

A Note on Cores and Quasi Relative Interiors in Partially Finite Convex Programming

Scott B. Lindstrom

Hong Kong Polytechnic University, Hong Kong
scott.b.lindstrom@polyu.edu.hk

Received: May 14, 2020

Accepted: March 2, 2021

The problem of minimizing an entropy functional subject to linear constraints is a useful example of partially finite convex programming. In the 1990s, Borwein and Lewis provided broad and easy-to-verify conditions that guarantee strong duality for such problems. Their approach is to construct a function in the quasi-relative interior of the relevant infinite-dimensional set, which assures the existence of a point in the core of the relevant finite-dimensional set. We revisit this problem, and provide an alternative proof by directly appealing to the definition of the core, rather than by relying on any properties of the quasi-relative interior. Our approach admits a minor relaxation of the linear independence requirements in Borwein and Lewis' framework, which allows us to work with certain piecewise-defined moment functions precluded by their conditions. We provide such a computed example that illustrates how this relaxation may be used to tame observed Gibbs phenomenon when the underlying data is discontinuous. The relaxation illustrates the understanding we may gain by tackling partially-finite problems from both the finite-dimensional and infinite-dimensional sides. The comparison of these two approaches is informative, as both proofs are constructive.

Keywords: Quasi relative interior, core, strong duality, entropy functional optimization, partially finite convex programming.

2010 Mathematics Subject Classification: 46N10.

1. Introduction

In the 1990s, Borwein and Lewis introduced the *quasi-relative interior* (qri) of a set C [4], defined by

$$\text{qri } C = \{x \in C \text{ such that } \text{cl cone}(C - x) \text{ is a subspace}\},$$

where $\text{cone}(C - x)$ denotes the cone of $C - x$ (Definition 2.1) and $\text{cl cone}(C - x)$ is its closure. The notion of the quasi-relative interior specifies to that of the relative interior (denoted ri ; see Definition 2.1) on $\mathbb{R}^n$, while $\text{qri } C \supset \text{ri } C$ holds in a Hausdorff topological vector space of arbitrary dimension. Borwein and Lewis demonstrated the value of the quasi-relative interior in the context of the partially finite convex programming problem of constrained entropy optimization [6, 8], which we now recall.

Let f be a proper, closed convex function. We define its associated entropy functional by

$$I_f : L^1([0, \tau]) \rightarrow \mathbb{R}, \quad x \mapsto \int_0^\tau f(x(s))ds.$$

Some commonly employed entropy functions and their conjugates are listed in [8, Table 1], some of which we recall in Table 1.1.

Name	Definition	Conjugate
L^2 Norm	$\frac{1}{2}u^2$	$\frac{1}{2}v^2$
(negative) Boltzmann-Shannon	$u \log(u)$	$\exp(v - 1)$
(^{negative} _{translated}) Boltzmann-Shannon	$u \log(u) - u =: \text{ent}(u)$	$\exp(v)$
Burg	$-\log(u)$	$-1 - \log(-v)$
Cosh	$\cosh(u)$	$\text{arcsinh}(v) - \sqrt{1 + v^2}$
Fermi Dirac	$u \log(u) + (1 - u) \log(1 - u)$	$\log(1 + \exp(v))$

Table 1.1: A selection of entropies that includes those from [8].

Borwein and Lindstrom have recently illuminated the role that real principal branch of the Lambert $\mathcal{W}$ function plays in the convex conjugates for weighted sums of entropies [9]; such sums admit new entropies. The entropy optimization problem is to minimize I_f subject to the n continuous linear constraints of the form

$$\langle a_k, x \rangle = \int_0^\tau a_k(s)x(s) = b_k, \quad k = 1, \dots, n.$$

We may describe these linear equality constraints concisely in terms of the linear operator A as follows:

$$A : L^1([0, \tau]) \rightarrow \mathbb{R}^n, \quad x \mapsto \left(\int_0^\tau a_1(s)x(s), \dots, \int_0^\tau a_n(s)x(s) \right) = b \quad (1)$$

where $b := A\rho$, $\rho, a_k \in L^\infty([0, \tau])$ and ρ is a given function used to generate the linear equality constraint vector b . When f^* is smooth and everywhere finite on $\mathbb{R}$, the problem

$$\inf_{x \in L^1} \{I_f(x) \mid Ax = b\} \quad (2)$$

reduces to solving the finite nonlinear equation

$$\int_0^\tau (f^*)' \left(\sum_{j=1}^n \mu_j a_j(s) \right) a_k(s) = b_k \quad (1 \leq k \leq n). \quad (3)$$

We will recall the reasons for this in Section 2, and additional information about primal attainment may be found in [3].

The structure of the chosen entropy function f may impose hard constraints on the solution. For example, when $f = \text{ent}$ is the negative, translated Boltzmann-Shannon entropy, the feasible region consists of nonnegative functions. In a follow-up to [9], Bauschke and Lindstrom computed entropy optimization problems using *proximal* averages of entropy functions [2]. In particular, they showed how proximal averages allow the practitioner to exercise flexibility with the barriers that entropies afford, in order to obtain a larger feasible region.

1.1. Outline

The remainder of this paper is outlined as follows. In Section 2, we recall the circumstances under which (2) is equivalent to (3), excepting a detailed discussion of how to verify the strong duality. In Section 3, we provide our new approach

to developing easy-to-verify sufficient conditions for strong duality, by appealing directly to the definition of the core. In Section 4, we briefly recall the classical approach that relies on finding a function in a quasi-relative interior. In Section 5, we provide a computed example that illustrates the minor relaxation we have obtained. We conclude in Section 6.

2. On the equivalence of the dual problem and the Lagrange multiplier problem

This discussion closely follows the one found in [15], which is an augmented version of those in [2, 9]. To see why solving (2) reduces to solving (3), we will first recall several concepts and results. Throughout, X is a Banach space and X^* denotes its dual.

