Mathematical Programming 57 (1992) 49-83
North-Holland

Partially finite convex programming, Part II: Explicit lattice models

J.M. Borwein*

Department of Mathematics, Statistics and Computing Science, Dalhousie University, Halifax, N.S., Canada B3H 3J5

A.S. Lewis

Department of Combinatorics and Optimization, Faculty of Mathematics, University of Waterloo, Waterloo, Ont., Canada N2L 3G1

Received 17 February 1988
Revised manuscript received 31 June 1989; 25 January 1990

In Part I of this work we derived a duality theorem for partially finite convex programs, problems for which the standard Slater condition fails almost invariably. Our result depended on a constraint qualification involving the notion of *quasi relative interior*. The derivation of the primal solution from a dual solution depended on the differentiability of the dual objective function: the differentiability of various convex functions in lattices was considered at the end of Part I. In Part II we shall apply our results to a number of more concrete problems, including variants of semi-infinite linear programming, L^1 approximation, constrained approximation and interpolation, spectral estimation, semi-infinite transportation problems and the generalized market area problem of Lowe and Hurter (1976). As in Part I, we shall use lattice notation extensively, but, as we illustrated there, in concrete examples lattice-theoretic ideas can be avoided, if preferred, by direct calculation.

AMS 1985 Subject Classifications: Primary 90C25, 49B27; Secondary 90C48, 52A07, 65K05.

Key words: Convex programming, duality, constraint qualification, semi-infinite programming, constrained approximation, spectral estimation, transportation problem.

Introduction to Part II

The main result of Part I of this work (Quasi Relative Interiors and Duality Theory) was a duality theorem for a class of problems we have called 'partially finite convex programs'. We developed the following notion of 'quasi relative interior' which appears in the constraint qualification for this result. Suppose X is a topological vector space.

Definition. For convex $C \subset X$, the *quasi relative interior* of C (qri C) is the set of those $x \in C$ for which cl cone$(C-x)$ is a linear subspace.

The main duality result is then the following. For $P \subset \mathbb{R}^n$, the dual cone is denoted by $P^+ := \{y \in \mathbb{R}^n \mid y^T\lambda \geq 0, \forall \lambda \in P\}$. The indicator function of C is denoted by $\delta(\cdot \mid C)$.

* Present address: Department of Combinatorics and Optimization, Faculty of Mathematics, University of Waterloo, Waterloo, Ont., Canada N2L 3G1.

Theorem (Corollary 4.8 in Part I). *Let X be locally convex, $f: X \to]-\infty, \infty]$ convex, $C \subset \operatorname{dom} f$ convex, $A: X \to \mathbb{R}^n$ continuous and linear, $b \in \mathbb{R}^n$ and $P \subset \mathbb{R}^n$ a polyhedral cone. Consider the following dual pair of problems:*

$$\begin{array}{lll}
\text{(CM)} & \inf & f(x) \\
& \text{subject to} & Ax \in b + P, \\
& & x \in C, \\
\text{(DCM)} & \max & -(f + \delta(\cdot \mid C))^*(A^{\mathrm{T}}\lambda) + b^{\mathrm{T}}\lambda \\
& \text{subject to} & \lambda \in P^+.
\end{array}$$

If the following constraint qualification is satisfied,

(CQ) *there exists an $\hat{x} \in \operatorname{qri} C$ which is feasible for* (CM),

then the values of (CM) *and* (DCM) *are equal* (*with attainment in* (DCM)).

Suppose further that $f + \delta(\cdot \mid C)$ is closed. If $\bar{\lambda}$ is optimal for the dual, and $(f + \delta(\cdot \mid C))^*$ is differentiable at $A^{\mathrm{T}}\bar{\lambda}$ with Gateaux derivative $\bar{x} \in X$, then $\bar{x}$ is optimal for (CM), *and is furthermore the unique optimal solution.*

Numerous properties and examples of quasi relative interior were discussed in Part I (Sections 2, 3). As may be seen from the above theorem, the derivation of primal solutions depends on the differentiability of $(f + \delta(\cdot \mid C))^*$. In Section 5 of Part I we therefore studied the differentiability of various convex functions in lattices.

In Part II of this work we shall concentrate on applying these ideas to more concrete models. The first special case of the problem (CM) we consider is when the set C is a cone. If C is the positive cone of a partially ordered vector space, writing down the dual problem involves computing the *monotone conjugate* of the convex function f. As an example we derive duality results for semi-infinite linear programming. We also consider the semi-infinite linear program with an additional norm constraint; the dual problem involves one of the standard penalty functions used in the solution of semi-infinite linear programs. As another example we consider certain quadratic programs in the Hilbert space of square-integrable functions $L^2(T, \mu)$.

The next section (7) deals with another important example, constrained approximation problems. These arise when the function f is a norm; our model includes the constrained interpolation problems considered in Irvine, Marin and Smith (1986), and spectral estimation (see Ben-Tal, Borwein and Teboulle, 1988 and 1989). We consider briefly the numerical treatment of such problems.

The second important special case of the problem (CM) that we consider is when the set C is of the form

$$\left\{ (x_1, \ldots, x_m) \in E^m \,\middle|\, \sum_{i=1}^{m} x_i = e,\ x_i \geqslant 0 \text{ for each } i \right\},$$

where $e \geqslant 0$ is some fixed element of the partially ordered vector space E, and $X = E^m$. When the function f is linear our theory gives an interesting analogue of classical linear programming (including a simple characterization of extreme points), and when E is actually a vector lattice the dual problem has a particularly straightforward structure. As an example we consider semi-infinite linear programming with an additional upper-bound constraint; again it is interesting to observe how a well-known penalty function arises naturally in the dual of this problem. A further example is furnished by L^1-approximation. The final section (9) deals with two more practical examples: the semi-infinite transportation problem considered in Kortanek and Yamasaki (1982), and the *generalized market area problem* (see Lowe and Hurter, 1976).

As in Part I, we shall frequently use ideas and terminology from the theory of vector lattices, which provides a unifying framework for much of this work. However, as we observed in Part I, the reader will find that calculations we perform in lattice notation may be easily followed through in concrete spaces, with no knowledge of vector lattices.

6. The conical case

We are now ready to consider more concrete examples of the convex model (CM) of the introduction. In this section and the next we shall consider the case where the set $C \subset X$ is a cone, partially ordering X. In this case the function $(f + \delta(\cdot \mid C))^*$ appearing in the dual problem (DCM) is the 'monotone conjugate' of the convex function f. We will therefore begin by identifying some circumstances under which this is easy to evaluate, and give some examples. We will then apply our results to certain semi-infinite linear and quadratic programs.

Throughout this section X will be a topological vector space (which we always understand to be Hausdorff) partially ordered by a convex cone K, the topological dual X^* partially ordered by the dual cone,

$$K^+ := \{\phi \in X^* \mid \phi(x) \geqslant 0 \text{ for all } x \in K\},$$

and $f: X \to]-\infty, \infty]$, convex. As always, f^* denotes the *convex conjugate* of $f: f^*(\phi) = \sup\{\phi(x) - f(x) \mid x \in X\}$, for $\phi \in X^*$. The domain of f, $\operatorname{dom} f$, is the set on which f is finite, and we denote the set where f is continuous by $\operatorname{cont} f$. We say f is closed if the epigraph of f,

$$\operatorname{epi} f := \{(x, r) \mid x \in \operatorname{dom} f,\ r \in \mathbb{R},\ r \geqslant f(x)\}$$

is a closed set. The core of a set $C \subset X$ is its algebraic interior: $x \in C$ lies in the core if for all $y \in X$ there exist $\delta > 0$ with $x + ty \in C$ for all $t[0, \delta]$.

The following Fenchel duality result has appeared in various contexts (see for example Rockafellar, 1974; Borwein, 1981b; and Ekeland and Temam, 1976). For completeness we provide a direct proof. A locally convex space X is Fréchet if it is complete metrizable. In particular, Banach spaces are Fréchet.

Theorem 6.1. *Suppose $g, h : X \to]\infty, \infty]$ are convex, and any of the following three conditions holds*:

(i) $\operatorname{cont} g \cap \operatorname{dom} h \neq \emptyset$,

(ii) $\operatorname{dom} g \cap \operatorname{cont} h \neq \emptyset$,

(iii) *X Fréchet, g and h closed and* $0 \in \operatorname{core}(\operatorname{dom} g - \operatorname{dom} h)$.

Then $\inf\{g(x)+h(x) \mid x \in X\} = \max\{-g^*(\phi) - h^*(-\phi) \mid \phi \in X^*\}$, (*when the left-hand side is finite*).

Proof. Define convex relations $F : X \times \mathbb{R} \to \mathbb{R}$, and $H : X \times \mathbb{R} \to X$, by

$$F(x, r) := \{s \in \mathbb{R} \mid s \geq r\},$$

$$H(x, r) := \{y \in X \mid g(x) + h(x+y) \leq r\}.$$

F is lower semi-continuous (Borwein, 1981a, Lemma 4). We wish to show H is open at 0. Suppose first that (ii) holds ((i) will follow by symmetry). Let $x_0 \in \operatorname{dom} g \cap \operatorname{cont} h$, and set $r_0 := g(x_0) + h(x_0)$, so $0 \in H(x_0, r_0)$. Since H is convex it suffices to show H is open at $((x_0, r_0), 0)$ (Borwein, 1981a, Proposition 2), in other words, given any neighbourhood U_1 of 0 in X and U_2 of 0 in $\mathbb{R}$ there exists a neighbourhood V of 0 in X with

$$V \subset \{y \in X \mid g(x) + h(x+y) \leq r, \text{ some } x \in x_0 + U_1,\ r \in r_0 + U_2\}.$$

But for any $\delta > 0$, $0 \in \operatorname{int}\{y \in X \mid g(x_0) + h(x_0 + y) \leq r_0 + \delta\}$, by the continuity of h at x_0, and the desired conclusion follows.

Suppose on the other hand that (iii) holds. Since g and h are closed, H has a closed graph, and also range $H = \operatorname{dom} h - \operatorname{dom} g$, so $0 \in \operatorname{core}(\operatorname{range} H)$. It then follows from the closed graph theorem (Borwein, 1981a, Theorem 8) that H is open at 0.

