* $a_1, \ldots, a_n$ *and* z_0 *are Lipschitz on* S. *Finally, suppose that* ϕ *is essentially strictly convex, with* $\underline{\lim}_{r \downarrow 0} r^m(\phi^*)'(q - r) > 0$ *if* $q < +\infty$, *and if* $p > -\infty$, $\underline{\lim}_{r \downarrow 0} -r^m(\phi^*)'(p + r) > 0$. *Then the Integrability Condition holds, so if in addition the Primal and Dual Constraint Qualifications hold, the original primal problem* (P) *has a unique optimal solution.*

Proof. By Lemma 4.7, for any s_0 in S, (4.8) holds. Now for any $k > 0$, if $q < +\infty$,

$$\lim_{\delta \downarrow 0} \inf_{0 < \varepsilon \leqq r \leqq \delta} r[(1/\varepsilon)(\chi_{s_0}(r + \varepsilon) - \chi_{s_0}(r))][(\phi^*)'(q - kr)]$$

$$\geqq \left\{ \lim_{\delta \downarrow 0} \inf_{0 < \varepsilon \leqq r \leqq \delta} r^{1-m}[(1/\varepsilon)(\chi_{s_0}(r + \varepsilon) - \chi_{s_0}(r))] \right\} \left\{ \lim_{\delta \downarrow 0} \inf_{0 < r \leqq \delta} r^m(\phi^*)'(q - r) \right\}$$

(since both factors are nonnegative). The first factor is strictly positive by (4.8), and the second is strictly positive by assumption. The first condition of Theorem 4.5 follows, and a similar argument shows the second condition. The result now follows from Theorem 4.5. $\square$

Probably the most important application of this result is when S is a compact interval of $\mathbb{R}$ and ϕ is the Burg entropy (1.3). In particular, we obtain the original existence result of [15].

The periodic case. In many cases in practice the moment conditions are given by Fourier coefficients. In other words, the constraint functions $a_1, \ldots, a_n$ are trigonometric polynomials (possibly multidimensional) and hence periodic. In these cases it is often possible to weaken the conditions for attainment in the original problem.

DEFINITION 4.10. Suppose $e^1, \ldots, e^m$ form a basis of $\mathbb{R}^m$. With respect to this basis, we say $S \subset \mathbb{R}^m$ is *covering* if $\bigcup_{\theta \in \mathbb{Z}^m} (S + \sum_{j=1}^m \theta_j e^j) = \mathbb{R}^m$. We say a function $z : S \to \mathbb{R}$ is *periodic* if $z(s) = z(s')$ whenever $s - s' = \sum_{j=1}^m \theta_j e^j$ for some θ in $\mathbb{Z}^m$.

The idea is to use the following simple result [8, Lemma 6.16].

LEMMA 4.11. *Suppose* $z : S \to \mathbb{R}$ *is periodic, with* ∇z *Lipschitz on* S, *and suppose* S *is covering. Suppose* $z(s) \leqq q$ *for all* s *in* S *and* $z(s_0) = q$. *Then for some* $k_1 > 0$, $z(s) \geqq q - k_1 \|s - s_0\|^2$ *for all* s *in* S.

Using this to estimate $(\phi^*)'(z(s))$ we arrive at the following refinement of Theorem 4.9.

THEOREM 4.12. *Suppose that* S *is a compact, DM-regular, covering subset of* $\mathbb{R}^m$, *and* ρ *dominates a positive multiple of Lebesgue measure on* S. *Suppose that* $\nabla a_1, \ldots, \nabla a_n$

and ∇z_0 are Lipschitz on S and $a_1, \ldots, a_n$ and z_0 are periodic. Finally, suppose that ϕ is essentially strictly convex, with $\varliminf_{r \downarrow 0} r^m (\phi^*)'(q - r^2) > 0$ if $q < +\infty$, and $\varliminf_{r \downarrow 0} - r^m (\phi^*)'(p + r^2) > 0$ if $p > -\infty$. Then the Integrability Condition holds, so if in addition the Primal and Dual Constraint Qualifications hold, the original primal problem (P) has a unique optimal solution.