Definition 2.1. Let $C \subset X$. Let $F : X \rightarrow \mathbb{R} \cup \{\infty\}$ be *proper*, meaning that $-\infty \notin F(X)$ and $\text{dom} F \neq \emptyset$. We make use of the following notation:

- (i) the *cone* of C is $\text{cone } C = \{\lambda c \mid (c, \lambda) \in C \times [0, \infty[\}$;
- (ii) the *relative interior* of C is $\text{ri } C = \{x \in C \mid \text{cone}(C - x) = \text{span}(C - x)\}$;
- (iii) the *domain* of F is $\text{dom } F := \{x \in X \mid F(x) < +\infty\}$;
- (iv) the *range* of F is $\text{ran } F := \{F(x) \mid x \in X\}$;
- (v) the *subdifferential* of F is

$$\partial F : X \rightarrow 2^{X^*} \text{ by } x \mapsto \{u \in X^* \mid (\forall y \in X) \langle y - x, u \rangle + F(x) \leq F(y)\};$$

- (vi) the *domain* of the set-valued operator ∂F is $\text{dom } \partial F = \{x \in X \mid \partial F(x) \neq \emptyset\}$;
- (vii) the *range* of the set-valued operator ∂F is $\text{ran } \partial F = \partial F(X) := \cup_{u \in X} \partial F(u)$;
- (viii) and the *conjugate* of F is

$$F^* : X^* \rightarrow [-\infty, \infty] \text{ by } y \mapsto \sup_{x \in X} \{\langle y, x \rangle - F(x)\}.$$

Lemma 2.2. (Fenchel-Young equality [1, Prop. 16.10]) *Let $F : X \rightarrow]-\infty, +\infty[$ be convex and proper. Let $u \in X, x \in X^*$. Then it holds that*

$$u \in \partial F^*(x) \iff F(u) + F^*(x) = \langle x, u \rangle \implies x \in \partial f(u).$$

From the Fenchel-Young equality, it is clear that for proper, convex F ,

$$\text{ran}(\partial F^*) \subset \text{dom}(\partial F) \subset \text{dom}(F).$$

We next recall a theorem (Theorem 2.4) from [11, Theorem 4.7.1] that provides sufficient conditions for strong duality to hold. For the present exposition, we are particularly interested in how to verify those sufficient conditions. We will introduce an easy-to-check criterion in Section 3, and then we will compare it with the existing theory in Section 4. Our process uses the definition of the *core*, which we now introduce.

Definition 2.3. (Core of a set) The *core* of a set $C \subset \mathbb{R}^n$, denoted by $\text{core } C$, is the set of points $x \in C$ such that for any direction vector $\eta \in \mathbb{R}^n$, $x + t\eta \in C$ for all $t > 0$ sufficiently small.

Theorem 2.4. ([11, Theorem 4.7.1]) *Let X be a Banach space and $F : X \rightarrow \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous convex function. Let $A : X \rightarrow \mathbb{R}^n$ be a continuous linear operator, and $b \in \text{core}(A \text{ dom} F)$. Then*

$$\inf_{x \in X} \{F(x) \mid Ax = b\} = \max_{\varphi \in \mathbb{R}^N} \{\langle \varphi, b \rangle - F^*(A^T \varphi)\},$$

where A^T denotes the adjoint map that satisfies

$$A^T : \mathbb{R}^n \rightarrow X \quad \text{by} \quad \langle Au, \varphi \rangle_{\mathbb{R}^n} = \langle u, A^T \varphi \rangle_X.$$

Our new contribution in Section 3 is an easy-to-verify sufficient condition to apply Theorem 2.4. For now, we suppose that the problem from (2) satisfies the sufficient conditions to apply Theorem 2.4, and so we may reformulate (2) as

$$\inf_{x \in L^2} \{I_f(x) \mid Ax = b\} = \max_{\varphi \in \mathbb{R}^N} \{\langle \varphi, b \rangle - (I_f)^*(A^T \varphi)\}, \quad (4)$$

where $A^T : \mathbb{R}^n \rightarrow L^1$ is the adjoint map satisfying the equality

$$\langle Au, \varphi \rangle_{\mathbb{R}^n} = \langle u, A^T \varphi \rangle_{L^2}. \quad (5)$$

To further simplify this, we use the following result from [10, Theorem 6.3.4].

Proposition 2.5. *If I_f is defined as above and $f : \mathbb{R} \rightarrow]-\infty, \infty]$ is convex, proper, closed, we have*

$$(I_f)^* = I_{f^*}.$$

Proposition 2.5 allows us to express the dual problem from the right hand side of (4) explicitly as follows:

$$\max_{\varphi \in \mathbb{R}^N} \{\langle \varphi, b \rangle - I_{f^*}(A^T \varphi)\} = \max_{\varphi \in \mathbb{R}^N} \left\{ \langle \varphi, b \rangle - \int_0^\tau (f^* \circ A^T \varphi)(s) \, ds \right\}. \quad (6)$$

We may express the adjoint map A^T from (5) explicitly as follows:

$$\sum_{k=1}^n \left(\varphi_k \int_0^\tau a_k(s) u(s) \, ds \right) = \langle Au, \varphi \rangle_{\mathbb{R}^n} = \langle u, A^T \varphi \rangle_{L^2} = \int_0^\tau (A^T \varphi)(s) u(s) \, ds. \quad (7)$$

From (7), it is clear why we may explicitly describe $A^T \varphi$ by:

$$A^T \varphi = \sum_{k=1}^n \varphi_k a_k(s).$$

Thus solving (6) reduces to finding $\varphi \in \mathbb{R}^n$ that maximizes

$$\sum_{k=1}^n \varphi_k b_k - \int_0^\tau f^* \left(\sum_{k=1}^n \varphi_k a_k(s) \right) \, ds. \quad (8)$$

We can subdifferentiate (8) and recover its maximum by finding the values of $\varphi_1, \dots, \varphi_n$ for which the subdifferential of (8) with respect to φ is zero. For that purpose, we recall the Fenchel-Young *inequality* of a convex function f ,

$$(\forall x, y \in X) \quad 0 \leq f(x) + f^*(y) - \langle x, y \rangle, \quad (9)$$

which follows from the definition of the conjugate and is weaker than the Fenchel-Young *equality* in Lemma 2.2.

