Mathematical Programming 65 (1994) 123–138

Facial reduction in partially finite convex programming

A.S. Lewis[1]

Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

Received 3 April 1992; revised manuscript received 11 June 1993

Abstract

We consider the problem of minimizing an extended-valued convex function on a locally convex space subject to a finite number of linear (in)equalities. When the standard constraint qualification fails a reduction technique is needed to derive necessary optimality conditions. Facial reduction is usually applied in the range of the constraints. In this paper it is applied in the domain space, thus maintaining any structure (and in particular lattice properties) of the underlying domain. Applications include constrained approximation and best entropy estimation.

AMS 1991 Subject Classification: Primary 49K27, 90C48; secondary 90C34, 52A41

Keywords: Partially finite convex program; Necessary conditions; Facial reduction; Quasi relative interior; Constrained approximation; Maximum entropy; Vector lattice

1. Introduction

Suppose X is a locally convex topological vector space, $f: X \to (-\infty, +\infty]$ is convex and $A: X \to \mathbb{R}^n$ is continuous and linear. Consider the following problem:

$$\text{(PFP)} \qquad \begin{cases} \inf & f(x) \\ \text{subject to} & Ax = b\,. \end{cases}$$

This 'partially-finite convex program' can be used to model a wide variety of interesting optimization problems, including constrained approximation, interpolation and smoothing

[1]E-mail: aslewis@orion.uwaterloo.ca. Research partially supported by the Natural Sciences and Engineering Research Council of Canada.

SSDI 0025-5610(93)E0078-S

problems, semi-infinite linear programming, best entropy estimation and semi-infinite transportation problems. Surveys of duality and existence theory may be found in [2–5].

The domain of f is denoted $\operatorname{dom} f = \{x \mid f(x) < +\infty\}$. When $\operatorname{dom} f$ is a cone, the image cone $A \operatorname{dom} f$ is often called the 'moment cone'. The constraint qualification that b lies in the relative interior of $A \operatorname{dom} f$ ensures that there is a Lagrange multiplier vector λ in $\mathbb{R}^n$ for (PFP), or equivalently that the dual problem has an optimal solution with value equal to the primal value. Under reasonable conditions we can then reconstruct the primal optimal solution as $x_0 = \nabla f^*(A^*\lambda)$, the (Gâteaux) derivative of the Fenchel conjugate evaluated at the image of λ under the adjoint map A^*.

When the constraint qualification fails it is still possible to derive more restricted optimality conditions (and hence characterizations of optimal solutions): [1] for example is an interesting survey. One approach is to restrict the variable x to lie in some subset E of X, chosen to include the feasible region. One interpretation of the 'facial reduction' technique used in [7] is that E is chosen to make $A(\operatorname{dom} f \cap E)$ the smallest face of $A \operatorname{dom} f$ which contains b: the constraint qualification is then satisfied. A similar technique is applied in [11].

A disadvantage with this technique, and with the constraint qualification, is the necessity of working in the *range* space of the constraints, using the set $A \operatorname{dom} f$. This is particularly the case when $\operatorname{dom} f$ has a simple structure such as a lattice cone, as occurs frequently in constrained approximation, or an order interval, which is of interest for example in best entropy estimation (see for example [12]). Rewriting the constraint qualification to exploit this structure is described in [3] and [4]. The idea of this paper is to use similar techniques to apply facial reduction in the *domain* space of the constraints, directly on the set $\operatorname{dom} f$.

A particularly striking example in support of this approach is the following constrained approximation problem:

$$
\text{(CAP)} \qquad \begin{cases} \inf & (1/p)\|x\|_p^p \\ \text{subject to} & \displaystyle\int_S a_i x = b_i, \quad (i = 1, \ldots, n) \\ & 0 \leqslant x \in L_p(S), \end{cases}
$$

where $L_p(S)$ is separable and $a_i \in L_q(S)$ $(i = 1, \ldots, n)$. An elegant result in [15] shows, assuming only that (CAP) is consistent, that there is a largest measurable subset S_0 on which every feasible x vanishes almost everywhere, and that the optimal solution of (CAP) has the form

$$
x_0(s) = \begin{cases} \left(\displaystyle\sum_{i=1}^{n} \lambda_i a_i(s) \right)_+^{q-1}, & (s \notin S_0) \\ 0, & (s \in S_0) \end{cases}
$$

where the subscript $+$ denotes the positive part (see also [8] and [10, 11]). We shall see how this result appears naturally as a consequence of the facial structure of the positive cone in $L_p(S)$, using the well-known structure of the closed ideals. These ideas extend to estimation problems having entropy-type integral functionals as objectives (c.f. [2]), and to problems involving order intervals.

2. Faces

We shall begin by surveying some of the properties of faces that we require in our analysis of convex programs (see Section 1.5 in [9] for comparison). Suppose initially that X is an arbitrary (real) vector space and $C \subset X$ is convex.

Definition 2.1 [19]. A convex set $E \subset C$ is a face of C if x, $y \in C$, $0 < \lambda < 1$ and $\lambda x + (1-\lambda)y \in E$ implies $x, y \in E$.

It is immediate that the intersection of any collection of faces is a face. For an arbitrary subset $D \subset C$ we can therefore consider the intersection of all faces of C containing D, and this will be the unique *smallest face* of C containing D, denoted $F(C, D)$. It is also clear that any face of a convex cone must be a convex cone.