Again, probably the most important application of the above result is when ϕ is the Burg entropy and $S = [-\pi, \pi]^2$ with trigonometric polynomials $a_1, a_2, \ldots, a_n, z_0$. In particular, we obtain the existence result in [56].

Theorems 4.9 and 4.12 involve growth conditions on $(\phi^*)'$. It is easy to translate these into conditions on ϕ', if so desired, using Lemma 2.6 (see [8, Lemma 6.19]).

5. Computation, primal uniqueness, and examples. In this section we will discuss how to solve the extended primal problem (P_E), and give conditions ensuring it has a unique solution. Suppose that the Primal and Dual Constraint Qualifications hold and that ϕ is essentially strictly convex, so ϕ^* is essentially smooth. From the Strong Duality Theorem (Corollary 3.6) we know that the dual problem (P^*) has an optimal solution $\bar{\lambda}$, and (P_E^1) has an optimal solution $(\bar{x}, \bar{\nu})$, where the absolutely continuous part $\bar{x}$ is given uniquely by

$$(5.1) \qquad \bar{x}(s) = (\phi^*)'(\bar{\lambda}^T a(s) - z_0(s))$$

(by Theorem 4.1), and the singular part $\bar{\nu}$ can be chosen arbitrarily, provided that it satisfies the optimality conditions (OCP_E^1). In order to compute a solution we first solve the dual problem. This is a concave maximization problem, and the objective function is continuously differentiable on the interior of its domain, so a wide variety of standard numerical techniques may be applied.

The continuous part of the primal solution $\bar{x}$ is now given by (5.1), while the singular part is any measure $\bar{\nu}$ which is singular with respect to ρ and satisfies

$$(5.2) \qquad \text{support } (\bar{\nu}^+) \subset \{s \in S \mid \bar{\lambda}^T a(s) - z_0(s) = q\},$$

$$(5.3) \qquad \text{support } (\bar{\nu}^-) \subset \{s \in S \mid \bar{\lambda}^T a(s) - z_0(s) = p\}, \text{ and}$$

$$(5.4) \qquad \int_S a_i(s) \, d\nu = b_i - \int_S a_i(s) \bar{x}(s) \, d\rho \quad \text{for } i = 1, \ldots, n.$$

(We know there exists a solution.) It can be shown using techniques analogous to those used in semi-infinite programming (see, for example, [1]) that we can restrict attention to $\bar{\nu}$ for which $\bar{\nu}^+$ and $\bar{\nu}^-$ are supported on $n + 1$ points in S (see [8, § 5]). Conditions (5.2)–(5.4) then form a semi-infinite linear problem for which standard numerical techniques are available (see, for example, [23]).

The idea of a Tchebycheff system will be useful for our discussion of uniqueness. Working on a fixed, finite interval S in $\mathbb{R}$, for a continuous function f we denote by $\tilde{Z}(f)$ the number of distinct zeros of f, counting twice the zeros in the interior of S at which f does not change sign. The following result [30, Thm. I.4.2] essentially characterizes Tchebycheff systems.

THEOREM 5.5. *If $\{a_1, \ldots, a_n\}$ is a Tchebycheff system on S then $\tilde{Z}(\lambda^T a) \leq n - 1$, provided that λ is nonzero.*

COROLLARY 5.6. *Suppose $\{a_1, \ldots, a_n\}$ is a Tchebycheff system on S, $a_1 \equiv 1$, $p \leq \lambda^T a(s) \leq q$ for all s in S, and $\lambda^T a$ is not identically p or q. Then we have $|\{s \in S \mid \lambda^T a(s) = p$ or $q\}| \leq n$.*

Proof. Denote the number of endpoints of S at which $\lambda^T a(s) = p$ or q by n_p^e and n_q^e, respectively, and the number of interior points by n_p^i and n_q^i, respectively. Then

$n_p^e + n_q^e \leqq 2$, and by Theorem 5.5,

$$n_p^e + 2n_p^i = \tilde{Z}(\lambda^T a - p) \leqq n - 1,$$

and

$$n_q^e + 2n_q^i = \tilde{Z}(q - \lambda^T a) \leqq n - 1.$$

Adding gives $2(n_p^e + n_p^i + n_q^e + n_q^i) \leqq 2n$, from which the result follows.   □