Because $(I_f)^* = I_{f^*}$, The Fenchel-Young equality, Lemma 2.2, guarantees that

$$(y \in \partial I_f(x)) \iff 0 = I_f(x) + I_{f^*}(y) - \langle y, x \rangle = \int f(x(s)) + f^*(y(s)) - x(s)y(s) \, ds.$$

By the Fenchel-Young *inequality* (9), the integrand on the right is nonnegative, and so must be zero almost everywhere. Using Lemma 2.2, we have that

$$f(x(s)) + f^*(y(s)) - x(s)y(s) = 0 \iff y(s) \in \partial f(x(s)).$$

Thus we subdifferentiate with respect to each φ_k in (8) and set the subdifferential equal to zero, which yields the n equations:

$$0 = b_k - \int_0^\tau (f^*)' \left(\sum_{j=1}^n \varphi_j a_j(s) \right) a_k(s) \, ds \quad \text{for } k = 1, \dots, n.$$

Thus we have reduced solving (2) to the challenge of solving (3). When $\mu_1, \dots, \mu_n$ are the optimal multipliers in (3), the solution x to the primal problem (2) is

$$x(s) = (f^*)' \left(\sum_{j=1}^n \varphi_j a_j(s) \right). \quad (10)$$

Sufficient conditions for attainment of this primal solution are given in [3]. Now we will introduce our process for verifying sufficient conditions for strong duality.

3. Approach using the core

This section is largely based on previously unpublished material from [15]. Our goal is to provide easy-to-check sufficient conditions to apply Theorem 2.4 with $F := I_f$. Recalling that I_f is convex (see, for example, [15]), we need only a convenient way to verify that $b \in \text{core}(A \text{ dom } I_f)$. The following description of the domain of I_f is from [15].

Lemma 3.1. *Let $[\alpha, \beta] \subset \text{dom } f$ with f convex. Let*

$$\mathbf{S}_f := \{x \in L^p([0, \tau]) \text{ such that } x(s) \in [\alpha, \beta] \text{ s-a.e.}\}.$$

Then $\mathbf{S}_f \subset \text{dom } I_f$.

Proof. By definition, $\text{dom } I_f = \left\{ x \in L^1([0, \tau]) \text{ such that } \int_0^\tau f(x(s)) \, ds < \infty \right\}$.

The conditions $x \in \mathbf{S}_f$, $[\alpha, \beta] \subset \text{dom } f$, and f a convex function ensure together that $f(x(s)) \leq \max\{f(\alpha), f(\beta)\} < \infty$ holds s-a.e. Thus

$$\int_0^\tau f(x(s)) \, ds \leq \int_0^\tau \max\{f(\alpha), f(\beta)\} \, ds \leq \tau \max\{f(\alpha), f(\beta)\} < \infty,$$

and so $x \in \text{dom } I_f$. □

The next example problem was used to showcase the role of the Lambert $\mathcal{W}$ function in entropy optimization in [2, 9], and it illustrates Lemma 3.1.

Example 3.2. Let f be the Boltzmann-Shannon entropy or the $\lambda \in]0, 1[$ -weighted average of the entropy and the L^2 norm: $u \mapsto \lambda u^2 + (1 - \lambda)u \log(u)$. In both cases, $[0, \infty[= \text{dom } f$. Now we define the pulse:

$$\mathbf{x} : [0, 1] \rightarrow \{0, 1\} : s \mapsto \begin{cases} 1 & \text{if } s \in [0, 1/2] \\ 0 & \text{otherwise.} \end{cases} \quad (11)$$

In this case, $\mathbf{x} \in \mathbf{S}_f$ with $\alpha = 0$ and $\beta = 1$ respectively. $\square$

We will use the definition of the core in order to develop easy-to-verify conditions for strong duality in Theorem 3.4. The following Lemma will allow us to choose functions $y_1, \dots, y_n$ that explicitly tie our functions $a_1, \dots, a_n$ with a general direction vector $\eta \in \mathbb{R}^n$ through the operator A .

Before we begin, we should note why we have chosen to work with the definition of the core instead of the interior. Since I_f is convex, its domain is convex. Since $A \text{ dom } I_f$ is then a convex, finite-dimensional set, we have $\text{int } A \text{ dom } I_f = \text{core } A \text{ dom } I_f$ ([7, Theorem 4.1.4]). However, we choose to speak in terms of the core, because our proofs will explicitly use direction vectors rather than open balls. This choice of definition allows us to find a value $t > 0$ that *depends on the direction vector* η , rather than seeking to find *one* value $t > 0$ so that the inclusion holds for *all* direction vectors as would be required by the definition of the interior.

Lemma 3.3. *Let $M \subset L^p([\zeta_1, \zeta_2])$ with $p \geq 2$ be a subspace of dimension n with basis $\{a_1, \dots, a_n\}$, and let $\eta \in \mathbb{R}^n$. Let $\langle \cdot, \cdot \rangle$ denote the inner product on $L^2([\zeta_1, \zeta_2])$. Then, for any $k \in \{1, \dots, n\}$, we may find $y_k \in M$ such that*

$$\langle y_k, a_k \rangle = \eta_k \quad \text{and} \quad \langle y_k, a_j \rangle = 0 \quad \text{for } j \neq k.$$

Proof. The case where $n = 1$ is obvious, so let $n \geq 2$. If $\eta_k = 0$, then let $y_k = 0$ and the proof is complete. Otherwise, without loss of generality, let $k = 1$. Let $v \in M \cap (\text{span}\{a_2, \dots, a_n\})^\perp \setminus \{0\}$. Then $\{v, a_2, \dots, a_n\}$ is a basis for M , and so we may write

$$a_1 = \lambda_1 v + \lambda_2 a_2 + \dots + \lambda_n a_n \quad (12)$$

for some $\lambda \in \mathbb{R}^n$. By the linear independence of $\{a_1, \dots, a_n\}$, (12) implies that

$$\lambda_1 \neq 0. \quad (13)$$

Thus we have that

$$\begin{aligned} \langle a_1, v \rangle &= \langle \lambda_1 v + \lambda_2 a_2 + \dots + \lambda_n a_n, v \rangle = \langle \lambda_1 v, v \rangle + \langle \lambda_2 a_2, v \rangle + \dots + \langle \lambda_n a_n, v \rangle \\ &= \langle \lambda_1 v, v \rangle + 0 + \dots + 0 = \lambda_1 \|v\|_{\langle \cdot, \cdot \rangle}^2. \end{aligned} \quad (14)$$

Since we have $\lambda_1 \neq 0$ and $v \neq 0$, we have $\lambda_1 \|v\|_{\langle \cdot, \cdot \rangle}^2 \neq 0$. Using the fact that $v \in M \subset L^p([\zeta_1, \zeta_2]) \subset L^2([\zeta_1, \zeta_2])$, we have that $\|v\|_{\langle \cdot, \cdot \rangle} = \|v\|_{L^2([\zeta_1, \zeta_2])} < \infty$.