Proposition 2.2. *Suppose E is a face of C, and H is a hyperplane (possibly dense) which supports E. Then $H \cap E$ is a face.*

Proof. For some linear functional ψ, and α in $\mathbb{R}$, $H = \{x \in X \mid \psi(x) = \alpha\}$ and $\psi(x) \leqslant \alpha$ for all x in E (see [13]). Suppose then that $x, y \in C$, $0 < \lambda < 1$ and $\lambda x + (1-\lambda)y \in H \cap E$. Since E is a face, $x, y \in E$. Thus $\psi(x) \leqslant \alpha$, $\psi(y) \leqslant \alpha$, and $\psi(\lambda x + (1-\lambda)y) = \alpha$, whence we obtain $x, y \in H$, so the result follows. □

Lattice ideas will be useful in what follows. We refer the reader to [19] for definitions and notation. If X is a vector lattice with $x \leqslant y$ in X then we denote the *order interval* $\{z \in X \mid x \leqslant z \leqslant y\}$ by $[x, y]$. A subset C of X is *saturated* if $x \leqslant y$ in C implies $[x, y] \subset C$, and is *solid* if $x \in C$ and $|y| \leqslant |x|$ implies $y \in C$. An *ideal* is a solid subspace.

A particularly significant type of convex set for what follows is the positive cone X_+ in a vector lattice X. The faces of such a cone are easily identified.

Proposition 2.3. *If X is a vector lattice then the faces of X_+ are precisely the sets $I \cap X_+$, where I is an ideal. Moreover, if E is a face of X_+ then the set $I = E - E$ is an ideal, and we have $E = I \cap X_+$.*

Proof. Suppose $x, y \in X_+$ $\lambda \in (0, 1)$ and $\lambda x + (1-\lambda)y \in I \cap X_+$, where I is an ideal. Then $0 \leqslant x \leqslant (1/\lambda)(\lambda x + (1-\lambda)y) \in I$, so $x \in I \cap X_+$. Similarly $y \in I \cap X_+$, so $I \cap X_+$ is a face.

Conversely, suppose E is a face of X_+, and define $I := E - E$. Suppose $x \in I$ and $|y| \leqslant |x|$. We wish to show $y \in I$. Since $x \in I$, $x = x_1 - x_2$ with $x_1, x_2 \in E \subset X_+$, so

$$0 \leqslant y^+, y^- \leqslant |y| \leqslant |x| = x^+ + x^- \leqslant x_1 + x_2 \in E$$

(because E must be a cone). But now

$$x_1 + x_2 = \tfrac{1}{2} y^+ + \tfrac{1}{2}(2x_1 + 2x_2 - y^+) ,$$

and since y^+, $2x_1 + 2x_2 - y^+ \in X_+$ and E is a face, $y^+ \in E$. Similarly $y^- \in E$, so $|y| = y^+ + y^- \in E \subset I$ as required. Thus I is an ideal.

Finally we need to show that $E = I \cap X_+$. Clearly $E \subset I \cap X_+$, so suppose $x \in I \cap X_+$, so there exist x_1, $x_2 \in E$ with $x_1 - x_2 = x \geqslant 0$. But now $\frac{1}{2}x + \frac{1}{2}x_2 = \frac{1}{2}x_1 \in E$, so since E is a face, $x \in E$. Thus $I \cap X_+ \subset E$ as required. $\square$

Thus if X is a vector lattice the smallest face of the positive cone containing an arbitrary set D is just the intersection of the positive cone with the ideal generated by D:

$$F(X_+, D) = I(D) \cap X_+ = \bigcup_{n=1}^{\infty} \{n[0, x] \mid x \in D\}$$

(see [19]).

Let us now suppose that X is a locally convex topological vector space, with topological dual space X^*. The *normal cone* to C at x_0 is

$$N_C(x_0) = \{x^* \in X^* \mid x^*(x - x_0) \leqslant 0 \text{ for all } x \in C\} .$$

The following idea is from [3].

Definition 2.4. The element x_0 of C lies in the *quasi relative interior* of C (denoted by qri C) if $N_C(x_0)$ is a subspace.

The set qri C shares many of the important properties of the relative interior in finite dimensions: in particular, qri C = ri C for finite-dimensional sets. If Y is a separable normed space then nonempty closed convex subsets of Y have nonempty quasi relative interior, and nonempty weak*-closed convex subsets of Y^* have nonempty weak* quasi relative interior. Geometrically, $x_0 \in$ qri C if there is no proper closed supporting hyperplane to C at x_0.

Theorem 2.5. *For convex sets $D \subset C \subset X$, we have* qri $D \subset$ qri $F(C, D)$.

Proof. Suppose $x_0 \in \text{qri } D \backslash \text{qri } F(C, D)$. Then there exists

$$x^* \in N_{F(C,D)}(x_0) \backslash (-N_{F(C,D)}(x_0)) ,$$

so for some x_1 in $F(C, D)$, $x^*(x_1 - x_0) < 0$. Since $D \subset F(C, D)$, $x^* \in N_D(x_0)$, and thus $x^* \in -N_D(x_0)$, or in other words $x^*(x - x_0) = 0$ for all x in D.

Let us consider the hyperplane $H := \{x \mid x^*(x - x_0) = 0\}$ which supports $F(C, D)$. Then by Proposition 2.2, $H \cap F(C, D)$ is a face, and the above remarks show it contains D and is properly contained in $F(C, D)$, contradicting the minimality of $F(C, D)$. $\square$

In finite dimensions faces are particularly well-behaved.

