On Entropy Maximization via Convex Programming

Jonathan M. Borwein
Mark A. Limber

Centre for Experimental and Construtive Mathematics
Department of Mathematics and Statistics
Simon Fraser University
Burnaby, B.C. V5A 1S6 CANADA

June 5, 1996

Key Words: Dual convex programs, Fenchel duality, integral functionals, interval bounds, maximum entropy estimates, signal reconstruction.

Abstract

Maximum entropy methods have received much attention in recent years. Many of the published derivations of characterizations of solutions to these infinite dimensional concave programming problems have contained subtle errors that in no way diminish the applications, and in many cases yield the correct solutions, but prompt this mathematical note. It is our intent to point out pitfalls in working with certain infinite dimensional function spaces in an optimization conmbox and offer tools to get around these problems.

1 Introduction

We shall study a fairly general constrained optimization problem that arises in a variety of applications. We begin by stating the problem in some generality, giving specific examples from the literature, and giving a standard (incorrect) argument that characterizes the solutions to this problem. The intent of this note is point out the difficulties in correctly solving this optimization problem and give some mathematically sound results.

1.1 The Problem

In the problems of entropy maximization, spectral estimation, signal restoration, concave interpolation, crystallography, information theory and other disciplines, there is a common theme of maximizing a concave integral function subject to a finite number of linear equality, and perhaps interval, constraints. In this note we shall study the equivalent minimization of the corresponding convex integral functional.

In the following T is a finite measure space with a regular measure μ. For example, $T = [-\pi, \pi]$ with Lebesgue measure.

One is typically given a finite number of measured "moments" b_k of the form

$$\int_T x_*(t)\psi_k(t)\, d\mu(t) = b_k, \qquad k = 1, \ldots, n \tag{1.1}$$

from an unknown function x_* with known functions ψ_k. These functions could be the complex exponentials in the classical moment problem, t^{k-1} in the Hausdorff moment problem, or some other measurement made physically, such as in the convex interpolation case [9], [22] or image reconstruction from projections [11], [12].

Take $A : X \to I\!R^n$ to be the continuous linear operator defined by

$$(Ax)_k = \int_T x_*(t)\psi_k(t) d\mu(t).$$

Then Eq. (1.1) becomes $Ax = b$ where $b \in I\!R^n$ is the vector $b = (b_1, \ldots, b_n)$.

Out of the infinity of feasible functions, those that satisfy Eq. (1.1), one would like some measure of which is the best. This is commonly done by minimizing a convex cost functional of the form

$$I_\phi(x) \triangleq \int_T \phi(x(t))\, d\mu(t) \tag{1.2}$$

where $\phi : I\!R \to (-\infty, +\infty]$ is convex, subject to the constraints in Eq. (1.1). One might also have physical constraints on the feasible functions x of the form

$$\alpha(t) \leq x(t) \leq \beta(t) \quad \text{a.e.} \tag{1.3}$$

where α and β are allowed to be infinite to include cases like $0 \leq x(t)$.

By adding the indicator function of the interval $[\alpha, \beta]_X$ we can assume that the interval constraint is incorporated into the integrand:

$$\phi_I(x,t) = \begin{cases} \phi(x(t)) & \text{if } \alpha(t) \leq x(t) \leq \beta(t) \\ +\infty & \text{else.} \end{cases} \tag{1.4}$$

That is,

$$I_{\phi_I}(x) \triangleq \begin{cases} \int_T \phi(x)\, dt & \text{if } \alpha \leq x \leq \beta \text{ a.e.} \\ +\infty & \text{else.} \end{cases} \tag{1.5}$$

We shall write ϕ instead of ϕ_I and understand that the interval constraints are included.

These problems fit the following form: find a feasible $\hat{x}$ that solves

$$V(P) \stackrel{\triangle}{=} \inf \{I_\phi(x) : Ax = b\}. \tag{P}$$

This is a *partially finite convex programming problem.* Implicit in this formulation is that there is a function space X from which our feasible functions are taken. The choice of this space is important in that reasonable solutions may be lost if the wrong space is used. In addition, the norm on the space, or the topology, can cause subtle mathematical difficulties.