Therefore, let
$$y_1 := \frac{\eta_1}{\lambda_1 \|v\|_{\langle \cdot, \cdot \rangle}^2} v. \quad (15)$$

Then we have that

$$\langle a_1, y_1 \rangle = \frac{\eta_1}{\lambda_1 \|v\|_{\langle \cdot, \cdot \rangle}^2} \langle a_1, v \rangle = \eta_1 \quad \text{and} \quad (j \neq 1) \implies \langle y_1, a_j \rangle = \frac{\eta_1}{\lambda_1 \|v\|_{\langle \cdot, \cdot \rangle}^2} \langle v, a_j \rangle = 0.$$

This is the desired result. $\square$

The following theorem provides a condition for strong duality that is, in practice, easy to check.

Theorem 3.4. *Let $[\alpha, \beta] \subset \text{dom } f$. Let $\mathbf{S}_f$ be defined as in Lemma 3.1, and let $x \in \mathbf{S}$ such that $Ax = b$. Moreover, let $0 \leq \zeta_1 < \zeta_2 \leq \tau$ and $x : [\zeta_1, \zeta_2] \rightarrow [\varepsilon_1, \varepsilon_2] \subset]\alpha, \beta[$. Let A be as in (1) with $a_1, \dots, a_n$ linearly independent in $L^\infty([\zeta_1, \zeta_2])$ and bounded on $[\zeta_1, \zeta_2]$. Then $b \in \text{core } A(\text{dom } I_f)$.*

Proof. Let $\eta \in \mathbb{R}^n$ be a direction. Recalling Definition 2.3, it suffices to find $t > 0$ such that $b + t\eta \in A(\mathbf{S}_f) \subset A(\text{dom } I_f)$.

For $k = 1, \dots, n$ let $y_k \in L^\infty([\zeta_1, \zeta_2])$ be functions bounded on $[\zeta_1, \zeta_2]$ satisfying:

$$\langle y_k, a_k \rangle_{L^2([\zeta_1, \zeta_2])} = \int_{\zeta_1}^{\zeta_2} y_k(s) a_k(s) ds = \eta_k \quad (16)$$

$$\text{and} \quad \langle y_k, a_j \rangle_{L^2([\zeta_1, \zeta_2])} = 0 \quad \text{for } j \neq k. \quad (17)$$

We may do this by the linear independence of $a_1, \dots, a_n$ on $[\zeta_1, \zeta_2]$ (simply apply Lemma 3.3 with $p = \infty$ and $M = \text{span}\{a_1, \dots, a_n\} \cap L^\infty([\zeta_1, \zeta_2])$). Since $y_1, \dots, y_n$ are bounded on $[\zeta_1, \zeta_2]$, there exists $\Delta \in]0, \infty[$ such that:

$$\max_{s \in [\zeta_1, \zeta_2], k \in \{1, \dots, n\}} \{|y_k(s)|\} < \Delta. \quad (18)$$

$$\text{Choose } t > 0 \text{ such that } t(n\Delta) < \min\{\varepsilon_1 - \alpha, (\beta - \varepsilon_2)\}. \quad (19)$$

Finally, let $y : [0, \tau] \rightarrow]-t(n\Delta), t(n\Delta)[$

$$\text{by} \quad y : s \mapsto \begin{cases} t \sum_{j=1}^n y_j(s) & \text{if } s \in [\zeta_1, \zeta_2] \\ 0 & \text{otherwise} \end{cases}.$$

We will show that:

1. $x + y \in \mathbf{S}_f$ and
2. $A(x + y) = b + t\eta$.

Conditions 1 and 2 combined demonstrate that

$$b + t\eta \in A(\mathbf{S}_f) \subset A(\text{dom } I_f), \quad \text{and thus } x \in \text{core}(A \text{ dom } I_f).$$

1: To see that $x + y \in \mathbf{S}_f$, we consider two cases.

Case 1: If $s \notin [\zeta_1, \zeta_2]$ then $(x + y)(s) = x(s) \in [\alpha, \beta]$.

Case 2: If $s \in [\zeta_1, \zeta_2]$, then we have from the assumptions that

$$x(s) \in [\varepsilon_1, \varepsilon_2]. \quad (20)$$

Additionally, from (18) and (19) we have

$$|y(s)| \leq t \sum_{k=1}^n |y_k(s)| \leq tn \left(\max_{\substack{s \in [\zeta_1, \zeta_2], \\ k \in \{1, \dots, n\}}} \{|y_k(s)|\} \right) \leq t(n\Delta) < \min\{\varepsilon_1 - \alpha, (\beta - \varepsilon_2)\}. \quad (21)$$

Now (20) and (21) combined show that $(x + y)(s) \in]\alpha, \beta[$. Combining the above two cases, we have that

$$(\forall s \in [0, \tau]) \quad (x + y)(s) \in [\alpha, \beta], \quad \text{and so } x + y \in \mathbf{S}_f.$$

2: To see that $A(x + y) = b + t\eta$, notice that by definition

$$A(x + y) = \left(\int_0^\tau a_1(s)(x(s) + y(s))ds, \dots, \int_0^\tau a_n(s)(x(s) + y(s))ds \right). \quad (22)$$

Thus the k th term of $A(x + y)$ is

$$\int_0^\tau a_k(s)(x(s) + y(s))ds = \int_0^\tau a_k(s)x(s)ds + \int_{\zeta_1}^{\zeta_2} a_k(s)y(s)ds \quad (23a)$$

$$= b_k + \int_{\zeta_1}^{\zeta_2} a_k(s)y(s)ds \quad (23b)$$

$$= b_k + \int_{\zeta_1}^{\zeta_2} a_k(s) \left(t \sum_{j=1}^n y_j(s) \right) ds \quad (23c)$$

$$= b_k + \sum_{j=1}^n t \langle a_k, y_j \rangle_{L^2([\zeta_1, \zeta_2])} \quad (23d)$$

$$= b_k + t \langle a_k, y_k \rangle_{L^2([\zeta_1, \zeta_2])} \quad (23e)$$

$$= b_k + t\eta_k. \quad (23f)$$

Here (23a) uses the fact that $s \notin [\zeta_1, \zeta_2] \implies y(s) = 0$, (23b) uses the definition of $Ax = b$ to replace the left integral by b_k , (23c) uses only the definition of y , (23d) is the definition of the inner product on $L^2([\zeta_1, \zeta_2])$, (23e) uses (17), and (23f) uses (16). Finally, (22) and (23) combined yield

$$A(x + y) = (b_1 + t\eta_1, \dots, b_n + t\eta_n) = b + t\eta.$$

This completes the result. $\square$

The following example illustrate the application of Theorem 3.4.