Now consider the problem

$$\begin{aligned}\mu &:= \inf\{g(x) + h(x) \mid x \in X\} \\ &= \inf\{r \mid g(x) + h(x) \leq r,\ x \in X,\ r \in \mathbb{R}\} \\ &= \inf\{F(x, r) \mid 0 \in H(x, r),\ (x, r) \in X \times \mathbb{R}\}.\end{aligned}$$

Applying the Lagrange multiplier theorem in Borwein (1981b, 3.1), we deduce the existence of a $\theta \in X^*$ for which

$$F(x, r) + \theta(H(x, r)) \geq \mu \quad \text{for all } (x, r) \in X \times \mathbb{R}.$$

We can rewrite this as $s + \theta(y) \geq \mu$, for all $x, y \in X$, $r, s \in \mathbb{R}$, for which $s \geq r$, and $g(x) + h(x+y) \leq r$, so

$$\mu - \theta(y) \leq g(x) + h(x+y) \quad \text{for all } x, y \in X.$$

Thus we have

$$\mu \leq g(x) - \theta(x) + h(x+y) + \theta(x+y) \quad \text{for all } x, y \in X.$$

Taking infs over x and y we deduce $\mu \leq -g^*(\theta) - h^*(-\theta)$. But for all $x \in X$, $\phi \in X^*$,

$$g(x) + g^*(\phi) + h(x) + h^*(-\phi) \geq 0,$$

so $\mu = -g^*(\theta) - h^*(-\theta)$ as required. $\square$

We say f is (K-) *monotonically regular* if, for all $\theta \in X^*$,

$$(f + \delta(\cdot \mid K))^*(\theta) = \min\{f^*(\psi) \mid \psi \geq \theta,\ \psi \in X^*\}.$$

Corollary 6.2. *Under any of the following three conditions, f is K-monotonically regular*:

(i) $\operatorname{dom} f \cap \operatorname{int} K \neq \emptyset$,

(ii) $\operatorname{cont} f \cap K \neq \emptyset$,

(iii) *X Fréchet, f and K closed, and* $0 \in \operatorname{core}(\operatorname{dom} f - K)$ (*which holds in particular when K is generating and* $K \subset \operatorname{dom} f$). *In particular, if* (X, K) *is a Banach lattice and f is closed with* $K \subset \operatorname{dom} f$, *then f is K-monotonically regular.*

Proof. In Theorem 6.1, set $g := f - \theta$ and $h := \delta(\cdot \mid K)$. Then $g^*(\phi) = f^*(\phi + \theta)$ and $h^*(-\phi) = \delta(\phi \mid K^+)$, so the result follows immediately. $\square$

When f is monotonically regular we can use the above result to rewrite the dual pair of problems (CM) and (DCM) in the following way. The primal *conical convex model* becomes

$$\text{(CCM)} \quad \begin{array}{ll} \inf & f(x) \\ \text{subject to} & Ax \in b + P, \\ & x \in K, \quad x \in X, \end{array}$$

and the dual is

$$\text{(DCCM)} \quad \begin{array}{ll} \max & b^{\mathrm{T}}\lambda - f^*(\psi) \\ \text{subject to} & \psi - A^{\mathrm{T}}\lambda \in K^+, \\ & \lambda \in P^+, \quad \lambda \in \mathbb{R}^n, \quad \psi \in X^*. \end{array}$$

Theorem 6.3. *Let X be locally convex, partially ordered by a convex cone K, X^* partially ordered by K^+, $f : X \to]-\infty, \infty]$ a convex, K-monotonically regular function with $K \subset \operatorname{dom} f$, $A : X \to \mathbb{R}^n$ continuous, linear, $b \in \mathbb{R}^n$ and $P \subset \mathbb{R}^n$ a polyhedral cone.*

Suppose there exists $\hat{x} \in \operatorname{qri} K$ with $A\hat{x} - b \in P$. Then the values of (CCM) *and* (DCCM) *are equal, with attainment in* (DCCM). *Furthermore, a primal feasible $\bar{x}$ is optimal if and only if there exist $\bar{\lambda} \in P^+$, $\bar{\psi} \in \partial f(\bar{x})$, with $\bar{\psi} - A^{\mathrm{T}}\bar{\lambda} \in K^+$, $(\bar{\psi} - A^{\mathrm{T}}\bar{\lambda})(\bar{x}) = 0$ and $\bar{\lambda}^{\mathrm{T}}(Ax - b) = 0$.*

Proof. The duality result follows immediately from Corollary 4.8 and Corollary 6.2. It follows that a primal feasible $\bar{x}$ is optimal if and only if there exists dual feasible $(\bar{\lambda}, \bar{\psi})$ with $b^{\mathrm{T}}\bar{\lambda} - f^*(\bar{\psi}) = f(\bar{x})$. But

$$\begin{aligned} b^{\mathrm{T}}\bar{\lambda} - f^*(\bar{\psi}) &\leq b^{\mathrm{T}}\bar{\lambda} - \bar{\psi}(\bar{x}) + f(\bar{x}) \\ &\leq b^{\mathrm{T}}\bar{\lambda} - (A^{\mathrm{T}}\bar{\lambda})(\bar{x}) + f(\bar{x}) \\ &= f(\bar{x}) - (A\bar{x} - b)^{\mathrm{T}}\bar{\lambda} \\ &\leq f(\bar{x}). \end{aligned}$$

Thus $\bar{x}$ is optimal if and only if we have equality throughout, which gives the result, since $f(\bar{x}) + f^*(\bar{\psi}) = \bar{\psi}(\bar{x})$ if and only if $\bar{\psi} \in \partial f(\bar{x})$. □

In some instances the monotone conjugate of f can be computed with no extra effort than the conjugate:

Special cases 6.4. Suppose f is K-monotonically regular.

(i) If f^* is K^+*-isotone* ($\theta_1 \geq \theta_2$ implies $f^*(\theta_1) \geq f^*(\theta_2)$) then $(f + \delta(\cdot \mid K))^*(\phi) = f^*(\phi)$.

(ii) If (X, K) is a normed lattice and f^* is *absolute* ($f^*(|\theta|) = f^*(\theta)$, for all $\theta \in X^*$), and isotone on K^+, then $(f + \delta(\cdot \mid K))^*(\phi) = f^*(\phi^+)$.

Examples 6.5.

(i) X normed, $f(x) = (1/p)\|x\|^p$, $1 < p < \infty$.

$$\begin{aligned} f^*(\theta) &= \sup_x \left\{ \theta(x) - \frac{1}{p}\|x\|^p \right\} \\ &= \sup_{t \geq 0} \sup_{\|x\| = t} \left\{ \theta(x) - \frac{1}{p}\|x\|^p \right\} \\ &= \sup_{t \geq 0} \left\{ \|\theta\|_* t - \frac{1}{p} t^p \right\}. \end{aligned}$$

Differentiating implies the maximum occurs at $t = \|\theta\|_*^{1/(p-1)}$, giving $f^*(\theta) = (1/q)\|\theta\|_*^q$.

If X is a normed lattice then 6.4(ii) applies so

$$(f + \delta(\cdot \mid X_+))^*(\theta) = \frac{1}{q}\|\theta^+\|_*^q.$$

(ii) X normed, $f(x) = M\|x\|$, for some $M > 0$. As above,

$$f^*(\theta) = \sup_{t \geq 0}\{(\|\theta\|_* - M)t\} = \delta(\theta \mid MB_{X^*}),$$

where B_{X^*} is the unit ball in X^*.

Again, if X is a normed lattice then (6.4)(ii) applies:

$$(f+\delta(\cdot\,|\,X_+))^*(\theta)=\delta(\theta^+\,|\,MB_{X^*}).$$

(iii) X normed, $f(x)=\delta(x\,|\,MB_X)$, $M>0$.

$$f^*(\theta)=\sup\{\theta(x)\,|\,x\in MB_X\}=M\|\theta\|_*.$$

Again, if X is a normed lattice then 6.4(ii) applies:

$$(f+\delta(\cdot\,|\,X_+))^*(\theta)=M\|\theta^+\|_*.$$

(iv) $X=L^2(T,\mu), f(x)=\frac{1}{2}\|x-x_0\|_2^2$.

$$f^*(y)=\sup_x\{\langle x,y\rangle-\tfrac{1}{2}\|x-x_0\|^2\}.$$

Differentiating implies the maximum occurs at $x=x_0+y$, giving

$$f^*(y)=\langle x_0+y,y\rangle-\tfrac{1}{2}\langle y,y\rangle=\tfrac{1}{2}\|y+x_0\|^2-\tfrac{1}{2}\|x_0\|^2.$$

By Corollary 6.2,

$$\begin{aligned}(f+\delta(\cdot\,|\,X_+))^*(y)&=\min\{f^*(z)\,|\,z\geqslant y\}\\&=\min\{\tfrac{1}{2}\|z+x_0\|^2\,|\,z\geqslant y\}-\tfrac{1}{2}\|x_0\|^2\\&=\min\{\tfrac{1}{2}\|u\|^2\,|\,u\geqslant y+x_0\}-\tfrac{1}{2}\|x_0\|^2\\&=\tfrac{1}{2}\|(y+x_0)^+\|^2-\tfrac{1}{2}\|x_0\|^2.\end{aligned}$$

(v) $X=L^p(T,\mu)$, $1\leqslant p<\infty$,

$$f(x)=\begin{cases}\int_T-\log x(t)\,\mathrm{d}\mu(t), & x(t)>0 \text{ a.e.},\\ \infty, & \text{otherwise.}\end{cases}$$

The integrand above is a normal convex integrand in the sense of Rockafellar (1968), and so $f^*: L^q(T,\mu)\to\,]-\infty,\infty]$ is given by

$$f^*(y)=\begin{cases}\int_T(-1-\log(-y(t)))\,\mathrm{d}\mu(t), & y(t)<0 \text{ a.e.},\\ \infty, & \text{otherwise.}\end{cases}$$

In this case 6.4(i) holds, so $(f+\delta(\cdot\,|\,X_+))^*=f^*$.
(vi) $X=L^p(T,\mu)$, $1\leqslant p<\infty$,

$$f(x)=\begin{cases}\int_T x(t)(\log x(t)-1)\,\mathrm{d}\mu(t), & x(t)>0 \text{ a.e.},\\ \infty, & \text{otherwise.}\end{cases}$$

Again we have a normal convex integrand, so $f^*: L^q(T,\mu)\to\,]-\infty,\infty]$ is given by $f^*(y)=\int_T \mathrm{e}^{y(t)}\,\mathrm{d}\mu(t)$. Again 6.4(i) holds so $(f+\delta(\cdot\,|\,X_+))^*=f^*$.

More generally than (v) and (vi), we can consider *entropic objectives* (see Ben-Tal, Borwein and Teboulle, 1992; and Borwein and Lewis, 1991).

Semi-infinite linear programming

The first special case of (CCM) we shall consider will be the case when f is just a continuous linear functional. In this case (CCM) becomes a *semi-infinite linear program*:

$$\begin{array}{lll} \text{(SILP)} & \inf & \theta(x) \\ & \text{subject to} & Ax \in b+P, \\ & & x \geq 0, \quad x \in X, \\ \text{(DSILP)} & \max & b^{\mathrm{T}}\lambda \\ & \text{subject to} & A^{\mathrm{T}}\lambda \leq \theta, \\ & & \lambda \in P^{+}, \quad \lambda \in \mathbb{R}^{n}. \end{array}$$

Corollary 6.6. *Let X be locally convex, partially ordered by a convex cone K, X^* partially ordered by K^+, $\theta \in X^*$, $A: X \to \mathbb{R}^n$ continuous, linear, $b \in \mathbb{R}^n$ and $P \subset \mathbb{R}^n$ a polyhedral cone.*

Suppose there exists $\hat{x} \in \operatorname{qri} K$ with $A\hat{x} - b \in P$. Then the values of (SILP) *and* (DSILP) *are equal, with attainment in* (DSILP). *Furthermore, a primal feasible $\bar{x}$ is optimal if and only if there exists $\bar{\lambda} \in P^+$ with $A^{\mathrm{T}}\bar{\lambda} \leq \theta$, $(\theta - A^{\mathrm{T}}\bar{\lambda})(\bar{x}) = 0$ and $\bar{\lambda}^{\mathrm{T}}(Ax - b) = 0$.*

Proof. This follows directly from Theorem 6.3, using the fact that $\theta^*(\phi) = \delta(\phi \mid \{\theta\})$ and θ is clearly monotonically regular. $\square$

The cone $AK \subset \mathbb{R}^n$ is known as the "moment cone" in the semi-infinite linear programming literature (see for example Glashoff and Gustafson, 1983). When $\operatorname{qri} K \neq \emptyset$, by Proposition 2.10, $A \operatorname{qri} K = \operatorname{ri} AK$. If K is generating and A is onto it follows that our constraint qualification is equivalent to $(b+P) \cap \operatorname{int}(AK) \neq \emptyset$, which is the classical 'superconsistency' constraint qualification (see Karlin and Studden, 1966).