THEOREM 5.7. *Suppose the Primal and Dual Constraint Qualifications hold, ϕ is essentially strictly convex, S is a finite, closed interval in $\mathbb{R}$, $\{a_1, \ldots, a_n\}$ is a Tchebycheff system on S, $a_1 \equiv 1$, and $z_0 \equiv 0$. Then the extended primal problem (P_E) (or (P_E^1)) has a unique optimal solution.*

Proof. If $\bar{\lambda}$ is a dual optimal solution then, from the above discussion, the absolutely continuous part of any extended primal optimal solution $(\bar{x}, \bar{\nu})$ is given uniquely by $\bar{x} := (\phi^*)'(\bar{\lambda}^T a)$, and since ϕ^* is essentially smooth, $\bar{\lambda}^T a$ is not identically p or q. The singular part $\bar{\nu}$ must satisfy (5.2)-(5.4), and Corollary 5.6 shows that it is supported on at most n points, determined by $\bar{\lambda}$. The set of linear equations resulting from (5.4) then has a unique solution for $\bar{\nu}$ since $\{a_1, \ldots, a_n\}$ is a Tchebycheff system.   □

Analogous results could be proved when $\{a_1, \ldots, a_n\}$ is a periodic Tchebycheff system, as in the trigonometric moment problem.

Examples. We begin by discussing two examples from [6]. The first is a simple semi-infinite linear program, where $\phi(u) := u$ if $u \geqq 0$ and $+\infty$ otherwise (note this is not an affine function):

$$\text{inf} \qquad \int_0^1 x(s) \, ds,$$

$(E1)$ $\qquad$ subject to $\qquad \displaystyle\int_0^1 sx(s) \, ds = 1,$

$$0 \leqq x \in L_1[0, 1].$$

The dual problem is

$(E1^*)$ $\qquad$ $\displaystyle\sup\left\{ \lambda - \int_0^1 \delta(\lambda \,|\, (-\infty, 1]) \, ds \,\Big|\, \lambda \in \mathbb{R} \right\},$

where $\delta(\cdot \,|\, C)$ is the indicator function of C. The Primal and Dual Constraint Qualifications are both satisfied, the unique dual optimal solution is $\bar{\lambda} = 1$, and both problems have value 1, but the primal value is not attained. The extended primal problem is

$$\text{inf} \qquad \int_0^1 x(s) \, ds + \nu[0, 1],$$

$(E1_E)$ $\qquad$ subject to $\qquad \displaystyle\int_0^1 sx(s) \, ds + \int_0^1 s \, d\nu = 1,$

$$0 \leqq x \in L_1[0, 1], \quad 0 \leqq \nu \in M[0, 1], \quad d\nu \perp ds,$$

and our results show that the unique optimal solution is a unit point mass at 1, giving the value 1.

The second example uses the objective function $\phi(u) := 1/u$ if $u > 0$ and $+\infty$ otherwise:

$$\inf \quad \int_0^{2\pi} (1/x(s))\, ds,$$

(E2) $\qquad$ subject to $\quad \int_0^{2\pi} \sin(s)x(s)\, ds = 1,$

$$0 \leqq x \in L_1[0, 2\pi].$$

The dual problem is

(E2*) $\qquad \sup\left\{\lambda - \int_0^{2\pi} \phi^*(\lambda \sin(s))\, ds \,\Big|\, \lambda \in \mathbb{R}\right\},$

where $\phi^*(v) = -2(-v)^{1/2}$ if $v \leqq 0$ and $+\infty$ otherwise. The only dual feasible solution is $\bar{\lambda} = 0$, which is therefore optimal, with value zero. Note that, although the Primal Constraint Qualification is satisfied, the value of the primal problem is zero and is unattained. Furthermore, the extended primal problem (not considered in [6]) is

$$\inf \quad \int_0^{2\pi} (1/x(s))\, ds,$$

(E2$_E$) $\qquad$ subject to $\quad \int_0^{2\pi} \sin(s)x(s)\, ds + \int_0^{2\pi} \sin(s)\, d\nu = 1,$

$$0 \leqq x \in L_1[0, 2\pi], \quad 0 \leqq \nu \in M[0, 2\pi], \quad d\nu \perp ds,$$

and this problem also does not attain its value of zero. The reason is, of course, that the Dual Constraint Qualification is not satisfied.