1.2 Examples

For example, to minimize the energy of the signal $x(t)$ one might use

$$\phi(u) = \begin{cases} \frac{1}{2}u^2 & \text{if } u \geq 0 \\ +\infty & \text{else} \end{cases} \tag{1.6}$$

Here $\alpha \equiv 0$ and $\beta \equiv +\infty$. This is found, for example, in power spectrum estimation [14], signal reconstruction [1], [11], [18], and convex interpolation [9], [22]. Implicit here is the space $X = L^2(T)$. Or we might use the (negative) Boltzmann-Shannon entropy

$$\phi(u) = \begin{cases} u\ln(u) & \text{if } u > 0 \\ 0 & \text{if } u = 0 \\ +\infty & \text{if } u < 0 \end{cases} \tag{1.7}$$

or the (negative) Burg entropy

$$\phi(u) = \begin{cases} -\ln(u) & \text{if } u > 0 \\ +\infty & \text{if } u \leq 0 \end{cases} \tag{1.8}$$

as found in crystallography [7], [24], information theory [16], spectral estimation [1], [11], [17], [18], general engineering [28], and the references therein. The choice of space is not so obvious in these cases.

Given a convex function $\varphi : I\!R \to (-\infty, +\infty]$ differentiable on the interior of it's domain (as defined in Section 3.1) and a fixed number u_0, one can form a <u>directed distance</u> entropy

$$\phi(u) = \varphi(u) - \varphi(u_0) - \varphi'(u_0)(u - u_0). \tag{1.9}$$

This gives a measure of the distance from a prior function x_0 to an approximated function x as suggested in [12], [17], and [23]. Obviously, it suffices to consider

$$\phi(u) = \varphi(u) - \varphi'(u_0)(u). \tag{1.10}$$

In the case φ is the Boltzmann-Shannon entropy, this has been called the cross-entropy.

1.3 A common solution

A common technique to solve (P) is the following:

1. Let $\overline{x}(t)$ denote the solution to (P).

2. Let $\lambda = \{\lambda_k\}_{k=1}^n$ be Lagrange multipliers for the linear constraints $Ax = b$. Let

$$L_\lambda(x) = \int_T \phi(x(t))\, d\mu(t) - \sum_{k=1}^n \lambda_k \left(\int_T \psi_k(t) x(t)\, d\mu(t) - b_k \right). \tag{1.11}$$

3. At $\overline{x}$ the Fréchet derivative of L_λ is zero:

$$\frac{\mathrm{d}}{\mathrm{dx}} L_\lambda|_{\overline{x}} = 0. \tag{1.12}$$

4. Write this out differentiating under the integral to obtain

$$0 = \int_T \left[\phi'(\overline{x}(t)) - \sum_{k=1}^n \lambda_k \psi_k(t) \right] h(t)\, d\mu(t) \quad \forall h. \tag{1.13}$$

5. Therefore the integrand must be zero and thus

$$\overline{x}(t) = (\phi')^{-1} \left(\sum_{k=1}^n \lambda_k \psi_k(t) \right). \tag{1.14}$$

For example, if ϕ is the cross-entropy, $\phi(u) = u \ln(u) - u$, then the solution in Eq. (1.14) would be

$$\overline{x}(t) = \exp \left(\sum_{k=1}^n \lambda_k \psi_k(t) \right). \tag{1.15}$$

We shall see that this is usually the solution, but each step in the above derivation is suspect and many are wrong without certain assumptions.

In much of the literature the wrong space X is used or invalid assumptions of continuity and interiority are made that make the above derivation seem correct. We shall give conditions under which many of the results in the literature can be derived rigorously and point out the pitfalls many fall into. We propose an appropriate ambient space X, characterize solutions to (P) using an extension of Fenchel duality, and give some numerical results. We include references to some mathematical papers that more fully address the issues in this note.

2 The Choice of Space

There are three basic issues that come up in decided upon an appropriate space to work over. The first is to find the "largest" of the natural choices. The second is that of attainment: over which spaces are we guaranteed to have solutions to (P)? The final issue is that the interior of the positive cone is frequently empty. This relates to the continuity and differentiability of the objective function. We shall see that a proper resolution of these issues make the standard (differentiable) optimization theory inapplicable.