Norm constrained semi-infinite linear programming

The next program we consider does not strictly speaking fit the model (CCM), but behaves in a very similar fashion. We consider the previous problem (SILP), but with an additional norm constraint, $\|x\| \leq M$ for some $M > 0$.

$$\begin{array}{lll} \text{(NLP)} & \inf & \theta(x) \\ & \text{subject to} & Ax \in b+P, \\ & & x \geq 0, \quad \|x\| \leq M, \quad x \in X. \end{array}$$

We shall now assume X is a normed lattice. The dual problem then becomes

$$\begin{array}{lll} \text{(DNLP)} & \max & b^{\mathrm{T}}\lambda - M\|(A^{\mathrm{T}}\lambda - \theta)^{+}\|_{*} \\ & \text{subject to} & \lambda \in P^{+}, \quad \lambda \in \mathbb{R}^{n}. \end{array}$$

Notice that (DNLP) is a penalty-function version of the problem (DSILP). The larger the constant M, the more we penalize violations of the constraint $A^{\mathrm{T}}\lambda \leqslant \theta$.

Theorem 6.7. *Let X be a normed lattice, $\theta \in X^*$, $A: X \to \mathbb{R}^n$ continuous, linear, $b \in \mathbb{R}^n$ and $P \subset \mathbb{R}^n$ a polyhedral cone. Suppose there exists $\hat{x} \in \operatorname{qri} X_+$ with $\|\hat{x}\| < M$ and $A\hat{x} \in b + P$. Then the values of* (NLP) *and* (DNLP) *are equal, with attainment in* (DNLP). *Furthermore, a primal feasible $\bar{x}$ is optimal if and only if there exists $\bar{\lambda} \in P^+$ satisfying the following four conditions*:

(i) $\bar{\lambda}^{\mathrm{T}}(A\bar{x} - b) = 0$,

(ii) $(A^{\mathrm{T}}\bar{\lambda} - \theta)^-(\bar{x}) = 0$,

(iii) $\|(A^{\mathrm{T}}\bar{\lambda} - \theta)^+\|_* \|\bar{x}\| = (A^{\mathrm{T}}\bar{\lambda} - \theta)^+(\bar{x})$,

(iv) $\|\bar{x}\| = M$ *if* $A^{\mathrm{T}}\bar{\lambda} - \theta \not\leqslant 0$.

If in fact $X = Y^*$ for a normed lattice Y, $\theta \in Y$, and A is $\sigma(X, Y)$-$\mathbb{R}^n$ continuous, then the value of (NLP) *will also be attained.*

Proof. Apply Corollary 4.8 with $f := \theta$, $C := X_+ \cap MB_X$. Then

$$\begin{aligned}(f + \delta(\cdot \mid C))^*(\phi) &= \sup\{\phi(x) - \theta(x) \mid x \geqslant 0,\ \|x\| \leqslant M\} \\ &= \sup\{(\phi - \theta)(x) - \delta(x \mid MB_X) \mid x \geqslant 0\} \\ &= M\|(\phi - \theta)^+\|_*,\end{aligned}$$

by 6.5(iii). The form of the dual problem then follows.

The form of the constraint qualification follows from the fact that $\operatorname{qri} C = \operatorname{qri} X_+ \cap \operatorname{int}(MB_X)$, by Theorem 2.13. A primal feasible $\bar{x}$ is optimal if and only if there exists $\bar{\lambda} \in P^+$ with

$$b^{\mathrm{T}}\bar{\lambda} - M\|(A^{\mathrm{T}}\lambda - \theta)^+\|_* = \theta(\bar{x}).$$

But

$$\begin{aligned}\theta(\bar{x}) &\geqslant \theta(\bar{x}) - \bar{\lambda}^{\mathrm{T}}(A\bar{x} - b) \\ &= b^{\mathrm{T}}\bar{\lambda} - (A^{\mathrm{T}}\bar{\lambda} - \theta)(\bar{x}) \\ &\geqslant b^{\mathrm{T}}\bar{\lambda} - (A^{\mathrm{T}}\bar{\lambda} - \theta)^+(\bar{x}) \\ &\geqslant b^{\mathrm{T}}\bar{\lambda} - \|(A^{\mathrm{T}}\bar{\lambda} - \theta)^+\|_* \|\bar{x}\| \\ &\geqslant b^{\mathrm{T}}\bar{\lambda} - M\|(A^{\mathrm{T}}\bar{\lambda} - \theta)^+\|_*,\end{aligned}$$

so we must have equality throughout; in other words $\bar{x}$ is optimal if and only if conditions (i), (ii), (iii) and (iv) hold.

The final set of conditions is sufficient to ensure that the feasible region of (NLP) is a $\sigma(X, Y)$-closed subset of MB_X, and so is $\sigma(X, Y)$-compact by the Alaoglu–Bourbaki Theorem. Since θ is $\sigma(X, Y)$-continuous, the infimum will be attained. □

The question of when the dual objective function is differentiable can now be addressed, using the results of Section 5. In particular, if $A^{\mathrm{T}}\lambda \not\leq \theta$ and $\|\cdot\|_*$ is differentiable at $(A^{\mathrm{T}}\lambda-\theta)^+$ (with $\nabla\|(A^{\mathrm{T}}\lambda-\theta)^+\|_* \in X$), then the derivative of the dual objective function at λ is $b - MA\nabla\|(A^{\mathrm{T}}\lambda-\theta)^+\|_*$, by Proposition 5.6.

Suppose $\bar\lambda$ is dual optimal with $A^{\mathrm{T}}\bar\lambda \not\leq \theta$, and $\|\cdot\|_*$ differentiable at $(A^{\mathrm{T}}\bar\lambda-\theta)^+$, with $\nabla\|(A^{\mathrm{T}}\bar\lambda-\theta)^+\|_* \in X$. Set $\bar x := M\nabla\|(A^{\mathrm{T}}\bar\lambda-\theta)^+\|$. By Corollary 4.8, $\bar x$ is the unique optimal solution of the primal problem (NLP).

Example 6.8. $X = L^p(T,\mu)$, $1<p<\infty$. Suppose $\bar\lambda$ is optimal for (DNLP) with $A^{\mathrm{T}}\bar\lambda \not\leq \theta$. Then by the above and Examples 5.7(i), the optimal solution of (NLP) is given by

$$\bar x = M\|(A^{\mathrm{T}}\bar\lambda-\theta)^+\|_q^{1-q}((A^{\mathrm{T}}\bar\lambda-\theta)^+)^{q-1}.$$

Semi-infinite quadratic programming in L^2

The last example in this section will be the following quadratic program.

$$\text{(QP)} \quad \inf \quad \tfrac12\|x-x_0\|_2^2$$
$$\text{subject to} \quad Ax \in b+P,$$
$$x \geq 0, \quad x\in L^2(T,\mu).$$

Here, (T,μ) is a σ-finite measure space, $x_0\in L^2$, and as usual $A: L^2\to\mathbb{R}^n$ is continuous and linear, $b\in\mathbb{R}^n$ and $P\subset\mathbb{R}^n$ is a polyhedral cone.

Using Examples 6.5(iv) we obtain the dual problem (from Corollary 4.8):

$$\text{(DQP)} \quad \max \quad b^{\mathrm{T}}\lambda - \tfrac12\|(A^{\mathrm{T}}\lambda+x_0)^+\|_2^2 + \tfrac12\|x_0\|_2^2$$
$$\text{subject to} \quad \lambda\in P^+, \quad \lambda\in\mathbb{R}^n.$$

More general constraints of the form $x\geq 0$ on $T_1\subset T$, $x\leq 0$ on $T_2\subset T$, can be easily handled using the remarks on separable problems at the end of Section 4.

Theorem 6.9. *If there exists a feasible $\hat x$ for* (QP) *with $\hat x(t)>0$ μ-a.e., then the values of* (QP) *and* (DQP) *are equal and both are attained.*

The dual objective function is everywhere differentiable, with derivative $b - A(A^{\mathrm{T}}\lambda+x_0)^+$. Furthermore, if $\bar\lambda$ is optimal for the dual problem (DQP), *then the unique optimal solution of* (QP) *is given by $\bar x = (a^{\mathrm{T}}\bar\lambda + x_0)^+$.*

Proof. By Examples 3.11(i), $\mathrm{qri}(L^2(T,\mu)_+) = \{x>0 \text{ a.e.}\}$, and the duality result follows from Corollary 4.8. By Examples 5.7(i), $\nabla(\frac12\|x^+\|_2^2) = x^+$, for $x\not\leq 0$, and is clearly 0 for $x\leq 0$. The expression for the derivative follows, and the derivation of the unique primal solution follows from Corollary 4.8. $\square$

7. Constrained approximation

Throughout this section we shall concentrate on one particular example of the conical case of the convex model (the problem (CCM) of the previous section). This problem requires the minimization of the norm of the variable, subject to positivity constraints and a finite number of linear inequalities. Such problems arise in particular in questions concerning 'best' convex interpolants to given data (see Micchelli, Smith, Swetits and Ward, 1985 and Irvine, Marin and Smith, 1986), and in spectral estimation (Ben Tal, Borwein and Teboulle, 1988 and 1992). We shall see that many of the above authors' results are special cases of our general duality theorems.

Throughout this section we shall adopt the following notation (see Schaefer, 1971 and 1974, for definitions):

$$\begin{aligned}
&(X, Y) \text{ a dual pair,}\\
&\text{with } (X, \|\cdot\|) \text{ a normed lattice, and}\\
&Y \text{ a sublattice of } X^* \text{ (norm } \|\cdot\|_*, \text{ positive cone } (X_+)^+),\\
&A: X \to \mathbb{R}^n,\ \sigma(X, Y)\text{-}\mathbb{R}^n \text{ continuous, linear,}\\
&b \in \mathbb{R}^n, \quad P \subset \mathbb{R}^n \text{ a polyhedral cone.}
\end{aligned} \tag{7.1}$$

For $1 \leqslant p < \infty$ we consider the problem

$$\begin{aligned}
(\mathrm{CA}_p) \quad &\inf &&(1/p)\|x\|^p\\
&\text{subject to} &&Ax \in b + P,\\
& &&x \geqslant 0, \quad x \in X.
\end{aligned}$$

For $1 < p < \infty$, and $1/p + 1/q = 1$, by Examples 6.5(i) and Corollary 4.8, the dual becomes

$$\begin{aligned}
(\mathrm{DCA}_p) \quad &\max &&b^{\mathrm{T}}\lambda - (1/q)\|(A^{\mathrm{T}}\lambda)^+\|_*^q\\
&\text{subject to} &&\lambda \in P^+, \quad \lambda \in \mathbb{R}^n,
\end{aligned}$$

while for $p = 1$, using Examples 6.5(ii), we obtain

$$\begin{aligned}
(\mathrm{DCA}_1) \quad &\max &&b^{\mathrm{T}}\lambda\\
&\text{subject to} &&\|(A^{\mathrm{T}}\lambda)^+\|_* \leqslant 1,\\
& &&\lambda \in P^+, \quad \lambda \in \mathbb{R}^n.
\end{aligned}$$

The two most significant cases of the pair (X, Y) in (7.1) are when $Y = X^*$ (the norm case), and when Y is a normed lattice with $X = Y^*$ (the weak* case, cf. Theorem 3.5).