Example 3.5. (The pulse) Consider the case of Example 3.2 where $\mathbf{x}$ is the pulse. In terms of Theorem 3.4, $\mathbf{x} \in \mathbf{S}_f$ with $\alpha = 0$, $\beta = \infty$, $0 \leq \zeta_1 < \zeta_2 \leq \frac{1}{2}$, and $0 < \varepsilon_1 = 1 = \varepsilon_2 < \infty$. $\square$

4. Approach using the quasi-relative interior

Now we will compare the approach we used in Section 3 to obtain Theorem 3.4 against the process used in the existing theory. We have the following proposition from Borwein and Lewis' first paper on quasi-relative interiors.

Proposition 4.1. ([4, Proposition 2.10]) *Let X, Y be locally convex with $C \subset X$ convex and $A: X \rightarrow \mathbb{R}^n$ continuous and linear. If $\text{qri } C \neq \emptyset$, then $A(\text{qri } C) = \text{ri}(AC)$.*

As before, we want to verify the condition that $b \in \text{core}(A \text{ dom } F)$, so that we can apply Theorem 2.4 to guarantee strong duality. By Proposition 4.1, it suffices to find a feasible $x \in \text{qri}(\text{dom } I_f)$, because if such a feasible x exists, Proposition 4.1 tells us that $b = Ax \in A \text{ qri}(\text{dom } I_f) \subset \text{ri}(A \text{ dom } I_f) \subset \text{core}(A \text{ dom } I_f)$.

We denote by X_+ the positive cone in X . In the case where $X = L^p(\mathbb{R}, \mu)$, where $(\mathbb{R}, \mu)$ is the real numbers together with the usual Lebesgue measure, we simply have

$$X_+ = X_{\geq 0} := \{x(s) \mid x(s) \geq 0, s - a.e.\} \quad (24)$$

It may be seen from Lemma 3.1 and Table 1.1 that X_+ is prototypical of the domain for many entropy functionals of interest. The following example from [4] provides an important characterization of the quasi-relative-interiors of the domains of many entropy functionals.

Example 4.2. [4, Examples 3.11] Let $X = L^p(T, \mu)$ with (T, μ) a σ -finite measure space and $1 \leq p < \infty$. Then we have that

$$(Q1) \quad \text{qri}(X_+) = \{x \mid x(s) > 0, \text{ } s - a.e.\}.$$

Strong duality may be verified by finding a feasible point x (feasible in the sense that $Ax = b$) that is also in $\text{qri}(X_+)$. However, consider the example when f is the Boltzmann-Shannon entropy and $b = A\mathbf{x}$ where $\mathbf{x}$ is the pulse (11). Here we possess a priori knowledge of a point $\mathbf{x}$ that is *feasible* in the sense that $A\mathbf{x} = b$. However, (Q1) tells us that the feasible point $\mathbf{x}$ is not in $\text{qri}(X_+)$. For many problems when strong duality holds, it may be difficult to find a feasible $x \in \text{qri}(X_+)$ by using the characterization in (Q1).

Borwein and Lewis provided a result, [3, Theorem 2.9], which we recall as Theorem 4.4, that allows us to easily check strong duality for this particular example. Before we introduce the theorem, we first need a definition from [4].

Definition 4.3. (Pseudo-Haar [5, Definition 7.8]) Suppose that $a_i : [\alpha, \beta] \rightarrow \mathbb{R}$ for $i = 1, \dots, m$ are continuous and linearly independent on every non-null subset of $[\alpha, \beta]$. Then we say the a_i s are pseudo-Haar on $[\alpha, \beta]$.

For example, the monomials $a_i : s \mapsto s^k$, $i = 1, \dots, n$ are pseudo-Haar on $[\alpha, \beta]$ for any $\alpha < \beta$. This choice of a_i was used in, for example, [2, 9]. Now we recall the strong duality conditions of Borwein and Lewis.

Theorem 4.4. [3, Theorem 2.9] Suppose (T, μ) is a finite measure space and that

$$(Q2a) \quad 0 \leq \rho \in L^p(T) \text{ is nonzero } (1 \leq p \leq \infty).$$

$$(Q2b) \quad \text{and } a_i \in L^q(T), i = 1, \dots, n \text{ are pseudo-Haar}$$

Then there exists a $y \in L^\infty(T)$ and $\varepsilon > 0$ with $y(s) \geq \varepsilon$ almost everywhere and $\langle \rho, a_i \rangle = \langle y, a_i \rangle$, each i .

Proof. See [3, Theorem 2.9] or our proof below for Theorem 4.7, which is based thereon. $\square$

Clearly the choice of the pulse $\mathbf{x} \in L^1([0, 1])$ satisfies (Q2a); moreover, the condition (Q2b) holds for, say, monomials $a_i : s \mapsto s^i$ where $a_1, \dots, a_n \in L^\infty([0, 1])$. Now the condition $\langle \mathbf{x}, a_i \rangle = \langle y, a_i \rangle$ is the same as $A\mathbf{x} = Ay = b$, and the condition $y(s) \geq \varepsilon$, $s - a.e.$ assures that $y \in \text{ri}(X_+)$ by (Q1). Thus, the existence of this y guarantees strong duality by Proposition 4.1. Thus the conditions (Q2) are sufficient to guarantee strong duality for the choice $\rho = \mathbf{x}$, and they are easy to check.