Theorem 2.6. *For convex sets $D \subset C \subset \mathbb{R}^n$, the smallest face of C containing D is the largest convex subset of C whose relative interior intersects D.*

Proof. We can assume D is nonempty. In this case Theorem 2.5 shows that $\emptyset \neq \text{ri}\, D \subset \text{ri}\, F(C, D)$, so certainly ri $F(C, D)$ intersects D. Suppose on the other hand that E is a convex subset of C with $D \cap \text{ri}\, E$ nonempty. Then $F(C, D) \cap \text{ri}\, E$ is nonempty, so by Theorem 18.1 in [17], $E \subset F(C, D)$. The result follows. $\square$

In general the closed faces of a closed convex set C will have more tractable structure than arbitrary faces. In finite dimensions any face of a closed convex set is closed (Corollary 18.1.1 in [17]). However, this may fail in infinite dimensions. For example, $L_\infty[0, 1]_+$ is clearly a face in $L_1[0, 1]_+$ which is not closed. We therefore define the *smallest closed face* of C containing D, denoted $\text{CF}(C, D)$, to be the intersection of all the closed faces of C containing D. We then obtain the following result, parallel to Theorem 2.5. The proof is identical, simply replacing $F(C, D)$ with $\text{CF}(C, D)$ and observing that the hyperplane H is closed.

Theorem 2.7. *Suppose $C \subset X$ is closed and convex with convex $D \subset C$. Then* qri $D \subset$ qri $\text{CF}(C, D)$. $\square$

This result shows that a maximal convex subset of C whose quasi relative interior intersects D must be a closed face. Unfortunately though, except in finite dimensions (Theorem 2.6), in general it may be strictly larger than the smallest closed face of C containing D as the following example shows. The same phenomenon may be found for instance in Example 5.3 in [14].

Example. Consider the set C in l_3 defined by $C := \{x \in l_3 \mid \|x\|_2 \leqslant 1\}$. Since $\|x\|_r \leqslant \|x\|_p$ for all x provided $1 \leqslant p \leqslant r \leqslant +\infty$, we have $l_{3/2} \subset l_2 \subset l_3$, and the identity map embedding l_2 in l_3 is continuous. The unit ball in l_2 is weakly compact, so C is weakly compact in l_3.

Pick any $\bar{x}$ with $\|\bar{x}\|_2 = 1$ but $\bar{x} \notin l_{3/2}$, for example,

$$\bar{x}_n := \left(\sum_{m=1}^{\infty} m^{-4/3} \right)^{-1/2} n^{-2/3} .$$

Since $\|\cdot\|_2$ is a strictly convex norm, $\bar{x}$ is an extreme point of C, so

$$\text{CF}(C, \{\bar{x}\}) = F(C, \{\bar{x}\}) = \{\bar{x}\} .$$

On the other hand, suppose $x^* \in N_C(\bar{x})$, so $x^* \in l_{3/2}$ and $x^*(x - \bar{x}) \leqslant 0$ whenever $\|x\|_2 \leqslant 1$. Thus

$$\|x^*\|_2=\sup\{x^*(x) \mid \|x\|_2 \leqslant 1\} \leqslant x^*(\bar{x}) \leqslant \|x^*\|_2\|\bar{x}\|_2,$$

by Cauchy–Schwarz, so we have equality throughout. Hence $x^* = \mu\bar{x}$ for some $\mu \in \mathbb{R}$, and since $x^* \in l_{3/2}$ whereas $\bar{x} \notin l_{3/2}$ we deduce that $x^* = 0$. Thus $N_C(\bar{x}) = \{0\}$ so $\bar{x} \in \operatorname{qri} C$. Thus the largest convex subset of C whose quasi relative interior intersects $\{\bar{x}\}$ is simply C.

The case of a lattice cone is more straightforward.

Proposition 2.8. *If X is a normed lattice then the closed faces of X_+ are precisely the sets $\bar{I} \cap X_+$, where $\bar{I}$ is a closed ideal.*

Proof. Any such set is clearly a closed face by Proposition 2.3. On the other hand, any face has the form $I \cap X_+$ for some ideal I. If I is not closed then for some sequence $x_n \to x$ with x_n in I, $x \notin I$, so either x^+ or $x^- \notin I$. Without loss of generality suppose $x^+ \notin I$. Since the lattice operations are continuous, $x_n^+ \to x^+$, so $I \cap X_+$ is not closed. □

When X is a normed lattice the closure of an ideal is an ideal, so the closed ideal generated by D is $\operatorname{cl} I(D)$, and thus for any $D \subset X_+$,

$$\mathrm{CF}(X_+, D) = (\operatorname{cl} I(D)) \cap X_+ = \operatorname{cl} \bigcup_{n=1}^{\infty} \{n[0, x] \mid x \in D\} . \tag{2.1}$$

To see this, observe that $(\operatorname{cl} I(D)) \cap X_+$ is a closed face by Proposition 2.3. On the other hand, by Proposition 2.8, $\mathrm{CF}(X_+, D) = \bar{I} \cap X_+$ for some closed ideal $\bar{I}$. Since $D \subset \bar{I}$ it follows that $\operatorname{cl} I(D) \subset \bar{I}$, and hence that

$$(\operatorname{cl} I(D)) \cap X_+ \subset \bar{I} \cap X_+ = \mathrm{CF}(X_+, D) \subset \operatorname{cl} I(D) \cap X_+ .$$

Recall that for a convex cone K in X, the *polar cone* K° is defined by

$$K^\circ = \{x^* \in X^* \mid x^*(x) \leqslant 0 \text{ for all } x \in K\}.$$

Lemma 2.9. *Suppose X is a normed lattice and J is a closed ideal in X. Then $x_0 \in \operatorname{qri}(J \cap X_+)$ if and only if $J \cap X_+ = (\operatorname{cl} I(\{x_0\})) \cap X_+$.*