Theorem 7.1. *If $\|\cdot\|$ is strictly convex then any optimal solution of* (CA_p) *is unique. If there exists an $\hat{x} \in \sigma(X, Y) - \mathrm{qri}\, X_+$, feasible for* (CA_p), *then the values of* (CA_p) *and* (DCA_p) *are equal, with attainment in* (DCA_p). *Suppose further that $-b \notin P$. If $\|\cdot\|_*$ is differentiable at $(A^{\mathrm{T}}\bar{\lambda})^+$ (with derivative in X) for some $\bar{\lambda}$, optimal for* (DCA_p), *some $1 < p < \infty$, then the unique optimal solution of* (CA_p) *is* $\bar{x} = \|(A^{\mathrm{T}}\bar{\lambda})^+\|^{q-1}\nabla\|(A^{\mathrm{T}}\bar{\lambda})^+\|$.

Proof. The uniqueness follows by a standard argument and the duality result is a direct application of Corollary 4.8.

Notice that if $-b \in P$ then $\bar{x} = 0$ is the unique optimal solution of (CA_p). Suppose on the other hand that $-b \notin P$. Since $P = P^{++}$ it follows that there exists $\hat{\lambda} \in P^+$ with $b^T\hat{\lambda} > 0$. Thus the value of (DCA_p) is strictly positive, since $\delta\hat{\lambda} \in P^+$ with $b^{\mathrm{T}}(\delta\hat{\lambda}) - (1/q)\|(A^{\mathrm{T}}(\delta\hat{\lambda}))^+\|_*^q > 0$, for $\delta > 0$ sufficiently small.

Suppose now that $\bar{\lambda}$ is optimal for (DCA_p). If $A^{\mathrm{T}}\bar{\lambda} \leq 0$, then $b^{\mathrm{T}}\bar{\lambda} > 0$, and we obtain a contradiction since $k\bar{\lambda}$ is feasible for arbitrarily large $k > 0$, with arbitrarily large objective value. Thus $A^{\mathrm{T}}\bar{\lambda} \not\leq 0$, and we can apply Proposition 5.6 to obtain the result. □

It is clear that the problems (CA_p), $1 \leq p \leq \infty$ are all equivalent: $\bar{x} \in X$ is optimal for one problem if and only if it is optimal for all the problems. By contrast, the relationship between the dual problems (DCA_p), $1 \leq p < \infty$, is not so immediately evident. Our next result illuminates this relationship. First we need some preliminary results.

Definition 7.2 (Rockafellar, 1970). A convex function $g : \mathbb{R}^n \to [0, \infty]$ is called a *gauge* if g is positively homogeneous with $g(0) = 0$.

Definition 7.3 (Rockafellar, 1970). Let g be a gauge. The *polar* of g, g^0, is defined by

$$g^0(y) = \inf\{\mu \geq 0 \mid y^{\mathrm{T}}x \leq \mu g(x)\ \forall x \in \mathbb{R}^n\}.$$

Theorem 7.4. *Suppose g is a closed gauge and $C = \{x \in \mathbb{R}^n \mid g(x) \leq 1\}$. Then $g^0(\cdot) = \delta^*(\cdot \mid C)$.*

Proof. Rockafellar (1970, 15.1.2). □

Theorem 7.5. *Suppose g is a closed gauge and $1 < q < \infty$. Suppose f is defined by $f(x) := (1/q)g(x)^q$. Then $f^*(y) = (1/p)g^0(y)^p$, where $1/p + 1/q = 1$.*

Proof. Rockafellar (1970, 15.3.1). □

Theorem 7.6. *Suppose g is a closed gauge and $1 < q < \infty$. Consider the following problems*:

(PHP_q) $\sup\{b^{\mathrm{T}}\lambda - (1/q)g(\lambda)^q \mid \lambda \in \mathbb{R}^n\}$,

(PHP_∞) $\sup\{b^{\mathrm{T}}\lambda \mid g(\lambda) \leq 1, \lambda \in \mathbb{R}^n\}$.

These problems are equivalent in the following sense. The value of (PHP_q) *is* $(1/p)g^0(b)^p$, *and the set of optimal solutions is* $g^0(b)^{p-1}\partial(g^0)(b)$. *On the other hand, the value of* (PHP_∞) *is* $g^0(b)$ *and the set of optimal solutions is* $\partial(g^0)(b)$.

Proof. $V(\mathrm{PHP}_q)=((1/q)g^q)^*(b)$, by definition, and this is $(1/p)g^0(b)^p$ by Theorem 7.6. Now λ attains the value if and only if $\lambda\in\partial((1/p)(g^0)^p)(b)$ by Rockafellar (1970, 23.5), so by the chain rule (Clarke, 1983, 2.3.10), the solution set is $g^0(b)^{p-1}\partial(g^0)(b)$.

On the other hand, $V(\mathrm{PHP}_\infty)=\delta^*(b\,|\,C)=g^0(b)$, where $C=\{\lambda\in\mathbb{R}^n\,|\,g(\lambda)\le 1\}$, by Theorem 7.4. Thus λ is optimal for (PHP_∞) if and only if $\lambda\in\partial(\delta^*(\cdot\,|\,C))(b)=\partial(g^0)(b)$, as required. □

We can now apply this result to (DCA_p) by setting

$$g(\lambda):=\|(A^{\mathrm{T}}\lambda)^+\|_*+\delta(\lambda\,|\,P^+).$$

We now turn to examples of the application of Theorem 7.1. The case we shall consider is that of constrained L^p approximation (cf. Micchelli, Smith, Swetits and Ward, 1985).

Constrained L^p approximation

Example 7.7. (T,μ) a σ-finite measure space, $1<p<\infty$.

$$\begin{array}{lll}(L_p\mathrm{A}) & \inf & (1/p)\|x\|_p^p\\ & \text{subject to} & Ax\in b+P,\\ & & x\ge 0,\quad x\in L^p(T,\mu).\end{array}$$

Here, $A:L^p\to\mathbb{R}^n$ is defined by $(Ax)_i=\int_T a_ix\,\mathrm{d}\mu$, $i=1,\ldots,n$, for some a_i's $\in L^q$ $(1/q+1/q=1)$. The dual problem becomes

$$\begin{array}{lll}(\mathrm{D}L_p\mathrm{A}) & \max & b^{\mathrm{T}}\lambda-\dfrac{1}{q}\left\|\left(\displaystyle\sum_{i=1}^n\lambda_ia_i\right)^+\right\|_q^q\\ & \text{subject to} & \lambda\in P^+,\quad \lambda\in\mathbb{R}^n.\end{array}$$

Applying Theorem 7.1, Examples 3.11(i) and Examples 5.7(i), we obtain that if there exists a feasible $\hat{x}$ for (L_pA) with $\hat{x}(t)>0$ μ-a.e., then the values of (L_pA) and ($\mathrm{D}L_p$A) are equal, with attainment in ($\mathrm{D}L_p$A). If $-b\in P$, then $\bar{x}=0$ is optimal for (L_pA). If $-b\notin P$ then the unique optimal solution of (L_pA) is given by

$$\bar{x}=\left\|\left(\sum_i\bar{\lambda}_ia_i\right)^+\right\|_q^{q-1}\nabla\left\|\left(\sum_i\bar{\lambda}_ia_i\right)^+\right\|_q=\left(\left(\sum_i\bar{\lambda}_ia_i\right)^+\right)^{q-1}. \tag{7.2}$$

Definition 7.8. Suppose a set of functions $a_i:[\alpha,\beta]\to\mathbb{R}$, $i=1,\ldots,n$, are continuous and linearly independent on every non-null subset of $[\alpha,\beta]$. Then we say the a_i's are *pseudo-Haar* on $[\alpha,\beta]$.

As an example, if the a_i's are analytic and linearly independent on $[\alpha, \beta]$, then it is easy to see that they are pseudo-Haar (see Borwein and Lewis, 1991).

Now consider the case where $T=[\alpha, \beta]$ and μ is Lebesgue measure, and the a_i's are pseudo-Haar. Assuming $-b \notin P$, if λ and θ are any two optimal solutions of $(\mathrm{D}L_p\mathrm{A})$, then from (7.2) we must have $(\sum_i \lambda_i a_i)^+ = (\sum_i \theta_i a_i)^+$, a.e., and so by continuity there will exist a non-null subset $\Omega \subset [\alpha, \beta]$ such that $\sum_i (\lambda_i - \theta_i) a_i = 0$ on Ω. The pseudo-Haar condition now implies that $\lambda = \theta$. Thus the optimal dual solution is unique.

The case $p=2$, and $P=\{0\}$ (equality-constrained L^2 approximation, see Borwein and Wolkowicz, 1986) is particularly simple. The primal problem is then just

$$
\begin{array}{lll}
(\mathrm{E}L_2\mathrm{A}) & \inf & \frac{1}{2}\|x\|_2^2 \\
 & \text{subject to} & \langle x, a_i \rangle = b_i, \quad i=1, \ldots, n, \\
 & & x \geqslant 0, \quad x \in L^2,
\end{array}
$$

and the dual problem is $\max\{b^{\mathrm{T}}\lambda - \frac{1}{2}\|(\sum_i \lambda_i a_i)^+\|_2^2 \mid \lambda \in \mathbb{R}^n\}$. Solving this problem reduces to solving the equation $A((A^{\mathrm{T}}\lambda)^+) = b$ and the Newton iteration reduces to

$$
\sum_i \left[\int_{\{t \mid (A^{\mathrm{T}}\lambda^{\mathrm{old}})(t) > 0\}} a_i a_j \right] \lambda_i^{\mathrm{new}} = b_j, \quad j=1, \ldots, n.
$$

In particular, if $t=[\alpha, \beta]$, μ = Lebesgue measure, and $a_i(t) = t^{i-1}$, $i=1, \ldots, n$, and if we denote the Hessian Matrix

$$
(H(\lambda))_{ij} = \int_{\{t \mid \sum_k \lambda_k t^{k-1} > 0\}} t^{i+j-2},
$$

then the Newton step is simply $H(\lambda^{\mathrm{old}})\lambda^{\mathrm{new}} = b$. The following observations are easy to check.

(i) No numerical integration is necessary (only root-finding).
(ii) $H(\lambda)$ takes only $\mathrm{O}(n)$ steps to evaluate.
(iii) $H(\lambda)$ is positive definite if $\sum_k \lambda_k t^{k-1} \not\leqslant 0$.
(iv) $H(\lambda)$ is locally Lipschitz if $\sum_k \lambda_k t^{k-1}$ has no repeated roots.

The Newton method is therefore superlinear and generally quadratic, locally.

Let us return to the question of when the constraint qualification will be satisfied. For simplicity, suppose $P=\{0\}$, and let us also suppose that the a_i's are linearly independent on T. it follows that $\mathrm{ri}(A(L_+^p)) = \mathrm{int}(A(L_+^p))$, since otherwise there is a $y \neq 0$ with $\sum_i y_i \int_T a_i x \, \mathrm{d}\mu = 0$, for all $x \geqslant 0$, which implies $\sum_i y_i a_i = 0$ a.e. on T. But by Proposition 2.10, $\mathrm{ri}(A(L_+^p)) = A\,\mathrm{qri}(L_+^p)$. We thus see that the constraint qualification is satisfied if and only if $b \in \mathrm{int}(A(L_+^p))$. The cone $A(L_+^p)$ is called the 'moment cone'.