The final two examples are particularly interesting since the objective function is the Burg entropy, which is widely used in practice. The first problem is extremely simple, and demonstrates the importance of the assumption that the constraint functions are Lipschitz in Theorem 4.9. We consider the primal problem

$$\inf \quad \int_0^1 -\log(x(s))\, ds,$$

$$\text{subject to} \quad \int_0^1 x(s)\, ds = 1,$$

(E3)

$$\int_0^1 s^{1/2} x(s)\, ds = \alpha,$$

$$0 \leqq x \in L_1[0, 1],$$

where $\alpha \in (0, 1)$. The dual problem is

$$\sup \quad \lambda_0 + \alpha\lambda_1 + \int_0^1 [1 + \log(-\lambda_0 - \lambda_1 s^{1/2})]\, ds,$$

(E3*) $\qquad$ subject to $\quad -\lambda_0 \leqq 0, \quad \lambda_0 + \lambda_1 \leqq 0,$

$$\lambda_0, \lambda_1 \in \mathbb{R}$$

(where the extra constraint is implicit in the objective function).

It is straightforward to check that both the Primal and Dual Constraint Qualifications hold. We would expect, from the form of the constraints in (E3), that

the weight in the optimal density will shift from right to left in the interval $[0, 1]$ as we decrease α in $(0, 1)$, and this is indeed what happens. We know from (4.2) that the absolutely continuous part of any extended primal solution is given by $\bar{x}(s) :=$ $(-\bar{\lambda}_0 - \bar{\lambda}_1 s^{1/2})^{-1}$, where $\bar{\lambda}_0$ and $\bar{\lambda}_1$ are dual optimal, and it may be checked that for α in $(\frac{1}{2}, 1)$, $\bar{\lambda}_0 + \bar{\lambda}_1 s^{1/2} > 0$ on $[0, 1]$, so $\bar{x}$ is the unique primal solution. At $\alpha = \frac{2}{3}$ the optimal solution $\bar{x}(s) \equiv 1$, and as α decreases the weight shifts to the left until at $\alpha = \frac{1}{2}$ the unique optimal solution is $\bar{x}(s) = (1/2)s^{-1/2}$. For α in $(0, \frac{1}{2}]$ it can be checked that the dual optimum is $\bar{\lambda}_0 = 0$, $\bar{\lambda}_1 = -\alpha^{-1}$, and $\bar{x}$ is no longer feasible for $(E3)$.

What has happened is that part of the optimal solution has condensed into a point mass at the origin, as would be shown by discretization. The extended primal problem is

$$\inf \quad \int_0^1 -\log (x(s)) \, ds,$$

$$\text{subject to} \quad \int_0^1 x(s) \, ds + \nu[0, 1] = 1,$$

$(E3_E)$

$$\int_0^1 x(s)s^{1/2} \, ds + \int_0^1 s^{1/2} \, d\nu = \alpha,$$

$$0 \leq x \in L_1[0, 1], \quad 0 \leq \nu \in M[0, 1], \quad d\nu \perp ds,$$

and this has a unique optimal solution for α in $[0, \frac{1}{2}]$, $\bar{x}(s) = \alpha s^{-1/2}$, and $\bar{\nu}$ a point mass of $(1 - 2\alpha)$ at the origin.

The last example was presented in [40] to demonstrate the problems associated with the Burg entropy for three-dimensional density reconstruction, and it has also been discussed in [52] and [14]. The underlying set S is the unit cube in $\mathbb{R}^3$, $[0, 1]^3$, with Lebesgue measure ds, and the problem has simple trigonometric moment constraints:

$$\inf \quad \int_S -\log (x(s)) \, ds,$$

$$\text{subject to} \quad \int_S x(s) \, ds = 1,$$

$(E4)$

$$\int_S x(s) \cos (2\pi s_i) \, ds = \alpha, \quad \text{for } i = 1, 2, 3,$$

$$0 \leq x \in L_1(S),$$

where $\alpha \in [0, 1)$. The dual problem is

$$\sup \quad \lambda_0 + \alpha \sum_1^3 \lambda_i + \int_S \left[1 + \log \left(-\lambda_0 - \sum_1^3 \lambda_i \cos (2\pi s_i) \right) \right] ds,$$

$(E4^*)$ subject to $-\lambda_0 \geq \sum_1^3 |\lambda_i|,$

$$\lambda_0, \lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$$

(where again the extra constraint is implicit in the objective function).