Proof. Suppose $x_0 \in \operatorname{qri}(J \cap X_+)$. The containment in one direction is clear. In the opposite direction, by the Bipolar Theorem [13] we need to show

$$(J \cap X_+)^\circ \supset \left(\operatorname{cl} \bigcup_{n=1}^{\infty} n[0, x_0]\right)^\circ ,$$

so suppose $x^* \in (\mathrm{cl}\ \bigcup_{n=1}^{\infty} n[0, x_0])^\circ$, so $x^*(x) \leqslant 0$ for all x in $[0, x_0]$. It follows that $x_+^*(x_0) \leqslant 0$ (Propositions II.4.2 and II.5.5 in [19]), so $x_+^*(x - x_0) \geqslant 0$ for all $x \geqslant 0$, and thus in particular $x_+^* \in N_{J\cap X_+}(x_0)$. Then since $x_0 \in \mathrm{qri}(J\cap X_+)$, $-x_+^* \in N_{J\cap X_+}(x_0)$, or $x_+^*(x - x_0) = 0$ for all x in $J \cap X_+$. Setting $x = 0$ implies $x_+^*(x_0) = 0$ and $x_+^*(x) = 0$ for all x in $J \cap X_+$, so $x^*(x) \leqslant 0$ for all x in $J \cap X_+$, or $x^* \in (J \cap X_+)^\circ$ as required.

Conversely, suppose $J \cap X_+ = (\mathrm{cl}\ I(\{x_0\})) \cap X_+$, and $x^* \in N_{J\cap X_+}(x_0)$, so $x^*(x - x_0) \leqslant 0$ for all x in $J \cap X_+$. Setting $x := \frac{1}{2}x_0$ and $2x_0$ shows $x^*(x_0) = 0$ and $x^*(x) \leqslant 0$ for all x in $J \cap X_+$ and in particular in $[0, x_0]$. Thus $x_+^*(x_0) \leqslant 0$, so in fact $x_+^*(x_0) = 0$, and hence $x_-^*(x_0) = 0$. This says that $-x^*(x) \leqslant 0$ for all x in $[0, x_0]$ and thus in $J \cap X_+$. Thus $-x^* \in N_{J\cap X_+}(x_0)$, so $x_0 \in \mathrm{qri}(J \cap X_+)$. $\square$

We can now deduce the analogue of Theorem 2.6 in the case of a lattice cone.

Theorem 2.10. *Suppose X is a normed lattice with convex $D \subset X_+$ having* qri $D \neq \emptyset$. *Then the largest convex subset of X_+ whose quasi relative interior intersects D is* $\mathrm{cl}\ I(D) \cap X_+$.

Proof. We have

$$\emptyset \neq \mathrm{qri}\ D \subset D \cap \mathrm{qri}\ \mathrm{CF}(X_+, D) = D \cap \mathrm{qri}((\mathrm{cl}\ I(D)) \cap X_+),$$

by Theorem 2.7 and (2.1). On the other hand, if $E \subset X_+$ is convex with $x_0 \in D \cap \mathrm{qri}\ E$, then

$$x_0 \in \mathrm{qri}\ \mathrm{CF}(X_+, E) = \mathrm{qri}((\mathrm{cl}\ I(E)) \cap X_+),$$

by Theorem 2.7 and (2.1). Then by Lemma 2.9,

$$E \subset (\mathrm{cl}\ I(E)) \cap X_+ = (\mathrm{cl}\ I(\{x_0\})) \cap X_+ \subset (\mathrm{cl}\ I(D)) \cap X_+,$$

so $\mathrm{cl}\ I(D) \cap X_+$ is largest. $\square$

3. Facial reduction

Throughout this section we shall suppose X is a locally convex topological vector space, $f: X \to (-\infty, +\infty]$ is a closed, convex function, $A: X \to \mathbb{R}^n$ is continuous and linear, K is a polyhedral cone in $\mathbb{R}^n$ and $b \in \mathbb{R}^n$. We consider the partially-finite convex program

$$(\mathrm{P}) \qquad \begin{cases} \inf & f(x) \\ \text{subject to} & Ax - b \in K, \\ & x \in X. \end{cases}$$

Let us define $D = \{x \in \mathrm{dom}\, f \mid Ax - b \in K\}$, the *feasible region*. Thus x is *feasible* if $x \in D$. The following are the classical conditions for a Lagrange multiplier vector $\lambda \in \mathbb{R}^n$ at a proposed optimal solution x_0 for (P):

(A) *Dual feasibility*: $\lambda \in K^\circ$.

(B) x_0 *minimizes the Lagrangian*, $f(x) + \lambda^{\mathrm{T}} Ax$.

(C) *Complementary slackness*: $\lambda^{\mathrm{T}}(Ax_0-b)=0$.

Proposition 3.1 (Sufficient conditions). *If x_0 is feasible and there exists λ in $\mathbb{R}^n$ satisfying conditions* (A), (B) *and* (C) *then x_0 is optimal.*

Proof. For any feasible x,

$$f(x) \geqslant f(x)+\lambda^{\mathrm{T}}(Ax-b) \geqslant f(x_0)+\lambda^{\mathrm{T}}(Ax_0-b)=f(x_0)\ ,$$

by applying (A), (B) and (C) in turn. □

To ensure the existence of a Lagrange multiplier we need a constraint qualification:

(CQ) $\qquad (\mathrm{ri}(A\ \mathrm{dom} f))\cap(b+K)\neq\emptyset\ .$

Theorem 3.2 (Necessary conditions). *If x_0 is optimal and* (CQ) *holds then there exists $\lambda\in\mathbb{R}^n$ satisfying conditions* (A), (B) *and* (C).

When $\mathrm{qri}(\mathrm{dom} f)$ is nonempty we can rewrite the constraint qualification (CQ) equivalently as:

(QCQ) $\qquad$ There exists feasible $\hat{x}$ in $\mathrm{qri}(\mathrm{dom} f)$.

The general theory behind these results appears in [3].