2.1 The space $L^1(T)$

The most common function spaces used in the literature are $C(T)$, the set of continuous functions on T with the supremum or Chebyshev norm:

$$\|x\|_\infty = \sup\{|x(t)| : t \in T\}, \tag{2.16}$$

$L^\infty(T)$, the set of essentially bounded measurable functions on T with essentially the same norm as $C(T)$, and the $L^p(T)$ spaces $1 \le p < \infty$, the set of p^{th} power integrable measurable functions on T with norm

$$\|x\|_p = \left(\int_T |x(t)|^p \, d\mu(t) \right)^{\frac{1}{p}}. \tag{2.17}$$

A common choice is $p = 2$, the set of functions with finite energy.

The most compelling reason why $L^1(T)$ is the appropriate space to use is based on the following observation: of these spaces, $L^1(T)$ is in some sense the "largest". In fact, since $\mu(T) < \infty$, for any $p > 1$ we have (as sets)

$$C(T) \subseteq L^\infty(T) \subseteq L^p(T) \subseteq L^1(T) \tag{2.18}$$

usually with strict inclusion at each step. To solve (P) over $C(T)$ leaves open the possibility that there are "better" discontinuous functions; to solve (P) over $L^\infty(T)$ leaves open the possibility that there are "better" functions with finite energy; and so on. Even though the set of continuous functions on T is dense in $L^1(T)$, it is possible in principle that $V(P)$ over $C(T)$ is attained, but is larger than $V(P)$ over $L^1(T)$.

We shall see that the <u>norm dual</u> space X^* of X, the set of all continuous linear functionals on X, is important in the derivation of the solution. In the case $X = C(T)$ the dual is the set $M(T)$ all regular Borel measures on T[10], and when T is a finite interval this reduces to $NBV(T)$, all normalized functions of bounded variation on T[19]. In case $X = L^\infty(T)$, X^* is the space of all finitely additive measures[10], but in case $X = L^1(T)$, X^* is just $L^\infty(T)$[13]. This makes certain calculations easier, and in many cases will guarantee that solutions to (P) formulated in $L^1(T)$ actually lie in $L^\infty(T)$ or in $C(T)$ or better.

2.2 The attainment difficulty

To attempt the steps given in section 1.3 one must first show that an optimal feasible vector exists. This can be very difficult to justify depending on the ambient space X, and in some cases just not true. Although it is true that attainment can be substantiated by weak compactness arguments in $L^p(T)$ $(1 < p < \infty)$ for many but not all entropies, attainment is not so easy to justify for any of the entropies in $C(T)$ or $L^\infty(T)$.

For example, the lower level sets of the Boltzmann-Shannon entropy

$$\left\{x \in L^\infty(T) : \int_T x(t)\ln(x(t))\,d\mu(t) \leq M\right\} \tag{2.19}$$

are not even bounded in $L^\infty(T)$. Since compact sets (in any of the norm, weak or weak-$*$ topologies) must be norm bounded[15], this implies that the lower level sets are not compact and thus any sort of coercivity argument would fail to guarantee minima exist.

Even with some compactness the existence of an optimal solution to (P) cannot be insured. For example consider

$$\inf\left\{\int_0^1 x(t)\,dt : \int_0^1 tx(t)\,dt = 1,\ 0 \leq x \text{ a.e.},\ \ x \in L^1(T)\right\}. \tag{2.20}$$

It has been shown in [2] that the optimal value for this problem is not attained.

2.3 The interiority/continuity difficulty

One difficulty with working in any of the L^p spaces when the measure μ is non-atomic[1], $1 \leq p < \infty$, is that the positive cone,

$$C = \{x \in L^p(T) : 0 \leq x \text{ a.e.}\} \tag{2.21}$$

has *empty interior*. To see this, let $x \geq 0$ a.e. For $E \subseteq T$ small enough we define x_1 by

$$x_1(t) = \begin{cases} -1 & \text{if } t \in E \\ x(t) & \text{if } t \in E^c. \end{cases} \tag{2.22}$$

Then $\int_T |x - x_1|^p$ can be made small by choosing E small enough, but $x_1 \notin C$. From this it follows that, for example, the Boltzmann-Shannon entropy is $+\infty$ on a dense set in $L^p(T)$ and thus not continuous anywhere and certainly not differentiable (as some have assumed [12], [23]). So even if attainment were guaranteed, the Lagrange multiplier argument outlined in section 1.3 fails in $L^p(T)$. In particular, steps (3) and (4) are unjustified. Even when continuity can be assured, some care is needed to interchange the derivative and integral in step (4).