In certain cases we can be more explicit.

Theorem 7.9. *Suppose $T=[0,1]$, μ is Lebesgue measure, $b \in \mathbb{R}^{n'+1}$, $P=\{0\}$, and $a_i(t) = t^i$, $i=0, \ldots, n'$.*

(a) *Suppose $n' = 2m$. Consider the two quadratic forms*

$$\sum_{i,j=0}^{m} b_{i+j} y_i y_j, \qquad \sum_{i,j=0}^{m-1} (b_{i+j+1} - b_{i+j+2}) y_i y_j.$$

Then $(L_p\mathrm{A})$ is consistent if and only if these forms are positive definite, and in this case the constraint qualification is satisfied.

(b) *Suppose $n' = 2m+1$. Consider the two quadratic forms*

$$\sum_{i,j=0}^{m} b_{i+j+1} y_i y_j, \qquad \sum_{i,j=0}^{m} (b_{i+j} - b_{i+j+1}) y_i y_j.$$

Then $(L_p\mathrm{A})$ is consistent if and only if these forms are positive definite, and in this case the constraint qualification is satisfied only if they are positive definite.

In both cases there is a non-negative measure satisfying the constraints of $(L_p\mathrm{A})$ if and only if the relevant forms are positive semi-definite.

Proof. Karlin and Studden (1966, p. 106). □

For example, for $[\alpha, \beta] = [0, 1]$, $n' = 2$ and $b_0 = 1$ we obtain $1 > b_1 > b_2 > b_1^2$, and by a similar technique, for $[\alpha, \beta] = [-1, 1]$ and $n' = 2$, the constraint qualification becomes $b_0 > b_2 > 0$ and $b_0 b_2 > b_1^2$.

The following trigonometric case, which occurs in the context of spectral estimation (Ben-Tal, Borwein and Teboulle, 1988), may also be treated explicitly to find the form of the constraint qualification.

$$\text{(SEP)} \quad \inf \quad (1/p)\|x\|_p^p$$
$$\text{subject to} \quad \int_{-\pi}^{\pi} x(t)\cos(jt)\,\mathrm{d}t = b_j, \quad j = 0, \ldots, m,$$
$$\int_{-\pi}^{\pi} x(t)\sin(jt)\,\mathrm{d}t = c_j, \quad j = 1, \ldots, m,$$
$$x \geq 0, \quad x \in L^p[-\pi, \pi].$$

Theorem 7.10. *Set $c_0 := 0$ and $r_j := b_j + c_j\sqrt{-1}$, $r_{-j} := \bar{r}_j$, $j = 0, \ldots, m$. Then (SEP) is consistent if and only if the Toeplitz matrix $(r_{j-k})_{j,k=0}^{m}$ is positive definite, and in this case there is a feasible $\hat{x}(t) > 0$ a.e. on $[-\pi, \pi]$.*

Proof. Ben-Tal, Borwein and Teboulle (1988). See also Karlin and Studden (1966, p. 184). □

Constrained L^∞ approximation

Let us now consider the problem (CA_p) in the case where $(X, Y) = (L^\infty(T, \mu), L^1(T, \mu))$ for a σ-finite measure space (T, μ). The primal problem is thus (for

$1<p<\infty$)

$$\begin{array}{lll}(L_\infty\mathrm{A}) & \inf & (1/p)\|x\|_\infty^p \\ & \text{subject to} & \left(\int_T a_i\, \mathrm{d}\mu - b_i\right)_1^n \in P, \\ & & x \geqslant 0, \quad x \in L^\infty(T,\mu),\end{array}$$

where $a_i \in L^1(T,\mu)$, $i=1,\ldots,n$, and the dual is

$$\begin{array}{lll}(\mathrm{D}L_\infty\mathrm{A}) & \max & b^{\mathrm{T}}\lambda - \frac{1}{q}\left\|\left(\sum_{i=1}^n \lambda_i a_i\right)^+\right\|_1^q \\ & \text{subject to} & \lambda \in P^+, \quad \lambda \in \mathbb{R}^n.\end{array}$$

Applying Theorem 7.1, Examples 3.11(ii) and Examples 5.7(ii), we obtain that if there exists a feasible $\hat{x}(t) > 0$ μ-a.e. for $(L_\infty\mathrm{A})$ then the values of $(L_\infty\mathrm{A})$ and $(\mathrm{D}L_\infty\mathrm{A})$ are equal, with attainment in $(\mathrm{D}L_\infty\mathrm{A})$. Furthermore, if $\bar{\lambda}$ is optimal for $(\mathrm{D}L_\infty\mathrm{A})$ and $|\sum_i \bar{\lambda}_i a_i| > 0$ μ-a.e. then the unique optimal solution of $(L_\infty\mathrm{A})$ is given by

$$\bar{x} = \left\|\left(\sum_i \bar{\lambda}_i a_i\right)^+\right\|_1^{q-1} \chi_{\{t \mid \sum_i \bar{\lambda}_i a_i(t) > 0\}}.$$

The case of $(L_\infty\mathrm{A})$ with $P=\{0\}$ (i.e. equality constrained) and without the positivity constraint $x \geqslant 0$ was considered in Favard (1940), using a duality argument. The dual problem in this case becomes simply

$$\max\left\{ b^{\mathrm{T}}\lambda - \frac{1}{q}\left\|\sum_i \lambda_i a_i\right\|_1^q \,\middle|\, \lambda \in \mathbb{R}^n \right\},$$

or, in the case $p=1$,

$$\max\left\{ b^{\mathrm{T}}\lambda \,\middle|\, \left\|\sum_i \lambda_i a_i\right\|_1 \leqslant 1,\ \lambda \in \mathbb{R}^n \right\}.$$

Favard also considers the case when we do not have $|\sum_i \lambda_i a_i| > 0$ at the optimum, and show how to construct solutions in this case. A straight-forward adaptation of his ideas applies to our case too (see also De Boor, 1976). We defer discussion of this and other numerical questions to a later paper.

Constrained interpolation

We have already seen two interesting concrete examples of the constrained approximation problem (CA_p) in Theorems 7.9 and 7.10, where the function a_i were either t^{i-1} or $\mathrm{e}^{i\pi\sqrt{-1}}$ respectively. A third interesting example arises from problems of constrained interpolation. Typically we might be interested in interpolating some given set of data points with a convex function of minimum norm in some Sobolev space (see Irvine, Marin and Smith, 1986). In this case the a_i's become normalized B-splines.

Definition 7.11. The *Sobolev space* $L_p^k[0, 1]$ is defined as the set of $f \in C^{(k-1)}[0, 1]$ with $f^{(k-1)}$ absolutely continuous and $f^{(k)} \in L_p[0, 1]$ (for $1 \leqslant p \leqslant \infty$ and $k \in \mathbb{N}$).

Let us suppose that $0 \leqslant t_1 < t_2 < \cdots < t_n \leqslant 1$.

Definition 7.12 (Schumaker, 1981, p. 45). For any function $f:[0, 1] \to \mathbb{R}$ we define the kth *divided difference* of f by, for $i = 1, \ldots, n-k$,

$$[t_i, \ldots, t_{i+k}]f := \sum_{p=i}^{i+k} \left\{ \prod_{\substack{q=i \\ q \neq p}}^{i+k} (t_p - t_q) \right\}^{-1} f(t_p).$$

Definition 7.13 (De Boor, 1976, p. 29). The *normalized B-spline* is defined by

$$M_{i,k}(t) = k[t_i, \ldots, t_{i+k}]((\cdot - t)^{k-1})^+, \quad i = 1, \ldots, (n-k).$$

Examples 7.14.

(i) $\quad M_{i,1}(t) = (t_{i+1} - t_i)^{-1} \chi_{[t_i, t_{i+1}]}(t).$

(ii) $$M_{i,2}(t) = \begin{cases} 0, & t \leqslant t_i \text{ and } t \geqslant t_{i+2}, \\ 2(t_{i+2} - t_i)^{-1}, & t = t_{i+1}, \\ \text{linear}, & \text{on } [t_i, t_{i+1}] \text{ and } [t_{i+1}, t_{i+2}). \end{cases}$$

Since all the Sobolev norms are equivalent we restrict attention to the problem $(1 \leqslant p \leqslant \infty)$:

$$(\mathrm{Cl}_k) \quad \begin{array}{ll} \inf & \|f^{(k)}\|_p \\ \text{subject to} & f(t_i) = b_i, \quad i = 1, \ldots, n, \\ & f^{(k)} \geqslant 0, \quad f \in L_p^k[0, 1]. \end{array}$$

For $i = 1, \ldots, (n-k)$, define

$$d_i := \sum_{p=i}^{i+k} \left\{ \prod_{\substack{q=i \\ q \neq p}}^{i+k} (t_p - t_q) \right\}^{-1} b_p,$$

so $d_i = [t_i, \ldots, t_{i+k}]f$, for any feasible f. Consider the following problem (which is of the form (CA_1)).

$$(\mathrm{Cl}_k') \quad \begin{array}{ll} \inf & \|g\|_p \\ \text{subject to} & \displaystyle\int_0^1 M_{i,k}(t) g(t)\, \mathrm{d}t = d_i, \quad i = 1, \ldots, (n-k), \\ & g \geqslant 0, \quad g \in L_p[0, 1]. \end{array}$$

We then have the following result showing that the problems (Cl_k) and (Cl_k') are essentially equivalent.

Theorem 7.15. *If f is feasible for* (Cl_k) *then* $f^{(k)}$ *is feasible for* (Cl'_k), *with the same value. On the other hand, if g is feasible for* (Cl'_k) *then there exists feasible f for* (Cl_k) *with the same value and with* $f^{(k)} = g$.

Proof. See De Boor (1976) and Micchelli, Smith, Swetits and Ward (1985). □

Examples 7.16.

(i) The monotone case: $k=1$.

In this case $d_i = b_{i+1} - b_i$, $i = 1, \ldots, (n-1)$. (Cl′) is consistent if $d_i \geq 0$, each i, and the constraint qualification is satisfied if $d_i > 0$, each i. It is an easy and pleasant exercise to check from both the primal and dual problems that the optimal f for (Cl_1) is simply the piecewise linear interpolant through the data points.

(ii) The convex case: $k=2$. In this case

$$d_i = \frac{b_i}{(t_i - t_{i+1})(t_i - t_{i+2})} + \frac{b_{i+1}}{(t_{i+1} - t_i)(t_{i+1} - t_{i+2})} + \frac{b_{i+2}}{(t_{i+2} - t_i)(t_{i+2} - t_{i+1})}, \tag{7.3}$$

each $i = 1, \ldots, n-2$.