An important feature of the Necessary Conditions is that they give information as to the form of any optimal solution. If the constraint qualification (CQ) fails then there may exist no Lagrange multiplier λ and we lose this information. We resolve this by weakening condition (B) to:

(B_E) $\qquad x_0$ minimizes $f(x)+\lambda^{\mathrm{T}}Ax$ over x in E ,

for some convex subset E of X. Providing E contains the feasible region the conditions remain sufficient.

Proposition 3.3. *If x_0 is feasible, $E\supset D$ and there exists $\lambda\in\mathbb{R}^n$ satisfying* (A), (B_E) *and* (C) *then x_0 is optimal.* □

The proof is essentially the same as for Proposition 3.1. The conditions are also necessary, under a suitable constraint qualification,

(CQ_E) $\qquad (\mathrm{ri}(A(E\cap\mathrm{dom} f)))\cap(b+K)\neq\emptyset\ .$

Theorem 3.4. *If x_0 is optimal, the convex set E contains D, and* (CQ_E) *holds then there exists λ satisfying* (A), (B_E) *and* (C).

Proof. Since $E\supset D$, x_0 is optimal for

$$(\mathrm{P}_E) \qquad \begin{cases} \inf & (f+\delta_E)(x) \\ \text{subject to} & Ax-b\in K, \end{cases}$$

(where δ_E is the indicator function for E, taking the value 0 on E and $+\infty$ off E). Applying Theorem 3.2 now gives the result. □

Clearly the larger the set E, the more information is imparted by the optimality condition (B_E). We may as well assume E is a subset of dom f, and it is then natural to identify the largest convex subset E of dom f containing D and satisfying (CQ_E). We can answer this using the results of the last section.

In order to satisfy (CQ_E), the set $A(E\cap \operatorname{dom} f)=\mathrm{AE}$ must be a convex subset of A dom f whose relative interior intersects $(b+K)\cap A \operatorname{dom} f$. The largest convex set achieving this is $F(A \operatorname{dom} f, (b+K)\cap A \operatorname{dom} f)$, by Theorem 2.6. Thus we must have $E\subset A^{-1}F(A \operatorname{dom} f, (b+K)\cap A \operatorname{dom} f)$, so the largest possible set E will be given by

$$E=\operatorname{dom} f\cap A^{-1}F(A \operatorname{dom} f,\ (b+K)\cap A \operatorname{dom} f)\ . \tag{3.1}$$

The difficulty in using (3.1), just as with the original constraint qualification (CQ), is the need to work with A dom f. In practice dom f may have a simple structure (such as a lattice cone) whereas A dom f does not. This is the motivation for replacing (CQ) by (QCQ). We therefore seek a set E which may be easily computed using dom f rather than A dom f. One other point is worth emphasizing. In many of the examples that we wish to consider, the domain of f is not closed. In order to capitalize on underlying lattice structure, we need instead to work with the closure of the domain of f.

Theorem 3.5. *Let* $E=\mathrm{CF}(\mathrm{cl} \operatorname{dom} f, D)$. *If*

(i) qri $D\neq\emptyset$, *and*

(ii) $\mathrm{cl}(E\cap \operatorname{dom} f)=E$,

then (CQ_E) *holds, and so* (A), (B_E) *and* (C) *are necessary and sufficient for a feasible* x_0 *to be optimal for* (P).

Proof. By (i) we can choose x in qri D, and Theorem 2.7 shows $x\in$ qri E. Since $x\in E\cap$ dom f, (ii) shows that $x\in \mathrm{qri}(E\cap \operatorname{dom} f)$, so it follows that $Ax\in \mathrm{ri}(A(E\cap \operatorname{dom} f))$, by Proposition 2.10 in [3]. It follows that (CQ_E) holds, and Proposition 3.3 and Theorem 3.4 give the result. □

Corollary 3.6. *Suppose that X is either a separable normed space, or the dual of a separable normed space with the weak* topology, and* dom f *is closed. Let* $E=\mathrm{CF}(\operatorname{dom} f, D)$. *Then provided that the problem* (P) *is consistent,* (CQ_E) *holds. Thus conditions* (A), (B_E) *and* (C) *are necessary and sufficient for a feasible* x_0 *to be optimal for* (P).

Proof. In either case the feasible region D has nonempty quasi relative interior, by Theorem 2.19 in [3], while part (ii) of Theorem 3.5 is obviously satisfied. □

Condition (i) in Theorem 3.5 holds more generally.

Lemma 3.7. *Suppose that X is either a separable normed space, or the dual of a separable normed space with the weak* topology, and $g: X \to (-\infty, +\infty]$ is closed, convex and proper. Then* qri(dom g) *is nonempty.*

Proof. Choose α in $(\inf g, +\infty)$ and define the level set

$$L_\alpha = \{x \in X \mid g(x) \leqslant \alpha\} .$$

Since L_α is closed, convex and nonempty, as before qri L_α is nonempty. Furthermore it is easy to verify that L_α and dom g have the same affine span (Lemma 4.1 in [6]). It follows that $\emptyset \neq \text{qri}\, L_\alpha \subset \text{qri}(\text{dom}\, g)$. □

We can use this result to check (i) in Theorem 3.5. Assuming the conditions of the above lemma, and that (P) is consistent, we simply define

$$g(x) = \begin{cases} f(x) & \text{if } Ax - b \in K , \\ +\infty & \text{otherwise} . \end{cases} \tag{3.2}$$

Then $\emptyset \neq \text{qri}(\text{dom}\, g) = \text{qri}\, D$ as required.

As we shall see, many of the objective functions in which we are interested do not have closed domain. To validate condition (ii) for these objectives we return to the lattice setting.