[1]For example, when μ is Lebesgue measure on an interval.

For this reason, many authors have chosen to work in $C(T)$ or $L^\infty(T)$ where the positive cone does have interior [1], [6], [18], [21]. We have seen that in these spaces there remains the difficulty of proving that the infimum is attained and the solution over these spaces is not as strong as over $L^1(T)$.

Using duality arguments discussed below, the value of (P) is the same as the (attained) value of a concave maximization problem where differentiation can be mathematically justified with some careful convex analysis [4], [5].

3 The Characterization of Solutions

We now give reasonable conditions under which the form of the solutions to (P) can be rigorously obtained. We discuss the classical theory of Fenchel duality, show why it doesn't apply and give a modern resolution.

3.1 Definitions

To properly address problem (P) we need some definitions from convex analysis. Let X be a normed linear space, X^* its norm dual and for $x \in X$ and $x^* \in X^*$ denote the application of x^* on x by $\langle x, x^*\rangle$. A function $f : X \to (-\infty, +\infty]$ is <u>convex</u> provided

$$f(tx + (1-t)y) \le tf(x) + (1-t)f(y) \qquad \forall t \in (0,1), \quad x, y \in X. \tag{3.23}$$

The <u>domain</u> of f is $\mathrm{dom}(f) = \{x : f(x) < +\infty\}$ and f is <u>proper</u> if $\mathrm{dom}(f) \neq \emptyset$. Similarly $g : Y \to [-\infty, \infty)$ is <u>concave</u> if $-g$ is convex and $\mathrm{dom}(g) = \mathrm{dom}(-g)$. Define $\mathrm{cont}(f)$ to be the set of points where f is continuous. The Fenchel <u>convex conjugate</u> of the proper convex function f, $f^* : X^* \to (-\infty, +\infty]$ is defined by

$$f^*(x^*) = \sup_{x \in X}\{\langle x, x^*\rangle - f(x)\} \tag{3.24}$$

and the concave conjugate is $g_*(x^*) = (-g)^*(-x^*)$. Note that for $\hat{f}(x) \triangleq f(x) - \langle x, x_1^*\rangle$ a linear translate of f, the conjugate is just

$$\hat{f}^*(x^*) = f^*(x_1^* + x^*). \tag{3.25}$$

This is useful when finding the conjugate of the directed distance entropies. Simple calculations yield the conjugates found in Table 1.

Given a convex set C in $\mathbb{R}^n$, the <u>relative interior</u> of C, $\mathrm{ri}(C)$ is the interior of C considered as a subset of the smallest (closed) affine subspace containing C.

Entropy	$\phi(u)$	$\phi^*(v)$
L^2 Norm	$\frac{1}{2}u^2$	$\frac{1}{2}v^2$
Boltzmann-Shannon	$u\ln(u)$	$\exp(v-1)$
Burg	$-\ln(u)$	$-1-\ln(-v)$
Cosh	$\cosh(u)$	$\operatorname{arcsinh}(v)-\sqrt{1+v^2}$
Directed Distance	$\varphi(u)-\varphi'(u_0)u$	$\varphi^*(v+\varphi'(u_0))$
Fermi-Dirac	$u\ln(u)+(1-u)\ln(1-u)$	$\ln(1+\exp(v))$

Table 1: Entropies with their conjugates.

3.2 Fenchel duality

For reference we state a classical form of the Fenchel duality theorem.

Theorem 3.1 *Let X and Y be linear normed spaces, $f : X \to (-\infty, +\infty]$ proper and convex, $g : Y \to [-\infty, \infty)$ proper and concave, and $A : X \to Y$ continuous and linear. Define*

$$V(P) \triangleq \inf\{f(x) - g(Ax)\}. \tag{3.26}$$

If the following constraint qualification holds,

$$A[\mathrm{dom}(f)] \cap \mathrm{cont}(g) \neq \emptyset, \tag{FCQ}$$

and if we let

$$V(D) \triangleq \max\{g_*(y^*) - f^*(A^T y^*) : y^* \in Y^*\} \tag{3.27}$$

then $V(P) = V(D)$ and $V(D)$ is attained when finite.