Straightforward calculations will now verify the following assertions. The Primal and Dual Constraint Qualifications both hold, and the unique dual optimal solution has the form $(\bar{\lambda}_0(\alpha), \bar{\lambda}(\alpha), \bar{\lambda}(\alpha), \bar{\lambda}(\alpha))$ for each α. Thus the absolutely continuous part $\bar{x}_\alpha$ of any extended primal solution $(\bar{x}_\alpha, \bar{\nu}_a)$ is given uniquely (see (4.2)) by

$$(5.8) \qquad \bar{x}_\alpha(s) := \left(-\bar{\lambda}_0(\alpha) - \bar{\lambda}(\alpha) \sum_1^3 \cos(2\pi s_i) \right)^{-1}.$$

The interesting phenomenon is to observe what happens as α increases. The trigonometric polynomial in (5.8) is strictly positive for small α, and $\bar{x}_\alpha$ is feasible for the primal problem $(E4)$, as is the unique optimal solution. As α approaches a certain critical value $\bar{\alpha}$, the minimum value of the polynomial decreases to zero, until the point when $\alpha = \bar{\alpha}$, where the polynomial has a zero when $s_i = 0$ or 1 for $i = 1, 2, 3$. The unique optimal solution of the primal is still $\bar{x}_{\bar{\alpha}}$. However, as α increases past $\bar{\alpha}$ the character of the solution changes. For $\alpha \in (\bar{\alpha}, 1)$ the unique dual optimal solution is $(1/[1-\alpha]) \times (-1, \frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, so (5.8) becomes

$$(5.9) \qquad \bar{x}_\alpha(s) = \frac{1-\alpha}{(1 - \frac{1}{3}\sum_1^3 \cos(2\pi s_i))},$$

which is no longer primal feasible.

The extended primal problem is

$$\inf \qquad \int_S -\log(x(s))\, ds,$$

$$(E4_E) \qquad \text{subject to} \qquad \int_S x(s)\, ds + \nu(S) = 1,$$

$$\int_S x(s) \cos(2\pi s_i)\, ds + \int_S \cos(2\pi s_i)\, d\nu = \alpha \quad \text{for } i = 1, 2, 3,$$

$$0 \leq x \in L_1(S), \quad 0 \leq \nu \in M(S), \quad d\nu \perp ds.$$

Our results show that the absolutely continuous part of the optimal solution is given by (5.9), and the optimality conditions ensure that the singular part $\bar{\nu}_\alpha$ is supported on the zeros of the denominator of (5.9), namely, $s_i = 0$ or 1 for $i = 1, 2, 3$. These points are equivalent up to periodicity, so essentially the unique singular part is a point mass at the origin with weight $((\alpha - \bar{\alpha})/(1 - \bar{\alpha}))$.

The critical value of α is given by

$$(5.10) \qquad \bar{\alpha} = 1 - \left(\int_S \left(1 - \frac{1}{3}\sum_1^3 \cos(2\pi s_i) \right)^{-1} ds \right)^{-1} \approx .34.$$

(This integral is actually Green's integral for the cubic lattice, and has the closed form $\Gamma(1/24)\Gamma(5/24)\Gamma(7/24)\Gamma(11/24)(6)^{1/2}/32\pi^3$; see [24].) In the case discussed in [52] and [40] an optimal solution was proposed informally for the case $\alpha = .5$; our solution agrees exactly.

As a final comment, numerous different measures of entropy ϕ have appeared in the literature. A survey of some of these, with their conjugates and associated p and q, may be found in [6].

REFERENCES

[1] E. J. ANDERSON AND P. NASH, *Linear Programming in Infinite-Dimensional Spaces*, Wiley-Interscience, Chichester, 1987.