Lemma 3.8. *Suppose X is a separable, order complete Banach lattice. Suppose G is a dense, saturated subset of X containing* 0, *and J is a closed ideal in X. Then $J \cap G$ is dense in $J \cap X_+$.*

Proof. We use some results from [19]. The closed ideal J is a band, by Theorem II.5.14 and its corollary, and therefore a projection band by Theorem II.2.10. Let $P: X \to J$ be the associated band projection, which is positive by Proposition II.2.7 and continuous by Proposition II.5.2. Suppose $x \in J \cap X_+$. Since G is dense in X_+ there is a sequence $(x_n)_1^\infty$ in G with $x_n \to x$, so $Px_n \to Px = x$. Since the projection onto $J^\perp$ is also positive, $Px_n \in [0, x_n]$, and since G is saturated, $Px_n \in J \cap G$ for each n, which proves the result. □

Theorem 3.9. *Suppose X is a separable, order complete Banach lattice and* dom f *is a dense, saturated subset of X_+ containing* 0. *Let $E = X_+ \cap \text{cl}\, I(D)$. Then provided that* (P) *is consistent,* (CQ_E) *holds. Thus* (A), (B_E) *and* (C) *are necessary and sufficient for a feasible x_0 to be optimal for* (P).

Proof. Defining g by (3.2) and applying Lemma 3.7 shows condition (i) in Theorem 3.5 holds. By (2.1), $E = \text{CF}(\text{cl dom}\, f, D)$. Applying Lemma 3.8 with $G = \text{dom}\, f$ and $J = \text{cl}\, I(D)$ gives

$$\mathrm{cl}(E\cap \mathrm{dom}\, f) = \mathrm{cl}(J\cap X_+ \cap \mathrm{dom}\, f) = \mathrm{cl}(J\cap G) = J\cap X_+ = E,$$

so condition (ii) in Theorem 3.5 holds and the result follows. □

Notice that Theorem 2.10 shows that this result is the best possible in the sense that the chosen set E is the largest convex subset of X_+ whose quasi relative interior contains a feasible point.

Notice also that (CQ_E) is computationally significant. When it is satisfied we can compute a Lagrange multiplier λ satisfying conditions (A), (B_E) and (C) by solving the dual of the problem (P_E):

$$(\mathrm{P}_E^*) \qquad \begin{cases} \sup & -b^{\mathrm{T}}\lambda - (f+\delta_E)^*(-A^*\lambda) \\ \text{subject to} & \lambda \in K^{\circ}. \end{cases}$$

Notice that this problem is finite-dimensional, and a primal optimal solution may often be recaptured from a dual optimal solution (see [3] and [4]).

4. Examples in L_p

In this section we shall examine the consequences of the previous results in the specific case of L_p spaces. We shall suppose throughout that (S, μ) is a totally σ-finite measure space, with no atoms of infinite measure, and $1 \leqslant p < +\infty$. The closed ideals of $L_p(S, \mu)$ are then precisely the sets $\{x \mid x(s) = 0$ a.e. on $S_0\}$ for measurable subsets S_0 of S (see p. 157 in [19]). Note furthermore that if $L_p(S, \mu)$ is separable (as is the case for example if S is a Lebesgue measurable subset of $\mathbb{R}^m$ with μ Lebesgue measure) then if D is nonempty, closed and convex we know qri $D \neq \emptyset$, so we can apply Theorem 2.10.

Theorem 4.1. *Suppose D is an arbitrary subset of $L_p(S, \mu)$, and for some u in $L_p(S, \mu)$, $x \geqslant u$ for all x in D. Then there exists a largest measurable subset S_u^D on which $x(s) = u(s)$ a.e. for all x in D. Furthermore,*

$$\mathrm{cl}\, I(D-u) = \{x \mid x(s) = 0 \text{ a.e. on } S_u^D\},$$

and if D is convex with x_0 in qri D *then $x_0(s) > u(s)$ a.e. on $(S_u^D)^c$.*

Proof. Suppose without loss of generality $u = 0$. From the structure of the closed ideals we know cl $I(D) = \{x \mid x(s) = 0$ a.e. on $S_0\}$, for some measurable subset S_0 of S. Clearly S_0 is the largest measurable subset of S on which $x(s) = 0$ a.e. for all x in D: if S_1 is any other such set, $D \subset \{x \mid x(s) = 0$ a.e. on $S_1\}$, which is a closed ideal, so contains cl $I(D)$, and thus $\mu(S_1 \backslash S_0) = 0$.

The last assertion follows from the fact that

$$\mathrm{qri}\, D \subset \mathrm{qri}\, \mathrm{CF}((L_p)_+, D) = \mathrm{qri}((L_p)_+ \cap \mathrm{cl}\, I(D)),$$

by Theorem 2.7 and (2.1). Thus by Lemma 2.9 we have

$$\begin{aligned}\{x\geqslant 0 \mid x(s)=0 \text{ a.e. on } S_0\} &= (L_p)_+ \cap \operatorname{cl} I(D) \\ &= (L_p)_+ \cap \operatorname{cl} I(\{x_0\}) \\ &= \{x\geqslant 0 \mid x(s)=0 \text{ a.e. on } S_2\}\,,\end{aligned}$$

where $S_2=\{s\in S \mid x_0(s)=0\}$. Thus $x_0(s)>0$ a.e. on $(S_0)^c$ as required. □

Example. Suppose that $L_p(S,\mu)$ is separable and that D is a nonempty closed convex subset of the order interval $[u, v]$. We know from Theorem 4.1 that there is a largest measurable subset S_u^D on which every x in D satisfies $x(s)=u(s)$ a.e., and a largest measurable subset S_v^D on which every x in D satisfies $x(s)=v(s)$ a.e.: we identify these sets via the closed ideals generated by $D-u$ and $v-D$ respectively. We now claim that

$$\operatorname{CF}([u,v],D)=\left\{x\in[u,v]\,\middle|\,x(s)=\begin{cases}u(s) \text{ a.e. on } S_u^D\\ v(s) \text{ a.e. on } S_v^D\end{cases}\right\}. \tag{4.1}$$

This is clearly a closed face containing D.