See Rockafellar[27], for example. When Y is finite dimensional, we can replace (FCQ) with

$$\mathrm{ri}[A(\mathrm{dom}(f))] \cap \mathrm{ri}(\mathrm{dom}(g)) \neq \emptyset, \tag{3.28}$$

see [26]. An often quoted constraint qualification by Luenberger[19] for a special case of Eq. (3.26) in terms of the epigraphs of f and g reduces to (3.28) in our case.

The Fenchel duality approach is popular (see e.g. [1], [18], [21]) for solving problems like (P) when Y is finite dimensional because the dual problem in Eq. (3.27) is a finite dimensional unconstrained maximization problem, and can be solved using a variety of unconstrained optimization numerical schemes[8].

In most infinite dimensional incantations of problem (P), (FCQ) fails. For example, consider the positive minimum L^2 norm spectral estimation problem with ϕ as in Eq. (1.6). In the Fenchel duality context with $X = L^2(T)$

$$f(x) = \begin{cases} \int_T \frac{1}{2}x(t)^2\, d\mu & \text{if } 0 \leq x \text{ a.e.,} \\ +\infty & \text{else,} \end{cases} \qquad g(y) = \begin{cases} 0 & \text{if } Ax = b, \\ -\infty & \text{else.} \end{cases} \tag{3.29}$$

With this formulation (FCQ) fails since g is not continuous anywhere. Other common approaches to solving (P), using Lagrange multipliers on the interval constraints (1.3) [6], [14] and Lagrangian duality [9], require a constraint qualification similar to (FCQ) and cannot be applied without some careful analysis.

3.3 The infinite dimensional problem treated rigorously

Recently, Borwein and Lewis[3] introduced a generalization of the relative interior of a convex set, the *quasi relative interior* (qri), and suggested a constraint qualification for Fenchel duality that generalizes (FCQ) to work in our examples. They show that given a problem of the form in Eq. (3.26) where $A : X \to I\!R^n$ with constraint qualification

$$A[\mathrm{qri}(\mathrm{dom}(f))] \cap \mathrm{ridom} g \neq \emptyset, \tag{$BLCQ$}$$

then $V(P) = V(D)$ and $V(D)$ is attained if finite.

The definition of the quasi relative interior can be found in [3], but for problem (P) $(BLCQ)$ becomes

$$\exists \hat{x} \text{ such that } A\hat{x} = b, \quad I_\phi(\hat{x}) \in I\!R, \quad (\alpha \vee \iota) < \hat{x} < (\beta \wedge \sigma)\text{a.e.} \tag{CQ}$$

where $\iota \triangleq \inf \mathrm{dom}(\phi)$, $\sigma \triangleq \sup \mathrm{dom}(\phi)$ and for two extended real numbers, $a \vee b = \max\{a, b\}$, $a \wedge b = \min\{a, b\}$, see [5]. We have included explicit reference to α and β to emphasize the fact that the upper and lower bounds can be functions.

We shall make certain assumptions on ϕ that insure that ϕ is a *normal convex integrand* [25], and thus I_{ϕ^*} is well defined and equal to $(I_\phi)^*$ as one would hope. With these conditions, the dual problem in Eq. (3.27) becomes

$$V(D) \triangleq \sup\{\langle b, \lambda\rangle - I_{\phi^*}(A^T\lambda)\}. \tag{D}$$

As we have seen by the example in Eq. (2.20), there are some additional technical assumptions that need to be made on ϕ in order to guarantee solutions exist to (P) and are unique. We shall assume that $\phi : I\!R \to (-\infty, +\infty]$ is lower semicontinuous, proper, and strictly convex on its domain. Define

$$\Delta \triangleq \lim_{u\to\infty} (\phi(u)/u), \tag{3.30}$$

and if $\Delta < \infty$ define $\kappa \triangleq \lim_{v\uparrow\Delta}(\Delta - v)(\phi^*)'(v)$. We restrict attention to functions ϕ for which either $\Delta = \infty$ or $\Delta < \infty$ and $\kappa > 0$. We say such functions posses regular growth.