However, (Q2) does not tell us what to do when X_+ properly contains the domain of our functional f . For example, when f is the Fermi Dirac entropy, we have $\text{dom } f = [0, 1]$. For such a case, we can also define the sets $X_{\geq \alpha}, X_{\leq \beta}$ by

$$X_{\geq \alpha} := \{x \mid x(s) \geq \alpha, \text{ } s - a.e.\}, \text{ and } X_{\leq \beta} := \{x \mid x(s) \leq \beta, \text{ } s - a.e.\}.$$

It is clear that $X_{\leq\beta} = -X_+ + \beta$ where β denotes the constant function $\beta : s \mapsto \beta$. Naturally, we have that

$$\text{qri } X_{\leq\beta} = -\text{qri } X_+ + \beta = \{x \mid x(s) < \beta, \text{ } s - a.e.\} \quad (25a)$$

$$\text{and analogously } \text{qri } X_{\geq\alpha} = -\text{qri } X_+ + \alpha = \{x \mid x(s) > \alpha, \text{ } s - a.e.\}. \quad (25b)$$

Clearly we then have that

$$\text{qri } X_{\geq\beta} \cap \text{qri } X_{\leq\alpha} = \{x \mid \alpha < x(s) < \beta, \text{ } s - a.e.\}. \quad (26)$$

This leads us to consider Theorem 4.6, for which we employ another result of Borwein and Lewis.

Theorem 4.5. ([4, Theorem 2.13]) *Let X be locally convex and $C, D \subset X$ be convex. If $C \cap \text{int} D \neq \emptyset$, then $(\text{qri } C) \cap (\text{int} D) = \text{qri}(C \cap D)$.*

Theorem 4.6. *Let*

$$Ay = b \quad \text{and} \quad \alpha < y(s) < \beta, \text{ } s - a.e., \quad \text{and} \quad [\alpha, \beta] \subset \text{dom } f. \quad (27)$$

Then $b \in \text{core}(A \text{ dom } I_f)$.

Proof. We have that

$$b = Ay \in A(\text{qri } X_{\geq\alpha} \cap \text{qri } X_{\leq\beta}), \quad (28a)$$

$$= A \text{ qri } X_{\geq\alpha} \cap A \text{ qri } X_{\leq\beta}, \quad (28b)$$

$$= \text{ri } AX_{\geq\alpha} \cap \text{ri } AX_{\leq\beta}, \quad (28c)$$

$$= \text{qri}(AX_{\geq\alpha}) \cap \text{int}(AX_{\leq\beta}), \quad (28d)$$

$$= \text{qri}(AX_{\geq\alpha} \cap AX_{\leq\beta}), \quad (28e)$$

$$= \text{qri}(A(X_{\geq\alpha} \cap X_{\leq\beta})), \quad (28f)$$

$$\subset \text{qri } A \text{ dom } I_f \quad (28g)$$

$$= \text{core } A \text{ dom } I_f. \quad (28h)$$

Here the inclusion from (28a) is true because of (26) and the condition $\alpha < y(s) < \beta$, (28b) uses the linearity of A , and (28c) is true by Proposition 4.1. We have that (28d) is true, because $AX_{\geq\alpha}$ and $AX_{\leq\beta}$ are nonempty cones with dimension n in $\mathbb{R}^n$, and so we have that their interiors and relative interiors coincide. We have that (28e) is true by Theorem 4.5, and (28f) again uses the linearity of A . Next, (28g) holds because $\text{aff } A(X_{\leq\alpha} \cap X_{\leq\beta}) = \text{aff } A \text{ dom } I_f = \mathbb{R}^n$ and $X_{\leq\alpha} \cap X_{\leq\beta} \subset \text{dom } I_f$ by Lemma (3.1). Finally, (28h) is true by convexity. Altogether, (28) shows the desired result. $\square$

The key, then, is to find a y that satisfies (27). The existence of this y does not immediately follow from Theorem 4.4, but Theorem 4.4 may be extended by slightly modifying Borwein and Lewis' proof (from [3, Theorem 2.9]) as follows in Theorem 4.7.

Theorem 4.7. (A simple extension of [3, Theorem 2.9]) Suppose (T, μ) is a finite measure space and $x \in L^P(T)$ satisfies $\alpha \leq x \leq \beta$ where

$$\mu(\{t \mid \alpha < x(s) < \beta\}) > 0. \quad (29)$$

Suppose further that $a_i \in L_q(T)$, $i = 1, \dots, n$, are pseudo-Haar. Then there exists a $y \in L^\infty(T)$ and $\varepsilon > 0$ with $y(t) \geq \varepsilon$ almost everywhere and $\langle x, a_i \rangle = \langle y, a_i \rangle$ for each i .

Proof. Without loss of generality, let $0 = \alpha, 1 = \beta$. Since $\mu(\{t \mid 0 < x(s) < 1\}) > 0$, there exists $T_1 \subset T$ with $\mu(T_1) > 0$ and a $\delta > 0$ such that $x(t) \geq \delta$ almost everywhere on T_1 . We claim

$$C1 := \left\{ \left(\int_{T_1} u a_i d\mu \right)_{i=1}^n \mid u \in L^\infty(T_1) \right\} = \mathbb{R}^n. \quad (30)$$

Suppose to the contrary. As $C1$ is a subspace and is linearly dependent on $\mathbb{R}^n$, there exists a linear function $\lambda \in \mathbb{R}^n$ with $C1$ contained in its nullspace. This means that

$$\sum_{i=1}^n \lambda_i \int_{T_1} u a_i d\mu = 0 \quad \text{for all } u \in L^\infty(T_1).$$

Thus we have that $\sum_{i=1}^n \lambda_i a_i(t) = 0$ almost everywhere on T_1 . This is a contradiction, because the a_i are pseudo-Haar, and so they are linearly independent on T_1 . Thus (30) holds.

Now define $C2 = \left\{ \left(\int_{T_1} u a_i d\mu \right)_{i=1}^n \mid u \in L^\infty(T_1), \|u\|_\infty < \delta/2 \right\}$.