[2] J.-P. AUBIN AND H. FRANKOWSKA, *Set-Valued Analysis*, Birkhäuser, Boston, 1990.

[3] A. BEN-TAL, J. M. BORWEIN, AND M. TEBOULLE, *A dual approach to multi-dimensional L_p spectral estimation problems*, SIAM J. Control Optim., 26 (1988), pp. 985–996.

[4] M.-F. BIDAUT, *Un problème de contrôle optimal à fonction coût en norme L_1*, Comptes Rend. Acad. Sci. Paris, Sér. A, 281 (1975), pp. 273–276.

[5] J. M. BORWEIN AND A. S. LEWIS, *Convergence of best entropy estimates*, SIAM J. Optim., 1 (1991), pp. 191–205.

[6] ———, *Duality relationships for entropy-like minimization problems*, SIAM J. Control Optim., 29 (1991), pp. 325–338.

[7] ———, *On the convergence of moment problems*, Trans. Amer. Math. Soc., 325 (1991), pp. 249–271.

[8] ———, *Partially finite programming in L_1: Entropy maximization*, Tech. Rep. CORR 91-05, Dept. of Combinatorics and Optimization, Univ. of Waterloo, Ontario, Canada, 1991.

[9] ———, *Partially finite convex programming*, Math. Programming, 57 (1992), pp. 15–83 (in two parts).

[10] J. M. BORWEIN AND H. WOLKOWICZ, *A simple constraint qualification in infinite dimensional programming*, Math. Programming, 35 (1986), pp. 83–96.

[11] P. BORWEIN AND A. S. LEWIS, *Moment-matching and best entropy estimation*. Tech. Rep. CORR 91-03, Univ. of Waterloo, Ontario, Canada, 1991; J. Math. Anal. Appl., submitted.

[12] J. P. BURG, *Maximum entropy spectral analysis*, Paper presented at 37th Meeting of the Society of Exploration Geophysicists, Oklahoma City, OK, 1967.

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

[14] A. DECARREAU, D. HILHORST, C. LEMARÉCHAL, AND J. NAVAZA, *Dual methods in entropy maximization: Application to some problems in crystallography*, SIAM J. Optim., 2 (1992), pp. 173–197.

[15] J. A. EDWARD AND M. M. FITELSON, *Notes on maximum entropy processing*, IEEE Trans. Inform. Theory, 19 (1973), pp. 232–234.

[16] I. EKELAND AND R. TEMAN, *Convex Analysis and Variational Problems*, North-Holland, Amsterdam, 1976.

[17] G. J. ERICKSON AND C. R. SMITH, EDS., *Maximum-Entropy and Bayesian Methods in Science and Engineering*, Vols. I and II, Kluwer, Dordrecht, The Netherlands, 1988.

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

[19] F. GAMBOA, *Methode du Maximum d'Entropie sur la Moyenne et Applications*, Ph.D. thesis, Univ. Paris Sud, Centre d'Orsay, France, 1989.

[20] F. GAMBOA AND E. GASSIAT, *Extension of the maximum entropy method on the mean and a Bayesian interpretation of the method*, preprint, Statistics Laboratory, Univ. of Orsay, France, 1991.

[21] ———, *M.E.M. techniques for solving moment problems*, preprint, Statistics Laboratory, Univ. of Orsay, France, 1991.

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

[23] K. GLASHOFF AND S.-A. GUSTAFSON, *Linear Optimization and Approximation*, Springer-Verlag, New York, 1983.

[24] M. L. GLASSER AND I. J. ZUCKER, *Extended Watson integrals for the cubic lattices*, Proc. Nat. Acad. Sci., USA, 74 (1977), pp. 1800–1801.

[25] R. K. GOODRICH AND A. STEINHARDT, *L_2 spectral estimation*, SIAM J. Appl. Math., 46 (1986), pp. 417–428.

[26] L. D. IRVINE, S. P. MARIN, AND P. W. SMITH, *Constrained interpolation and smoothing*, Constructive Approx., 2 (1986), pp. 129–151.

[27] E. T. JAYNES, *Prior probabilities*, IEEE Trans. 4 (1968), pp. 227–241.