There are plenty of examples of convex integrands with regular growth. The L^2 norm, Boltzmann-Shannon and Burg entropies all posses regular growth. It is easy to see that if φ has regular growth and the prior function x_0 is constant, then its directed distance ϕ from Eq. (1.10) also has regular growth. The example given in Eq. (2.20) does not posses regular

Entropy	Δ	κ	Form of Solution
L^2 Norm	$+\infty$	–	$\alpha(t) \vee A^T\lambda(t) \wedge \beta(t)$
Boltzmann-Shannon	$+\infty$	–	$\alpha(t) \vee \exp[A^T\lambda(t) - 1] \wedge \beta(t)$
Burg	0	1	$\alpha(t) \vee [-A^T\lambda(t)]^{-1} \wedge \beta(t)$
Cosh	$+\infty$	–	$\alpha(t) \vee \text{arcsinh}(A^T\lambda(t)) \wedge \beta(t)$
Directed Distance	$\Delta - \varphi(x_0)$	κ	$\alpha(t) \vee (\varphi^*)'[A^T\lambda(t) + \varphi'(x_0(t))] \wedge \beta(t)$
Fermi-Dirac	$+\infty$	–	$\alpha(t) \vee \exp[A^T\lambda(t)][1 + \exp(A^T\lambda(t))]^{-1} \wedge \beta(t)$

Table 2: Entropies with regular growth and their associated solutions

growth. More such functions can be found in [2]. In Table 2 we give some common entropies and their associated Δ and κ constants[2].

The questions of existence, uniqueness and form of solution can generally be answered with the following theorem due to Borwein and Lewis[2].

Theorem 3.2 *Suppose $T = [a, b]$ is a finite interval, μ is Lebesgue measure, each ψ_k is locally Lipschitz (or in particular, continuously differentiable), and ϕ is proper, strictly convex with regular growth. Suppose (CQ) holds and*

$$\exists \tau \in I\!R^n \text{ such that } \sum_{i=1}^{n} \tau_i \psi_i(t) < \Delta \quad \forall t \in [a, b], \tag{3.31}$$

then the unique solution to (P) is given by

$$\overline{x}(t) = (\phi^*)'(\sum_{i=1}^{n} \overline{\lambda}_i \psi_i(t)) \tag{3.32}$$

where $\overline{\lambda}$ is any solution to the dual problem (D).

Eq. (3.31) always holds if one of the ψ_i is a nonzero constant function. These ideas can be extended to the multidimensional case, see [4].

We can also see that $C(T)$ is definitely not the correct space by observing that solutions need not be continuous when we use the directed distance entropy defined via Eq. (1.10) with φ the L^2 entropy from Eq. (1.6). Using Eqs. (3.25) and (3.32) we see that the solution will have the form

$$\overline{x}(t) = \max\left\{0, \sum_{i=1}^{n} \lambda_i \psi_i(t) + x_0(t)\right\} \tag{3.33}$$

[2]Here Δ and κ for the Directed Distance entropy are in terms of the original function φ and x_0 is assumed constant.

where x_0 is the prior function. If this prior function is not continuous, then there is no reason to suspect that $\overline{x}$ should be. We shall give numerical examples of this phenomenon in the next section.

It should be added that Fenchel duality can handle relaxation of the constraints to incorporate measurement error of the moments [1], [18], [21]. Let $\|\cdot\|$ be a norm on $I\!R^n$, $\|\cdot\|^*$ its dual norm and consider

$$V(P_r) \stackrel{\triangle}{=} \inf \{I_\phi(x) : \|Ax - b\| \leq \epsilon\}. \tag{P_r}$$

Provided there is a function $\hat{x} \in \mathrm{qridom}(I_\phi)$ such that $\|A\hat{x} - b\| < \epsilon$ then $V(P_r) = V(D_r)$ where

$$V(D_r) \stackrel{\triangle}{=} \max_{\lambda \in I\!R^n} \left\{ \langle b, \lambda \rangle - I_{\phi^*}(A^T\lambda) - \epsilon\|\lambda\|^* \right\} \tag{D_r}$$

and $V(D_r)$ is attained when finite[3].

Note that since $\phi = \phi^{**}$, we have $(\phi^*)' = (\phi')^{-1}$ and the solution given in Eq. (3.32) agrees with the solution in Eq. (1.14). Also, in case $\psi_i \in C(T)$ then for many entropies the solution $\overline{x}$ is also in $C(T)$ and thus we have solved the same problem over $C(T)$ or $L^\infty(T)$. For example, if $\psi_k(t) = t^{k-1}$ and we use the Boltzmann-Shannon entropy (1.7) then the solution is analytic.