Since, by (30), $\text{cone } C2 = \mathbb{R}^n$, we have that $0 \in \text{int } C2$ by [16, Corollary 6.4.1]. Define the sequence $(x_m)_{m=3}^\infty \subset L^\infty(T)$ by

$$x_m(t) = \begin{cases} 1 - \frac{1}{m} & \text{if } x(t) > 1 - \frac{1}{m}; \\ x(t) & \text{if } \frac{1}{m} \leq x(t) \leq 1 - \frac{1}{m}; \\ \frac{1}{m} & \text{if } x(t) < \frac{1}{m}. \end{cases}$$

For $p < \infty$ we have

$$\begin{aligned} \|x_m - x\|_p^p &\leq \int_{\{t \mid x(t) > 1 - \frac{1}{m}\}} \left(1 - \left(1 - \frac{1}{m} \right) \right)^p d\mu + \int_{\{t \mid x(t) < \frac{1}{m}\}} \left(\frac{1}{m} \right)^p d\mu \\ &= \int_{\{t \mid \leq x(t) \notin [1/m, 1-1/m]\}} \left(\frac{1}{m} \right)^p d\mu \rightarrow 0 \text{ as } m \rightarrow \infty. \end{aligned}$$

Otherwise, in the $p = \infty$ case, we have $\|x_m - x\|_\infty \leq 1/m$ for all m . Thus we have that $(\langle x_m - x, a_i \rangle)_{i=1}^m \rightarrow 0$ as $m \rightarrow \infty$, and so, since $0 \in \text{int } C2$, we have that for $m > 1/\delta$ sufficiently large, $(\langle x_m - x, a_i \rangle)_{i=1}^n \in C2$. Thus we may find $v \in L^\infty(T)$ such that $\langle x_m - x, a_i \rangle = \langle v, a_i \rangle$ for $i = 1, \dots, n$ and $\|v\|_\infty < \delta/2$ with $v(t) = 0$ almost everywhere on T_1^c . Set $y = x_m - v$. Then we have $y \in L^\infty(T)$ and $y(t) = x_m(t) \in [1/m, 1 - 1/m]$ almost everywhere on T_1^c .

On T_1 we have $1 - \frac{1}{m} \geq 1 - \delta \geq x(t) \geq \delta > 1/m$ almost everywhere, so we have $x_m(t) \in [\delta, 1 - \delta]$ almost everywhere. Since $v(t) \leq \delta/2$ almost everywhere, we have that $y(t) \in [\delta/2, 1 - \delta/2]$ almost everywhere on T_1 .

Altogether we have that $y \in [\min\{\delta/2, 1/m\}, 1 - \min\{\delta/2, 1/m\}]$ almost everywhere on T . Finally, since $y = x_m - v$ and $\langle x_m - x, a_i \rangle = \langle v, a_i \rangle$ for $i = 1, \dots, n$ we have that

$$\langle y, a_i \rangle = \langle x_m - v, a_i \rangle = \langle x_m, a_i \rangle + \langle v, a_i \rangle = \langle x_m, a_i \rangle - \langle x_m - x, a_i \rangle = \langle x, a_i \rangle$$

for $i = 1, \dots, n$. This concludes the result. $\square$

While Theorem 4.7 is a straightforward extension of [3, Theorem 2.9], it is not actually proven therein. Borwein and Lewis simply point out [3, Example 5.6(vi)] that, in such cases, the problem reduces to finding a y that satisfies (27). In their later work [6], Borwein and Lewis list the extension from Theorem 4.7 as the primal constraint qualification (PCQ_3), and they attribute the first proof to [13]; see also Lewis' detailed exposition on the consistency of moment systems [14].

To conclude our discussion of how the quasi-relative interior approach here relates to our approach to developing the constraint qualification by using the definition of the core, we have the following proposition that shows how (29) implies the existence of our interval $[\zeta_1, \zeta_2]$ from Theorem 3.4.

Proposition 4.8. *Let $\mu(\{x \mid \alpha < f(x) < \beta\}) > 0$. Then there exists $\epsilon, \delta > 0$ such that*

$$\mu(\{x \mid \alpha + \epsilon \leq f(x) \leq \beta - \epsilon\}) = \delta > 0. \quad (31)$$

Proof. Suppose for a contradiction that such a (ϵ, δ) pair does not exist. Let

$$\epsilon_n := \frac{\beta - \alpha}{2^n} \quad \text{for } n \in \mathbb{N};$$

$$A_n := \{x \in T \mid \alpha + \epsilon_n \leq f(x) \leq \beta - \epsilon_n\} \quad \text{and} \quad B_n := A_{n+1} \setminus \cup_{j=1}^n A_j, \quad n \in \mathbb{N}.$$

Since we have supposed that (31) does not hold for any $(\epsilon, \delta) \in \mathbb{R}_{++}^2$, we have that $\mu(A_n) = 0$ for all n , and so $\mu(B_n) = 0$ for all n . Since $(B_n)_{n \in \mathbb{N}}$ is a disjoint, countable family of members of a σ -algebra,

$$\mu\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mu(B_n) = 0. \quad (32)$$

See, for example, [12]. However, we also have that

$$\bigcup_{n=1}^{\infty} B_n = \{x \mid \alpha \leq f(x) \leq \beta\}. \quad (33)$$

Together, (32) and (33) imply $\mu(\{x \mid \alpha < f(x) < \beta\}) = 0$. A contradiction. $\square$

From Proposition 4.8, it is clear why the constraint qualification conditions of Theorem 4.7 are more restrictive than those of Theorem 3.4 for the problem (2). We only require that $a_1, \dots, a_n$ be linearly independent on one set: $[\zeta_1, \zeta_2]$.

We gain different insights through the two different approaches to the problem. By working in the core, we obtained a sufficiency condition that is no more restrictive in practice than the state-of-the-art in the literature. Moreover, we obtain it with very little scaffolding.

5. A computed example

Consider the pulse from Examples 3.2 and 3.5. Theorem 3.4 admits the use of new moment functions for solving (2), provided that the moment functions are linearly independent on some nonempty $[\zeta_1, \zeta_2] \subset [0, 1/2]$. In particular, let

$$a_i(t) = \begin{cases} t^{i-1} & \text{if } t \leq 1/2; \\ 1 & \text{if } t \geq 1/2, \end{cases} \quad \text{for } j = 1, \dots, n. \quad (34)$$

We compute with f as the negative translated Boltzmann-Shannon entropy, and $\tau = 1$. In this case, $[\zeta_1, \zeta_2]$ is any subset of $[0, 1/2]$. In Figure 5.1, we show the solution obtained by computing with piecewise-defined a_i in (34). For comparison, we also show the solution obtained when the $a_i : t \rightarrow t^{i-1}$ are monomials on the entire interval $[0, 1]$. The latter are the moments used in [2, 9, 15], to which the reader is referred for more details about computation. At left, we see that the use of the piecewise a_i apparently tames the observed Gibbs phenomenon near $1/2$. At right, we rescale for clearer comparison.