[28] R. W. JOHNSON AND J. E. SHORE, *Which is the better entropy expression for speech processing: $-S \log S$ or $\log S$?*, IEEE Trans. Acoust. Speech Signal Process., 32 (1984), pp. 129–137.

[29] L. K. JONES AND V. TRUTZER, *Computationally feasible high-resolution minimum distance procedures which extend the maximum-entropy method*, Inverse Problems, 5 (1989), pp. 749–766.

[30] S. KARLIN AND W. J. STUDDEN, *Tchebycheff Systems: With Applications in Analysis and Statistics*, Wiley-Interscience, New York, 1966.

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

[32] S. W. LANG AND J. H. MCCLELLAN, *Multidimensional MEM spectral estimation*, IEEE Trans. Acoust. Speech Signal Process., 30 (1984), pp. 880–887.

[33] A. S. LEWIS, *Pseudo-Haar functions and partially-finite programming*, manuscript.

[34] ———, *The convergence of entropic estimates for moment problems*, in Workshop/Miniconference on Functional Analysis/Optimization, S. Fitzpatrick and J. Giles, eds., Centre for Mathematical Analysis, Australian National University, Canberra, 1989, pp. 100–115.

[35] D. M. LIN AND E. K. WONG, *A survey on the maximum entropy method and parameter spectral estimation*, Phys. Rep., 193 (1990), pp. 41–135.

[36] J. H. MCCLELLAN AND S. W. LANG, *Multi-dimensional MEM spectral estimation*, in Spectral Analysis and Its Use in Underwater Acoustics, Institute of Acoustics, Imperial College, London, U.K., 1982, pp. 10.1–10.8.

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

[38] C. A. MICCHELLI, P. W. SMITH, J. SWETITS, AND J. D. WARD, *Constrained L_p approximation*, Constructive Approx., 1 (1985), pp. 93–102.

[39] J. NAVAZA, *The use of non-local constraints in maximum-entropy electron density reconstruction*, Acta Crystallographica, A42 (1986), pp. 212–223.

[40] R. NITYANANDA AND R. NARAYAN, *Maximum entropy image reconstruction—a practical non-information-theoretic appraoch*, J. Astrophys. Astron., 3 (1982), pp. 419–450.

[41] P. R. OLIVEIRA, *Contrôle de processus de vieillissement*, Ph.D. thesis, La Faculté des Sciences de Paris IX, France, 1977.

[42] M. POWELL, *Karmarkar's algorithm for semi-infinite programming*, Tech. Rep., University of Cambridge, Cambridge, UK, 1990.

[43] R. T. ROCKAFELLAR, *Duality and stability in extremum problems involving convex functions*, Pacific J. Math., 21 (1967), pp. 167–187.

[44] ———, *Integrals which are convex functionals*, Pacific J. Math., 24 (1968), pp. 525–539.

[45] ———, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

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

[47] ———, *Conjugate Duality and Optimization*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1974.

[48] W. RUDIN, *Real and Complex Analysis*, McGraw-Hill, New York, 1966.

[49] ———, *Functional Analysis*, McGraw-Hill, New York, 1973.

[50] A. SEGHIER, *Reconstruction de la densité spectrale par maximum d'entropie cas d-dimensionnel*, Comptes Rend. Acad. Sci. Paris, Sér. I, 305 (1987), pp. 517–520.

[51] J. SKILLING, ED., *Maximum-Entropy and Bayesian Methods*, Kluwer, Dordrecht, the Netherlands, 1989.

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

[53] C. R. SMITH AND G. J. ERICKSON, EDS., *Maximum-Entropy and Bayesian Spectral Analysis and Estimation Problems*, Kluwer, Dordrecht, the Netherlands, 1987.

[54] C. R. SMITH AND W. T. GANDY, EDS., *Maximum-Entropy and Bayesian Methods in Inverse Problems*, Kluwer, Dordrecht, the Netherlands, 1985.

[55] M. TODD, *Interior point methods for semi-infinite programming*, Tech. Rep., Cornell Univ., Ithaca, NY, 1991.

[56] J. W. WOODS, *Two-dimensional Markov spectral estimation*, IEEE Trans. Inform. Theory, 22 (1976), pp. 552–559.