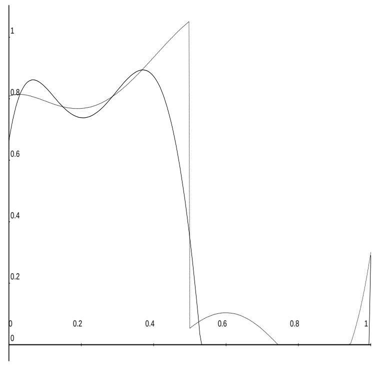
4 Numerical Results

Besides characterizing the solutions to (P), the theory of Fenchel duality has the added benefit that it replaces the infinite dimensional problem with the (usually) unconstrained finite dimensional problem (D). The dual problem is usually quite easy to solve with a standard numerical unconstrained maximization method, like Newton's method.

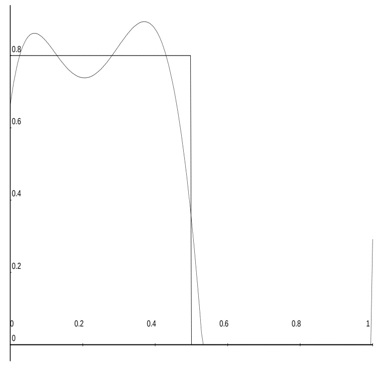
We implemented Newton's method in `Maple` to solve the one dimensional minimum entropy problem with user supplied entropy integrand ϕ, moment functions ψ_k, interval constraints α and β, and prior function $x_0(t)$, over a given finite interval $[a, b]$. We used an Armijo line search[20] with backtracking to stay in the domain of ϕ^* and exploited the explicit structure of the Hessian as much as possible. For the following examples $\psi_k(t) = t^{k-1}$, the interval is $[0, 1]$, we used 6 moments and the underlying function from which we generated the moments is $x_*(t) = 0.8\chi_{[0,1/2]}$, where $\chi_{[0,1/2]}$ is the characteristic function of $[0, 1/2]$.

The first example shows that if the prior function is discontinuous, the computed solution need not be continuous. We minimized the positive L^2 entropy with $x_0(t) = \chi_{[0,1/2]}$ and with the prior function identically zero (which corresponds to no prior function). We then graph the two solutions together in Fig. 1. Clearly, the solution of the directed distance problem is discontinuous while the other is continuous.

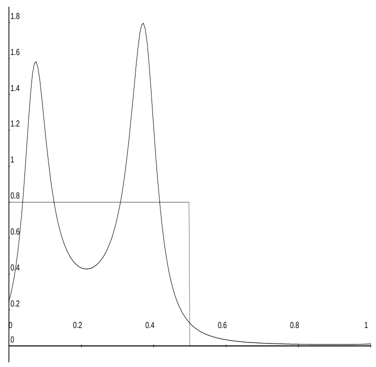
In Figs. 2, 3, and 4 we compare the positive L^2, Burg and Boltzmann-Shannon entropies with no prior function; or equivalently, since we are fixing $\int_T x(t)d\mu(t)$ in the linear constraints, with $x_0 \equiv 0$ for the L^2 and and $x_0 \equiv 1$ for the Burg and Boltzmann-Shannon



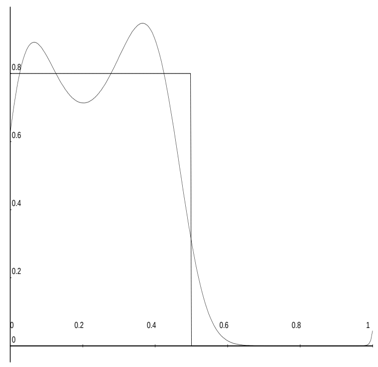
1. With and without a prior; L^2 Entropy



2. The Positive L^2 Entropy



3. The Burg Entropy



4. The Boltzmann-Shannon Entropy

entropies. It is a general observation that in examples where the support of the unknown function is not all of $[0, 1]$, the Burg entropy tends to do worse visually than the other two.

More numerical investigations are forthcoming in a future paper.