The relevant moment cone in (Cl'_2) is simply the positive orthant:

Lemma 7.17.

$$\left\{ \left(\int_0^1 M_{i,2} g \right) \middle| \, g(t) > 0, \text{ a.e., } g \in L_p[0,1] \right\} = \{ y \in \mathbb{R}^{n-2} \,|\, y_i > 0, \text{ each } i \}.$$

Proof. Denote the left-hand side by C. Clearly $C \subset \operatorname{int} \mathbb{R}_+^{n-2}$. We claim $\operatorname{cl} C = \mathbb{R}_+^{n-2}$. Suppose not, so there exists $0 \leq y \notin \operatorname{cl} C$. Therefore by separation there exists $\lambda \in \mathbb{R}^{n-2}$ for which $\lambda^{\mathrm{T}} y < 0 \leq \sum_{i=1}^{n-2} \lambda_i \int_0^1 M_{i,2} g$, for all $g(t) > 0$, a.e. Then $\sum_{i=1}^{n-2} \lambda_i M_{i,2} \geq 0$. However, by Examples 7.14(ii), $M_{i,2}(t_j) = 0$, $j \neq i+1$ and >0 for $j = i+1$, so $\lambda_i \geq 0$, each $i = 1, \ldots, n-2$. But this contradicts $\lambda^{\mathrm{T}} y < 0$. Thus $\operatorname{cl} C = \mathbb{R}_+^{n-2}$. It follows that $C \supset \operatorname{int} C = \operatorname{int} \mathbb{R}_+^{n-2}$ (Rockafellar, 1970, 6.3.1), and the result follows. □

This result shows that the constraint qualification for (Cl'_2) is satisfied if and only if each $d_i > 0$. This has a simple geometric interpretation, which may be checked from formula (7.3). Consider the lines L_i joining the data points (t_i, b_i) and (t_{i+2}, b_{i+2}). Then the constraint qualification for (Cl'_2) is satisfied if and only if L_i passes strictly above (t_{i+1}, b_{i+1}), for each i.

If, as is natural, we choose $p=2$ and use the objective function $\frac{1}{2}\|g\|_2^2$ in (Cl'_2), then we obtain a problem of the form $(\mathrm{E}L_2\mathrm{A})$. As we have already seen, Newton's method is well-suited to solving the corresponding dual problem, and in this case the Hessian will be tridiagonal and easy to compute exactly. It is worth noting that the proof of Lemma 7.17 breaks down for $k \geq 3$.

8. The bounded linear case

In the previous two sections we were concerned with the convex model (CM) of Section 4 in the case where the underlying constraint set was a cone. In the remaining two sections we shall concentrate on our other main example, the case when the underlying constraint set has the form (cf. (3.1))

$$F=\left\{(x_1,\ldots,x_m)\in X^m \;\middle|\; \sum_{i=1}^m x_i=e,\ x_i\geqslant 0,\ i=1,\ldots,m\right\}. \tag{8.1}$$

As we shall see, these sets arise frequently in connection with transportation-type problems. In an analogous fashion to the development of Section 6, we shall begin by identifying circumstances under which the function $(f+\delta(\cdot\,|\,C))^*$ appearing in the dual problem (DCM) is easy to evaluate. We shall primarily be concerned with the case where f is linear. In what follows, $\tau(\cdot,\cdot)$ denotes the Mackey topology (Schaefer, 1971).

Theorem 8.1. *Suppose X and Y are vector spaces partially ordered by convex cones S_X and S_Y respectively, and with $\langle\cdot,\cdot\rangle : X\times Y\to\mathbb{R}$ a bilinear form. Suppose $e\in S_X$ and $y_1,\ldots,y_m\in Y$, and consider the problems*

(PR) inf $\langle e, y\rangle$

subject to $y\geqslant y_i,\quad i=1,\ldots,m,$

$y\in Y,$

(DPR) sup $\sum_{i=1}^m \langle x_i, y_i\rangle$

subject to $\sum_{i=1}^m x_i=e,$

$x_i\geqslant 0,\ x_i\in X,\quad i=1,\ldots,m.$

If any of the following three conditions are met then the values of (PR) *and* (DPR) *are equal.* (*In each case $\langle\cdot,\cdot\rangle$ is simply evaluation.*)

(i) *(X, S_X) a vector lattice, $S_Y=S_X^+$, (Y,S_Y) a sublattice of X^b.*

(ii) *Y complete metrizable, S_Y closed, generating, $X=Y^*$ and $S_X=S_Y^+$.*

(iii) *(X,Y) a dual pair, S_Y generating, $S_X=S_Y^+$ and*

either (a) *$\tau(Y,X)-\operatorname{int} S_Y\neq\emptyset$,*

or (b) *S_Y is $\sigma(Y,X)$-closed and $e\in\tau(X,Y)-\operatorname{int} S_X$.*

Proof. (i) $y\geqslant y_i$, $i=1,\ldots,m$, if and only if $y\geqslant\bigvee_i y_i$. It follows that $\bar y=\bigvee_i y_i$ is optimal for (PR), with value $(\bigvee_i y_i)(e)$. But the value of (DPR) is also $(\bigvee_i y_i)(e)$, by Proposition 3.1.

(ii) Define a function $f: Y \to Y^m$ by $f(y) := (y_1 - y, \ldots, y_m - y)$. Then f is certainly S_Y^m-convex. Furthermore, given any $z_1, \ldots, z_m \in Y$, there exist $u_i, v_i \in S_Y$, $i = 1, \ldots, m$ with $y_i - z_i = u_i - v_i$, since S_Y is generating. It follows that $y_i - \sum_j u_j \leqslant_{S_Y} z_i$, each i, so $(z_1, \ldots, z_m) \in f(\sum_j u_j) + S_Y^m$. Since $(z_1, \ldots, z_m)$ was arbitrary, $f(Y) + S_Y^m = Y^m$, so certainly $0 \in \text{core}\,(f(Y) + S_Y^m)$.

Now define a convex relation $H: Y \to Y^m$ by $H(y) := (y_1 - y + S_Y, \ldots, y_m - y + S_Y)$. The same argument shows H is surjective, so certainly $H^{-1}(0, \ldots, 0) \cap \text{core}(\text{dom}(e)) \neq \emptyset$. Consider the problem (PR), which we can write

$$\mu := \inf\{e(y) \mid f(y) \leqslant 0, y \in Y\}.$$

Since $e \geqslant 0$, $\mu > -\infty$, so we apply Borwein (1987, 2.7) to deduce the existence of $\bar{x}_1, \ldots, \bar{x}_m \geqslant 0$ such that

$$\mu = \inf\left\{ e(y) + \sum_i \bar{x}_i(y_i - y) \,\middle|\, y \in Y \right\}.$$

It follows that

$$\sum_i \bar{x}_i = e \quad \text{and} \quad \mu = \sum_i \bar{x}_i(y_i).$$

But certainly for any feasible y for (PR) and feasible $(x_1, \ldots, x_m)$ for (DPR) we have $\sum_i x_i(y_i) \leqslant \sum_i x_i(y) = e(y)$, so $V(\text{DPR}) \leqslant V(\text{PR})$. The result follows.

(iii) (a) We can apply Theorem 3.13 in Anderson and Nash (1987). The problem (DPR) is the dual of (PR) in the linear programming sense. Since S_Y is generating, there exist u_i, $v_i \in S_Y$ with $u_i - v_i = y_i$, each i. Thus $\sum_{i=1}^m u_i$ is feasible for (PR). Furthermore, since $e \geqslant 0$, the value of (PR) is finite.

If $\hat{y} \in \tau(Y, X) - \text{int}\, S_Y$, then

$$\sum_{i=1}^m u_i + \hat{y} - y_j = \left(\sum_{i \neq j} u_i \right) + v_j + \hat{y} \in \tau(Y, X) - \text{int}\, S_Y,$$

so the Slater condition is satisfied, and the result follows.

(b) We can rewrite (DPR) in the following way:

$$(\text{DPR}') \quad \sup \quad \sum_{i=1}^{m-1} \langle x_i, y_i - y_m \rangle + \langle e, y_m \rangle$$

$$\text{subject to} \quad \sum_{i=1}^{m-1} x_i \leqslant e,$$

$$-x_i \leqslant 0, \ x_i \in X, \quad i = 1, \ldots, (m-1).$$

We shall apply Theorem 3.13 in Anderson and Nash (1987) to this problem. Since S_Y is closed, the dual of (DPR′) in the linear programming sense is

$$(\text{PR}') \quad \inf \quad \langle e, y \rangle + \langle e, y_m \rangle$$

$$\text{subject to} \quad y - z_i = y_i - y_m, \quad i = 1, \ldots, m-1,$$

$$0 \leqslant y, \quad z_1, \ldots, z_{m-1} \in Y,$$

which is clearly equivalent to (PR). Now as in part (a), $\sum_{i=1}^{m} u_i \geq y_j$, each j, so for any feasible x_i's for (DPR′),

$$\sum_{i=1}^{m-1} \langle x_i, y_i - y_m\rangle + \langle e, y_m\rangle \leq \sum_{i=1}^{m-1} \left\langle x_i, \sum_{j=1}^{m} u_j - y_m \right\rangle + \langle e, y_m\rangle \leq \left\langle e, \sum_{j=1}^{m} u_j \right\rangle,$$

so the value of (DPR′) is less than $+\infty$. Furthermore, by assumption, $\hat{x}_i = (1/m)e$, each i, satisfies the Slater condition, so the result follows. □

Definition 8.2. For X, Y, S_X, S_Y and $\langle\cdot,\cdot\rangle$ as in Theorem 8.1, we say $((X, S_X), (Y, S_Y), \langle\cdot,\cdot\rangle)$ (or more simply just (X, Y)) is a *pseudo-Riesz pair* if the values of (PR) and (DPR) are equal for all $e \in S_X$ and $y_1, \ldots, y_m \in Y$.

Examples 8.3. In each of the following cases Theorem 8.1 shows that (X, Y) is a pseudo-Riesz pair (with $\langle\cdot,\cdot\rangle$ evaluation):

(i) X a normed lattice, $Y = X^*$, $S_X = X_+$, $X_Y = (X^*)_+$ (by Proposition 3.2).

(ii) Y a normed lattice, $X = Y^*$, with the lattice cones (cf. Examples 3.7).

(iii) Y a Banach space, S_Y closed, generating, $X = Y^*$, $S_X = S_Y^+$. More generally than (i) and (ii), (X, Y) a dual lattice pair (Definition 3.6).