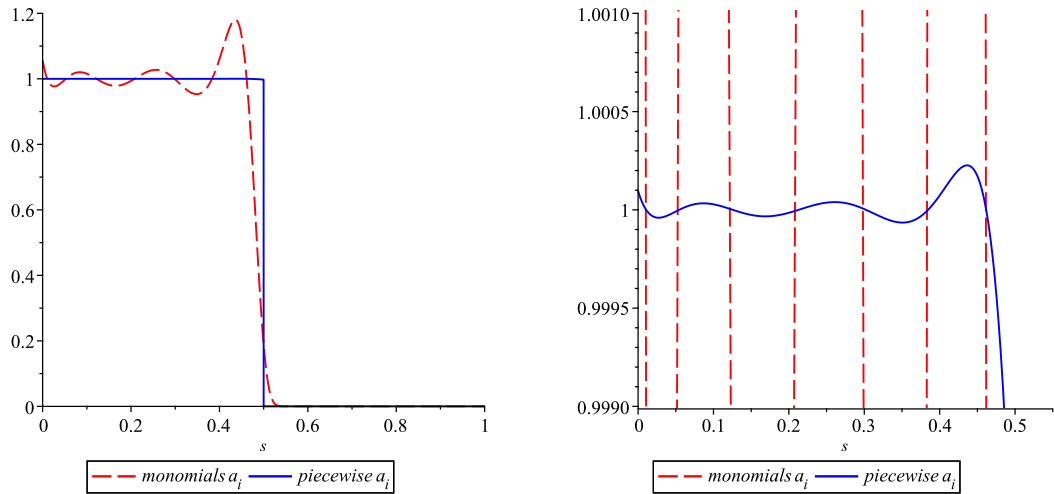


Figure 5.1: The computed example from Section 5.

6. Conclusion

Let us compare and contrast the two approaches to finding easy-to-check conditions for strong duality. In Section 3, we started with a direction vector $\eta \in \mathbb{R}^n$ and an x that satisfied

$$x : [\zeta_1, \zeta_2] \rightarrow [\varepsilon_1, \varepsilon_2] \subset]\alpha, \beta[\subset \text{dom } I_f, \quad (35)$$

and then we showed how we may construct a function $x + y$ and $t > 0$ that satisfies $x + y \in \text{dom } f$ and $A(x + y) = b + t\eta$. This shows that $b \in \text{core } A \text{ dom } I_f$.

In Section 4, we started with a function that satisfies (31) and so also satisfies (35) by Proposition 4.8. We then show how to construct another function y that satisfies $Ay = b$ with $y \in \text{qri dom } I_f$, which guarantees $b \in \text{core } A \text{ dom } I_f$ by Theorem 4.6.

The proof in Section 3 has two main advantages. The first is that it provides an interiority condition that is easy to understand from a geometric standpoint. We simply look for the interval $[\zeta_1, \zeta_2]$ and then can safely add a function y to the function x on this interval while keeping $x + y \in \text{dom } I_f$. The second advantage is

that this geometric construction makes clear why we only need the a_i to be linearly independent on $[\zeta_1, \zeta_2]$.

The proof in Section 4 has a significant theoretical advantage of its own, which is that it shows how one can construct a function y that lives in $\text{qri dom } I_f$. The two approaches are related by Proposition 4.8, which shows how we may find the set T_1 by simply taking $T_1 := [\zeta_1, \zeta_2]$, the existence of the latter being guaranteed. Revisiting the proof of Theorem 4.7, we realize that we have only used the pseudo-Haar assumption to guarantee linear independence of the a_i on the set $T_1 = [\zeta_1, \zeta_2]$, and so we only ever needed the linear independence on T_1 to begin with.

The study of which properties of relative interiors in finite dimensions will extend to quasi-relative interiors in infinite dimensions is an active area of current research. This study of the entropy minimization problem illustrates the insights that are to be gained by examining problems that are partially finite, and approaching them from both their finite-dimensional and infinite-dimensional sides.

References

- [1] H. H. Bauschke, P. L. Combettes: *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2nd ed., CMS Books in Mathematics, Springer, Cham (2017).
- [2] H. H. Bauschke, S. B. Lindstrom: *Proximal averages for minimization of entropy functionals*, Pure Appl. Funct. Analysis 5/3 (2020) 505–531.
- [3] J. M. Borwein, A. S. Lewis: *Duality relationships for entropy-like minimization problems*, SIAM J. Control Optimization 29 (1991) 325–338.
- [4] J. M. Borwein, A. S. Lewis: *Partially finite convex programming. I: Quasi relative interiors and duality theory*, Math. Programming 57 (1992) 15–48.
- [5] J. M. Borwein, A. S. Lewis: *Partially finite convex programming. II: Explicit lattice models*, Math. Programming 57 (1992) 49–83.
- [6] J. M. Borwein, A. S. Lewis: *Partially-finite programming in L_1 and the existence of maximum entropy estimates*, SIAM J. Optimization 3/2 (1993) 248–267.
- [7] J. M. Borwein, A. S. Lewis: *Convex Analysis and Nonlinear Optimization: Theory and Examples*, 2nd ed., Springer, Berlin (2006).
- [8] J. M. Borwein, M. A. Limber: *On Entropy Maximization via Convex Programming*, Faculty of Mathematics, University of Waterloo (1992).
- [9] J. M. Borwein, S. B. Lindstrom: *Meetings with Lambert W and other special functions in optimization and analysis*, Pure Appl. Func. Analysis 1/3 (2016) 361–396.
- [10] J. M. Borwein, J. D. Vanderwerff: *Convex Functions: Construction, Characterizations and Counterexamples*, Cambridge University Press, Cambridge (2010).
- [11] J. M. Borwein, Q. Zhu: *Techniques of Variational Analysis*, CMS Books in Mathematics, Springer, New York (2005).
- [12] H. Brezis: *Functional Analysis, Sobolev Spaces and Partial Differential Equations*, Springer, Cham (2010).
- [13] A. S. Lewis: *Pseudo-Haar functions and partially-finite programming*, manuscript.
- [14] A. S. Lewis: *Consistency of moment systems*, Can. J. Math. 47/5 (1995) 995–1006.
- [15] S. B. Lindstrom: *Proximal point algorithms, dynamical systems, and associated operators: Modern perspectives from experimental mathematics*, available at <https://nova.newcastle.edu.au/vital/access/manager/Repository/uon:34387> (2019).
- [16] R. T. Rockafellar: *Convex Analysis*, Princeton University Press, Princeton (1970).