References

[1] C. Auyeung and R. M. Mersereau. A dual approach to signal restoration. *IEEE International Conference on Acoustics, Speech and Signal Processing, Glasgow*, pages 1326–1329, May 1989.

[2] J. M. Borwein and A. S. Lewis. Duality relationships for entropy–like minimization problems. *SIAM Journal of Control Theory and Optimization*, 29(2):325–338, March 1991.

[3] J. M. Borwein and A. S. Lewis. Partially-finite convex programming, parts I and II. *Mathematical Programming*, pages 15–48,49–83, 1992.

[4] J. M. Borwein and A. S. Lewis. Partially-finite programming in L_1: entropy maximization. *SIAM Journal of Optimization*, to appear.

[5] J. M. Borwein, A. S. Lewis, and M. A. Limber. Entropy minimization with lattice bounds. *Journal of Approximation Theory*, to appear.

[6] R. Cole and R. K. Goodrich. L_p spectral analysis with an L_∞ upper bound. *Journal of Optimization Theory and Applications*, to appear.

[7] A. Decarreau, D. Hilhorst, C. Lemaréchal, and J. Navaza. Dual methods in entropy maximization. Application to some problems in crystallography. *SIAM J. Optimization*, 2(2):173–197, May 1992.

[8] J. E. Dennis, Jr. and R. B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall, Inc., New Jersey, 1983.

[9] A. L. Dontchev. Duality methods for constrained best interpolation. *Mathematica Balkanica*, 1:96–105, 1987.

[10] N. Dunford and J. T. Schwartz. *Linear Operators, Part 1: General Theory*. John Wiley & Sons, 1988.

[11] T. S. Durrani and C. E. Goutis. Optimisation techniques for digital image reconstruction from their projections. *IEE Proceedings Part E*, pages 161–169, September 1990.

[12] N. J. Dusaussoy and I. E. Abdou. The extended MENT algorithm: a maximum entropy type algorithm using prior knowledge for computerized tomography. *IEEE Transactions on Signal Processing*, 39(5):1164–1180, May 1991.

[13] G. B. Folland. *Reals Analysis, Modern Techniques and their Applications*. John Wiley & Sons, New York, 1984.

[14] R. K. Goodrich, A. Steinhardt, and R. Roberts. Spectral estimation via minimum energy correlation extension. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-33(6):1509–1515, December 1985.

[15] R. B. Holmes. *Geometric Functional Analysis and its Applications*. Springer–Verlag, New York, 1975.

[16] E. T. Jaynes. Information theory and statistical mechanics. *The Physical Review*, 106, 108:620–630, 171–190, 1957.

[17] L. K. Jones and V. Trutzer. Computationally feasible high-resolution minimum-distance procedures which extend the maximum-entropy method. *Inverse Problems*, pages 749–766, 1989.

[18] R. M. Leahy and C. E. Goutis. An optimal technique for constraint-based image restoration and reconstruction. *IEEE Transactions on Acoustics, Speech and Signal Processing*, pages 1629–1642, December 1986.

[19] D. G. Luenberger. *Optimization by Vector Space Methods*. John Wiley & Sons, Inc., New York, New York, 1969.

[20] D. G. Luenberger. *Linear and Nonlinear Programming*. Addison-Wesley Publishing Company, New York, New York, 2 edition, 1984.

[21] J. H. McClellan and S. W. Lang. Duality for multidimensional MEM spectral analysis. *IEE Proceedings Part F*, pages 230–235, April 1983.

[22] C. A. Micchelli, P. W. Smith, J. Swetits, and J. D. Ward. Constrained L_p approximation. *Constructive Approximation*, pages 93–102, 1985.

[23] C. Nadeu and M. Bertran. A flatness based generalized optimization approach to spectral estimation. *Signal Processing*, pages 311–320, 1990.

[24] J. Navaza. The use of non-local constraints in maximum-entropy electron density reconstruction. *Acta Cryst.*, 42:212–223, 1986.

[25] R. T. Rockafellar. Integrals which are convex functionals. *Pacific J. of Math.*, 24(3):525–539, 1968.

[26] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[27] R. T. Rockafellar. *Conjugate Duality and Optimization*. SIAM, Philadelphia, PA., 1974.

[28] A. B. Templeman. Entropy-based optimization methods for engineering design. Preprint.