We shall now work in the following setting:

$$\begin{aligned}
&(X, Y) \text{ a dual pair,}\\
&S_X \subset X,\ S_Y \subset Y, \quad \text{convex cones partially ordering } X \text{ and } Y,\\
&y_1, \ldots, y_m \in Y, \quad b \in \mathbb{R}^n, \quad e \in S_X, \qquad\qquad (8.2)\\
&P \subset \mathbb{R}^n \text{ a polyhedral cone,}\\
&A_i : X \to \mathbb{R}^n, \quad \sigma(X, Y)\text{-}\mathbb{R}^n \text{ continuous and linear,} \quad i = 1, \ldots, m.
\end{aligned}$$

We consider the following primal problem:

$$\begin{aligned}
\text{(BLP)} \quad \inf \quad & \sum_{i=1}^{m} \langle x_i, y_i\rangle\\
\text{subject to} \quad & \sum_{i=1}^{m} A_i x_i \in b + P,\\
& \sum_{i=1}^{m} x_i = e,\\
& 0 \leq x_i \in X, \quad i = 1, \ldots, m.
\end{aligned}$$

The dual problem becomes:

$$\begin{aligned}
(\text{DBLP}_1) \sup \quad & b^{\mathrm{T}}\lambda - \langle e, y\rangle\\
\text{subject to} \quad & A_i^{\mathrm{T}}\lambda - y \leq y_i, \quad i = 1, \ldots, m,\\
& \lambda \in P^+, \quad y \in Y.
\end{aligned}$$

Theorem 8.4. *With the notation of* (8.2), *suppose that* $((X, S_X), (Y, S_Y), \langle\cdot,\cdot\rangle)$ *is a pseudo-Riesz pair. Suppose further that for some* $(\hat{x}_1, \ldots, \hat{x}_m)$, *feasible for* (BLP), *we have*:

$$\sigma(X, Y) - \mathrm{cl}\,\mathbb{P}[0, \hat{x}_i] = \sigma(X, Y) - \mathrm{cl}\,\mathbb{P}[0, e], \quad \textit{each } i. \tag{8.3}$$

Then the values of (BLP) *and* (DBLP_1) *are equal. If* $[0, e]$ *is* $\sigma(X, Y)$*-compact and* (BLP) *is consistent then its value is attained at an extreme point of the feasible region. If* (8.3) *holds and* $(\bar{x}_1, \ldots, \bar{x}_m)$ *and* $(\bar{\lambda}; \bar{y})$ *are feasible for* (BLP) *and* (DBLP_1) *respectively, then they are optimal if and only if the following complementary slackness conditions hold*:

$$\begin{aligned} &\bar{\lambda}^{\mathrm{T}}\Big(\sum_i A_i\bar{x}_i - b\Big) = 0, \\ &\langle \bar{x}_i, y_i + \bar{y} - A_i^{\mathrm{T}}\bar{\lambda}\rangle = 0, \quad \textit{each } i. \end{aligned} \tag{8.4}$$

Proof. Apply Corollary 4.8 with underlying constraint set F as in (8.1). The constraint qualification becomes (8.3) by Theorem 3.12. In this case, $f(x_1, \ldots, x_m) = \sum_i \langle x_i, y_i\rangle$, so

$$\begin{aligned} (f + \delta(\cdot\,|\,F))^*(\phi_1, \ldots, \phi_m) &= \sup\Big\{\sum_i \langle x_i, \phi_i - y_i\rangle \,\Big|\, (x_1, \ldots, x_m) \in F\Big\} \\ &= \inf\{\langle e, y\rangle \,|\, y \geq \phi_i - y_i \ \forall i\}, \end{aligned}$$

by Definition 8.2. The duality result now follows.

If (BLP) is consistent, and $[0, e]$ is $\sigma(X, Y)$-compact, then the feasible region of (BLP) is clearly nonempty and $\sigma(X^m, Y^m)$-compact, so its value is attained at an extreme point by Holmes (1975, p. 74).

Finally, if $\bar{x}$ and $(\bar{\lambda}; \bar{y})$ are primal and dual feasible respectively, with equal value, then we have

$$\begin{aligned} b^{\mathrm{T}}\bar{\lambda} - \langle e, \bar{y}\rangle &= b^{\mathrm{T}}\bar{\lambda} - \Big\langle \sum_i \bar{x}_i, \bar{y}\Big\rangle \\ &\leq b^{\mathrm{T}}\bar{\lambda} - \sum_i \langle \bar{x}_i, A_i^{\mathrm{T}}\bar{\lambda} - y_i\rangle \\ &= \sum_i \langle \bar{x}_i, y_i\rangle + \bar{\lambda}^{\mathrm{T}}\Big(b - \sum_i A_i\bar{x}_i\Big) \\ &\leq \sum_i \langle \bar{x}_i, y_i\rangle = b^{\mathrm{T}}\bar{\lambda} - \langle e, \bar{y}\rangle. \end{aligned}$$

Thus equality holds throughout, and (8.4) follows. □

Suppose now that (X, Y) is a dual lattice pair (Definition 3.6). If $e \in \mathrm{qri}\, X_+$ then the constraint qualification reduces to finding a feasible $(\hat{x}_1, \ldots, \hat{x}_m)$ with $\hat{x}_i \in \mathrm{qri}\, X_+$

for each $i=1,\ldots,m$, by Corollary 3.14. The dual problem simplifies to the finite-dimensional problem

$$(\mathrm{DBLP}_2)\ \text{maximize} \quad b^{\mathrm{T}}\lambda-\left\langle e, \bigvee_{i=1}^{m}(A_i^{\mathrm{T}}\lambda-y_i)\right\rangle$$

$$\text{subject to} \quad \lambda\in P^{+}, \quad \lambda\in\mathbb{R}^n.$$

Define $g_i:\mathbb{R}^n\to Y$ and $N_i\subset X$, each $i=1,\ldots,m$, by $g_i(\lambda)=A_i^{\mathrm{T}}\lambda-y_i$, and $N_i(\lambda)=N(\bigvee_{k=1}^{m}g_k(\lambda)-g_i(\lambda))$, where $N(\cdot)$ denotes absolute kernel (cf. Section 5). For a specific $\bar\lambda$, set $\bar g_i=\bar g_i(\bar\lambda)$ and $\bar N_i=N_i(\bar\lambda)$, each i. As before, X_e is the principal ideal generated by e.

The following result shows how we can compute the solution of the original problem (BLP) by first solving (DBLP_2).

Theorem 8.5. *With the notation of* (8.2), *suppose* (X, Y) *is a dual lattice pair. If* (8.3) *holds for some* $(\hat x_1,\ldots,\hat x_m)$ *feasible for* (BLP), *then the values of* (BLP) *and* (DBLP_2) *are equal* (*with attainment in* (DBLP_2)).

Suppose further that (X, Y) *is a countably regular lattice pair* (*Definition* 5.18). *For* $\bar\lambda\in\mathbb{R}^n$, *consider the following condition*:

$$\left\langle x, \left|\bigvee_{k\neq i}\bar g_k-\bar g_i\right|\right\rangle>0, \quad \textit{for all } 0\neq x\in[0,e], \ \textit{each } i=1,\ldots,m. \tag{8.5}$$

If (8.5) *holds then* X_e *is an order direct sum*

$$X_e=\bigoplus_{i=1}^{m}(X_e\cap\bar N_i), \tag{8.6}$$

and the dual objective function is differentiable at $\bar\lambda$ *with derivative*

$$b-\sum_{i=1}^{m}A_iP_{X_e\cap\bar N_i}(e).$$

If furthermore $\bar\lambda$ *is optimal for* (DBLP_2) *then the unique optimal solution of* (BLP) *is given by* $\bar x_i=P_{X_e\cap\bar N_i}(e)$, *each* $i=1,\ldots,m$.

Proof. The duality result follows directly from Corollary 4.8. The criterion for the differentiability of the dual objective function follows from Corollary 5.15, since

$$\nabla_\lambda\left\langle e, \bigvee_i g_i(\lambda)\right\rangle=(\nabla_\lambda g(\lambda))^{\mathrm{T}}\left(\nabla_\phi\left\langle e,\bigvee_i\phi_i\right\rangle\right)\Big|_{\phi=g(\lambda)}.$$

The remainder of the result follows again from Corollary 4.8. Since (X, Y) is a countably regular lattice pair, X is an ideal in Y^{b}, so $X_+=(Y_+)^+$ and is thus $\sigma(X, Y)$-closed. It follows that the function $f+\delta(\cdot\,|\,C)$ in Corollary 4.8 is closed, and the result now follows. □

The interpretation of the various expressions in Theorem 8.5 for concrete spaces (X, Y) was discussed at the end of Section 5.

In Theorem 8.4 it was proved that if $[0, e]$ is $\sigma(X, Y)$-compact and (BLP) is consistent then its value is attained at an extreme point of the feasible region. If Y is a normed lattice with $X = Y^*$ then $[0, e]$ is $\sigma(X, Y)$-compact by the Alaoglu-Bourbaki Theorem. On the other hand, if X is a normed lattice with $Y = X^*$ then Examples 5.19 gives various conditions for X to have $\sigma(X, Y)$-compact order intervals. Our next result characterizes the extreme points of the feasible region of (BLP). We first need some definitions and lemmas.

Definition 8.6. Suppose X is an Archimedean vector lattice (i.e. $nx_1 \leqslant x_2$ for all $n \in \mathbb{N}$ implies $x_1 \leqslant 0$). For $0 \leqslant x \in X$, x is an *atom* if the principal ideal X_x is one-dimensional (Schaefer, 1974, p. 143).

In what follows we shall assume X is Archimedean. Notice in particular that any normed lattice is Archimedean.

Lemma 8.7. *$0 \leqslant x \in X$ is an atom if and only if X_x is a minimal ideal (i.e. $\{0\}$ is the only ideal properly contained in X_x).*

Proof. Schaefer (1974, p. 143). $\square$

Lemma 8.8. *If X has no atoms then every nonzero ideal of X is infinite-dimensional.*

Proof. Suppose $\{0\} \neq I \subset X$ is a minimal ideal. Take $0 \leqslant x \in I$, with $x \neq 0$ so $X_x \subset I$. Since I is minimal, $X_x = I$, so X_x is minimal, and thus x is an atom by Lemma 8.7. This is a contradiction, so X has no minimal ideals. Now suppose $\{0\} \neq J_1 \subset X$ is a finite-dimensional ideal. Since J_1 is not minimal, there exists a nonzero ideal $J_2 \subset J_1$ strictly, so $\dim J_2 < \dim J_1$. We can proceed indefinitely in this fashion to obtain a sequence of strict inclusions, $J_1 \supset J_2 \supset J_3 \cdots$, which contradicts $\dim J_1 < \infty$. $\square$

Examples 8.9.

(i) *If T is a normal topological space with no isolated points then $C(T)$ has no atoms.*

(ii) *If μ is a nonatomic measure on T (i.e. for measurable $T_2 \subset T$, with $\mu(T_2) > 0$, there exists measurable $T_1 \subset T_2$ with $\mu(T_1) \neq 0$ or $\mu(T_2)$), then $L^p(T, \mu)$ has no atoms, for $1 \leqslant p \leqslant \infty$.*

(iii) *In particular if $T \subset \mathbb{R}^n$ with μ absolutely continuous with respect to Lebesgue measure, then $L^p(T, \mu)$ has no atoms, for $1 \leqslant p \leqslant \infty$.*

(iv) *If $X = M(T)$ and $0 \leqslant e \in X$ is nonatomic on T, then the principal ideal X_e has no atoms.*

Proof. (i) Suppose $0 \leqslant x \in C(T) = X$ is an atom. Since T has no isolated points, $T_0 = \{t \mid x(t) > 0\}$ is not a singleton, so pick distinct $t_1, t_2 \in T_0$. By Urysohn's lemma, there exists $0 \leqslant y \in C(T)$ with $y(t_1) = 1$ and $y(t_2) = 0$. Now set $z = x \wedge y$. Clearly $\{0\} \neq X_z \subset X_x$ strictly, so x is not an atom, which is a contradiction.

(ii) Suppose $0 \leqslant x \in L^p(T, \mu) = X$ is an atom. Set $T_0 = (t \mid x(t) > 0\}$. T_0 is not an atom in (T, μ), so there exists $T_1 \subset T_0$ with $0 < \mu(T_1) < \mu(T_0)$. Now set $y(t) := x(t)\chi_{T_1}(t)$, for each $t \in T$. Then $\{0\} \neq X_y \subset X_x$ strictly, so x is not an atom, which is a contradiction.

(iii) The fact that any measure which is absolutely continuous with respect to Lebesgue measure is nonatomic follows from the definition of Lebesgue measure.