We know by the separability assumption that we can choose an x_0 in qri D. By Theorem 4.1, $x_0(s)\in(u(s), v(s))$ a.e. on $S_w=(S_u^D\cup S_v^D)^c$. Denote the right hand side of (4.1) by F, and suppose G is a closed face of $[u, v]$ containing D. We claim that $F\subset G$.

Suppose then that $x\in F$. For each n define

$$S_n=\{s\in S \mid (x_0+n^{-1}(x_0-x))(s)\notin[u(s), v(s)]\}\,,$$

and

$$x_n(s)=\begin{cases}x_0(s) & \text{if } s\in S_n\,,\\ x(s) & \text{otherwise}\,.\end{cases}$$

Thus $x_n\in[u,v]$ and also $x_0+n^{-1}(x_0-x_n)\in[u,v]$, and since

$$x_0=\left(\frac{1}{1+n}\right)x_n+\left(\frac{n}{1+n}\right)(x_0+n^{-1}(x_0-x_n))\,,$$

and G is a face, $x_n\in G$. It remains to show $x_n\to x$ as $n\to\infty$.

By assumption, $S_n\subset S_w$ for all n. But $0\in(u(s)-x_0(s), v(s)-x_0(s))$ a.e. on S_w, so $n^{-1}(x_0-x)(s)\in((u-x_0)(s), (v-x_0)(s))$ for n sufficiently large. Thus $\mu(\bigcap_{n=1}^{\infty}S_n)=0$. But now we have

$$\|x_n-x\|_p^p=\int_{S_n}|x_n-x|^p\,\mathrm{d}\mu\leqslant\int_{S_n}(v-u)^p\,\mathrm{d}\mu,$$

and since $(v-u)^p\chi_{S_n}\downarrow 0$ a.e. in L_1, the monotone convergence theorem implies $x_n\to x$. Thus, as we claimed, $F\subset G$. □

We shall be particularly interested in objective functions f which are normal convex integrals in the sense of [16]. Let us assume then that the integrand $\phi:\mathbb{R}\to(-\infty,+\infty]$ is a closed, proper, convex function satisfying $\phi(0)=0$, and suppose that for some y in $L_q(S, \mu)$ (where $p^{-1}+q^{-1}=1$), $\phi^*(y(\cdot))$ is summable (this holds automatically if (S, μ) is finite). Then $I_\phi : L_p(S, \mu) \to (-\infty, -\infty]$ defined by $I_\phi(x) = \int_S \phi(x(s))\, \mathrm{d}\mu$ is well-defined and closed.

Lemma 4.2. *If* dom $\phi = [0, +\infty)$ *then* dom I_ϕ *is a dense, saturated subset of* $L_p(S, \mu)_+$ *containing* 0.

Proof. Clearly $0 \in \operatorname{dom} I_\phi$. Any nonnegative, simple function x with $\mu\{s \mid x(s) > 0\} < +\infty$ is clearly in dom I_ϕ, and these functions are dense in $L_p(S, \mu)_+$, by Theorem 3.13 in [18]. Finally. suppose $u, v \in \operatorname{dom} I_\phi$ and $u \leqslant z \leqslant v$ in $L_p(S, \mu)$. Then we have

$$y(s)z(s) - \phi^*(y(s)) \leqslant \phi(z(s)) \leqslant \max\{\phi(u(s)), \phi(v(s)\} \quad \text{a.e.}$$

and both the right and left hand sides lie in $L_1(S, \mu)$. Hence so does $\phi(z(\cdot))$, so $z \in \operatorname{dom} I_\phi$. Thus dom I_ϕ is saturated. $\square$

Example. Suppose $L_p(S, \mu)$ is separable, dom $\phi = [0, \infty)$ and $a_1, \ldots, a_n \in L_q(S, \mu)$. Consider the problem

$$\text{(MEM)} \qquad \begin{cases} \inf & \displaystyle\int_S \phi(x(s))\, \mathrm{d}\mu \\ \text{subject to} & \displaystyle\left(\int_S a_i(s)x(s)\, \mathrm{d}\mu\right)_{i=1}^n - b \in K, \\ & x \in L_p(S, \mu)\,, \end{cases}$$

which we assume is consistent.

(i) There is a largest measurable subset S_0 of S on which every feasible x is 0 a.e.

(ii) A feasible x_0 is optimal for (MEM) if and only if there exists a Lagrange multiplier vector λ satisfying

(A′) Dual feasibility: $\lambda \in K^\circ$.

(B′) x_0 minimizes the Lagrangian

$$\int_S \left(\phi(x(s)) + x(s) \sum_{i=1}^n \lambda_i a_i(s)\right) \mathrm{d}\mu \tag{4.2}$$

subject to $x(s) = 0$ a.e. on S_0.

(C′) Complementary slackness:

$\sum_{i=1}^n \lambda_i (\int_S x_0(s) a_i(s)\, \mathrm{d}\mu - b_i) = 0$.

(iii) A Lagrange multiplier may be found by solving the dual problem:

$$\text{(MEM*)} \qquad \begin{cases} \sup & \displaystyle -b^{\mathrm{T}}\lambda - \int_{(S_0)^c} \phi^*\left(-\sum_{i=1}^n \lambda_i a_i(s)\right) \mathrm{d}\mu \\ \text{subject to} & \lambda \in K^\circ\,. \end{cases}$$

Proof. We apply Lemma 4.2, Theorem 3.9 and Theorem 4.1. The Banach lattice $L_p(S, \mu)$ is order complete by Propositions II.8.3 and II.5.5 in [19]. The dual problem (MEM*) is derived as remarked in the previous section, using the fact that $(I_\phi)^* = I_{\phi^*}$ [16]. □

Notice that in the case $p=1$ for example, (MEM) has an optimal solution providing $\lim_{u\to+\infty}\phi(u)/u=+\infty$. If ϕ is strictly convex on $[0, +\infty)$, the unique optimal solution x_0 of (MEM) is given by

$$x_0(s) = \begin{cases} (\phi^*)'\left(-\sum_{i=1}^{n} \lambda_i a_i(s)\right) & \text{if } s \notin S_0, \\ 0 & \text{if } s \in S_0, \end{cases}$$

where λ is a solution of (MEM*). This is easily checked using the ideas in [2] and [5], for example.

The case where

$$\phi(u) = \begin{cases} \dfrac{1}{p} u^p & \text{if } u \geqslant 0, \\ +\infty & \text{if } u < 0, \end{cases}$$

and $K=\{0\}$ now gives the result in [15].

In general, identifying the set S_0 may not be straightforward. In some instances however it is easy to compute. For example, in the convex interpolation problem considered in [15] we could choose S to be the interval $[0, 1]$ with Lebesgue measure, and define the constraint functions a_i to be B-splines:

$$a_i(s) = \begin{cases} 0 & \text{if } s \leqslant s_{i-1} \text{ or } s \geqslant s_{i+1}, \\ 1 & \text{if } s = s_i, \\ \text{linear on } [s_{i-1}, s_i] \text{ and } [s_i, s_{i+1}], \end{cases}$$

for given nodes $0=s_0<s_1<\cdots<s_{n+1}=1$. Now for $K=\{0\}$ it is straightforward to check that

$$S_0 = \bigcup \{[s_{i-1}, s_{i+1}] \mid b_i = 0\},$$

again giving results conciding with those in [15].

Example. To conclude we consider the following upper-bounded version of (MEM):

$$\text{(UMEM)} \qquad \begin{cases} \inf & \int_S \phi(x(s))\, \mathrm{d}\mu \\ \text{subject to} & \left(\int_S a_i(s)x(s)\, \mathrm{d}\mu\right)_{i=1}^n - b \in K, \\ & x \in L_1(S, \mu), \end{cases}$$

where we now assume (S, μ) is finite and dom $\phi = [c, d]$, a compact interval. As before, we assume (UMEM) is consistent. Applying Corollary 3.6 and the example following Theorem 4.1 we obtain

(i) There are largest measurable subsets S_c and S_d of S with every feasible x equal to c a.e. on S_c and d a.e. on S_d.

(ii) A feasible x_0 is optimal for (UMEM) if and only if there exists a Lagrange multiplier λ satisfying conditions (A′) and (C′) of the previous example and such that x_0 minimizes the Lagrangian (4.2) subject to $x(s) = c$ a.e. on S_c and $x(s) = d$ a.e. on S_d.

(iii) A Lagrange multiplier may be found by solving the appropriate dual problem, as remarked in the previous section.

A typical example of a measure of entropy with bounded domain of interest in practice is the Fermi–Dirac entropy:

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

See for instance [12] for other examples.

References

[1] A. Ben-Israel, A. Ben-Tal and S. Zlobec, *Optimality in Nonlinear Programming: a Feasible Directions Approach* (Wiley-Interscience, New York, 1981).

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

[3] J.M. Borwein and A.S. Lewis, "Partially finite convex programming, Part I, Duality theory," *Mathematical Programming B* 57 (1992) 15–48.

[4] J.M. Borwein and A.S. Lewis, "Partially finite convex programming, Part II, Explicit lattice models," *Mathematical Programming B* 57 (1992) 49–84.

[5] J.M. Borwein and A.S. Lewis, "Partially-finite programming in L_1 and the existence of maximum entropy estimates," *SIAM Journal on Optimization* 3 (1993) 248–267.

[6] J.M. Borwein and A.S. Lewis, "Strong rotundity and optimization," *SIAM Journal on Optimization* 4 (1994) 146–158.

[7] J.M. Borwein and H. Wolkowicz, "Facial reduction for a cone-convex programming problem," *Journal of the Australian Mathematical Society (Series A)* 30 (1981) 369–380.

[8] J.M. Borwein and H. Wolkowicz, "A simple constraint qualification in infinite dimensional programming," *Mathematical Programming* 35 (1986) 83–96.

[9] A. Brøndsted, *An introduction to convex polytopes* (Springer-Verlag, New York, 1983).

[10] C.K. Chui, F. Deutsch and J.D. Ward, "Constrained best approximation in Hilbert space," *Constructive Approximation* 6 (1990) 35–64.

[11] C.K. Chui, F. Deutsch and J.D. Ward, "Constrained best approximation in Hilbert space II," *Approximation Theory* 71 (1992) 213–238.

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

[13] R.B. Holmes, *Geometric Functional Analysis and its Applications* (Springer-Verlag, New York, 1975).

[14] V.L. Klee, "Extremal structure of convex sets II," *Mathematische Zeitschrift* 69 (1958) 90–104.

[15] C.A. Micchelli, P.W. Smith, J. Swetits and J.D. Ward, "Constrained L_p approximation," *Constructive Approximation* 1 (1985) 93–102.

[16] R.T. Rockafellar, "Integrals which are convex functionals," *Pacific Journal of Mathematics* 24 (1968) 525–539.
[17] R.T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, N.J., 1970).
[18] W. Rudin, *Real and Complex Analysis* (McGraw-Hill, New York, 1966).
[19] H.H. Schaefer, *Banach Lattices and Positive Operators* (Springer-Verlag, Berlin, 1974).