(iv) By the Radon–Nikodym theorem, X_e is isomorphic with $L^\infty(T, e)$ and the result follows by (ii). □

Definition 8.10. An element $x \in [0, e]$ is a *characteristic element* of $[0, e]$ if $x \wedge (e - x) = 0$.

If X is $C(T)$, $M(T)$ or $L^p(T, \mu)$, $1 \leqslant p \leqslant \infty$, then x is a characteristic element of $[0, e]$ if and only if x is of the form $x = e$ on T_1, and 0 on T_1^c, for some $T_1 \subset T$.

Lemma 8.11. *An element x is an extreme point of $[0, e]$ if and only if it is a characteristic element.*

Proof. Schaefer (1974, p. 65). □

We are now ready to prove the main result.

Theorem 8.12. *With the notation of (8.2), suppose that (X, S_X) is an Archimedean vector lattice, and that the principal ideal X_e has no atoms. Then a feasible $(\bar{x}_1, \ldots, \bar{x}_m)$ for (BLP) is an extreme point of the feasible region if and only if each $\bar{x}_i$ is a characteristic element of $[0, e]$, $i = 1, \ldots, m$.*

Proof. If each $\bar{x}_i$ is a characteristic element of $[0, e]$ then $(\bar{x}_1, \ldots, \bar{x}_m)$ is an extreme point of $[0, e]^m$ (by Lemma 8.11), and therefore of the feasible region.

On the other hand, suppose without loss of generality that $u := \bar{x}_1 \wedge (e - \bar{x}_1) \neq 0$. Certainly $u \geqslant 0$.

Suppose that $\bar{x}_i \wedge u = 0$, each $i = 2, \ldots, m$. Then we would have $u = u \wedge (e - \bar{x}_1) = u \wedge \sum_{i=2}^m \bar{x}_i = 0$, which is a contradiction. Without loss of generality therefore, suppose that $v := \bar{x}_2 \wedge u \neq 0$. We then have $0 \leqslant v \leqslant u \leqslant \bar{x}_1$, $(e - \bar{x}_1)$ and $v \leqslant \bar{x}_2$. Furthermore, $e - \bar{x}_2 \geqslant \bar{x}_1 \geqslant v$. Thus we have constructed a nonzero $v \geqslant 0$ with $\bar{x}_1, \bar{x}_2 \in [v, e - v]$.

Now the principal ideal X_v is infinite-dimensional by Lemma 8.8 so there exists nonzero $w \in X_v$ with $(A_1 - A_2)w = 0$. Since $X_v = \bigcup_{j=1}^\infty j[-v, v]$, we can assume (by scaling if necessary) that $w \in [-v, v]$. It then follows that $(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_m) \pm (w, -w, 0, \ldots, 0)$ are both feasible, so $(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_m)$ is not extreme. □

Notice that if the hypotheses of Theorem 8.5 are satisfied then from (8.6) we know that $X_e = \bigoplus_{i=1}^m (X_e \cap \bar{N}_i)$ is a direct sum of ideals and so the solution $\bar{x}_i = P_{X_e \cap \bar{N}_i}(e)$ satisfies $\bar{x}_i \wedge \bar{x}_j = 0$ for all $i \neq j$. Thus $\bar{x}_i \wedge (e - \bar{x}_i) = 0$, for each i, so this solution is an extreme point.

Notice also that under the conditions of the theorem, if $(\bar{x}_1, \ldots, \bar{x}_m)$ is an extreme point of the feasible region then $0 = \bar{x}_i \wedge (e - \bar{x}_i) = \bar{x}_i \wedge \sum_{j \neq i} \bar{x}_j = \sum_{j \neq i} (x_i \wedge x_j)$, so $x_i \wedge x_j = 0$ for $i \neq j$. If X is $C(T)$, $M(T)$ or $L^p(T, \mu)$, $1 \leq p \leq \infty$ then this says that the supports of x_i and x_j are disjoint. It follows that the extreme points correspond exactly with solutions of the form $\bar{x}_i = e$ on T_i, and 0 on T_i^c, each $i = 1, \ldots, m$, where $T = \bigcup_{i=1}^m T_i$ is a partition of T. Thus by restricting attention to the extreme points of the problem we have reduced it to a set-partitioning problem.

Upper bound constrained semi-infinite linear programming

To conclude this section we shall illustrate Theorem 8.5 by applying it to the problem of semi-infinite linear programming with an upper-bound constraint, and to the problem of best L^1-approximation. Suppose first that (X, Y) is a dual lattice pair, with $c \in Y$, $A: X \to \mathbb{R}^n$ a $\sigma(X, Y)$-$\mathbb{R}^n$ continuous linear mapping, $b \in \mathbb{R}^n$, $P \subset \mathbb{R}^n$ a polyhedral cone and $0 \leq e \in X$. The problem we wish to consider is

$$(\mathrm{ULP}_1) \quad \inf \quad \langle x, c \rangle \quad \text{subject to} \quad Ax \in b + P, \quad 0 \leq x \leq e, \ x \in X.$$

We can write this in the form (BLP) by adding a slack variable:

$$(\mathrm{ULP}_2) \quad \inf \quad \langle x_1, c \rangle + \langle x_2, 0 \rangle \quad \text{subject to} \quad Ax_1 + 0x_2 \in b + P, \quad x_1 + x_2 = e, \quad 0 \leq x_1, x_2 \in X.$$

The dual problem becomes

$$(\mathrm{DULP}) \quad \text{maximize} \quad b^{\mathrm{T}}\lambda - \langle e, (A^{\mathrm{T}}\lambda - c)^+ \rangle \quad \text{subject to} \quad \lambda \in P^+, \ \lambda \in \mathbb{R}^n.$$

Notice again how this dual problem is reminiscent of penalty function approaches to the solution of the dual semi-infinite linear program (DSILP) (see Section 6). The larger the element e becomes, the more we penalize violations of the constraint $A^{\mathrm{T}}\lambda \leq c$.

The constraint qualification requires the existence of a feasible $\hat{x}$ for (ULP_1) with $\mathrm{cl}\, \mathbb{P}[0, \hat{x}] = \mathrm{cl}\, \mathbb{P}[0, e - \hat{x}] = \mathrm{cl}\, \mathbb{P}[0, e]$ (in the $\sigma(X, Y)$ topology), and if $e \in \mathrm{qri}\, X_+$ then this is equivalent to $\hat{x}, (e - \hat{x}) \in \mathrm{qri}\, X_+$. If this holds then we know from Theorem 8.5 that the values of (ULP_1) and (DULP) are equal, with attainment in (DULP).

If $\bar{x}$ and $\bar{\lambda}$ are feasible for (ULP$_1$) and (DULP) respectively in this case, then necessary and sufficient conditions for optimality are the complementary slackness conditions

$$\bar{\lambda}^{\mathrm{T}}(A\bar{x}-b)=0,$$

$$\langle \bar{x}, (A^{\mathrm{T}}\bar{\lambda}-c)^{-}\rangle = 0,$$

$$\langle (e-\bar{x}), (A^{\mathrm{T}}\bar{\lambda}-c)^{+}\rangle = 0,$$

by Theorem 8.4. By Theorem 8.12, if the principal ideal X_e has no atoms then the extreme points of (ULP$_1$) are just the feasible characteristic elements of $[0, e]$.

Finally, suppose that (X, Y) is a countably regular lattice pair. Suppose the constraint qualification holds and that $\bar{\lambda}$ is optimal for (DULP). Suppose further that

$$\langle x, |A^{\mathrm{T}}\bar{\lambda}-c|\rangle > 0, \quad \text{for all } 0 \neq x \in [0, e]. \tag{8.7}$$

It then follows by Theorem 8.5 that if we define $N_+, N_- \subset X$ by $N_{\pm} = N((A^{\mathrm{T}}\bar{\lambda}-c)^{\pm})$ then $X_e = (X_e \cap N_+) \oplus (X_e \cap N_-)$ and the unique optimal solution of (ULP$_1$) is $\bar{x} = P_{X_e \cap N_-}(e)$. We defer a discussion of numerical techniques for the solution of (DULP) to a later paper, except to observe that the objective function is differentiable at any $\bar{\lambda}$ for which (8.7) holds, with derivative $b - AP_{X_e \cap N_-}(e)$. In particular, suppose that $X = L^p([\alpha, \beta], \mu)$, $Y = L^q([\alpha, \beta], \mu)$, $1 \leq p \leq \infty$, with μ Lebesgue measure, and that $A: X \to \mathbb{R}^n$ is defined by $(Ax)_i = \langle x, a_i\rangle$, for some $a_i \in Y$, $i = 1, \ldots, n$. Then if the set $\{a_1, \ldots, a_n, c\}$ is pseudo-Haar on $[\alpha, \beta]$ (Definition 7.8), condition (8.7) will always hold, so the dual objective function will be everywhere differentiable.

Best L^1-approximation

Finally, let us turn to the problem of best L^1-approximation. Suppose that (X, Y) is a dual lattice pair (we shall primarily be concerned with the case $X = L^\infty(T, \mu)$, $Y = L^1(T, \mu)$ with (T, μ) a σ-finite measure space). Suppose that $a_1, \ldots, a_n, c \in Y$ and $e \in X_+$, and consider the problem below, which is in the form of (BLP):

$$(\mathrm{D}L^1\mathrm{P}) \quad \inf \quad \langle x_1, c\rangle + \langle x_2, -c\rangle$$

$$\text{subject to} \quad \langle x_1, a_i\rangle + \langle x_2, -a_i\rangle = 0, \ i = 1, \ldots, n,$$

$$x_1 + x_2 = e,$$

$$x_1, x_2 \geq 0, \ x_1, x_2 \in X.$$

The dual problem is, from Theorem 8.5,

$$(L^1\mathrm{P}) \quad \text{maximize} \quad -\left\langle e, \left| \sum_{i=1}^{n} \lambda_i a_i - c \right| \right\rangle$$

$$\text{subject to} \quad \lambda \in \mathbb{R}^n.$$

When $X = L^\infty$, $Y = L^1$ and $e \equiv 1$ this is exactly the problem of finding the best approximation in the L^1 norm to c from the subspace spanned by $\{a_1, \ldots, a_n\}$, and for more general e we obtain weighted best L^1-approximation.

The constraint qualification for $(\mathrm{D}L^1\mathrm{P})$ is always met by $\hat{x}_1 = \hat{x}_2 = \frac{1}{2}e$, by Corollary 3.13, so by Theorem 8.5 the values of $(L^1\mathrm{P})$ and $(\mathrm{D}L^1\mathrm{P})$ are equal, with attainment in $(L^1\mathrm{P})$ (implying in the (L^∞, L^1) case the existence of a best L^1-approximation). Furthermore, if $[0, e]$ is $\sigma(X, Y)$-compact (as in the case when $(X, Y) = (L^\infty, L^1)$), then the value of $(\mathrm{D}L^1\mathrm{P})$ will also be attained. In this case it follows by complementary slackness (Theorem 8.4) that $\bar{\lambda}$ is optimal for $(L^1\mathrm{P})$ if and only if there exist feasible $\bar{x}_1, \bar{x}_2$ for $(\mathrm{D}L^1\mathrm{P})$ satisfying $\langle \bar{x}_1, (\sum_i \bar{\lambda}_i a_i - c)^- \rangle = 0 = \langle \bar{x}_2, (\sum_i \bar{\lambda}_i a_i - c)^+ \rangle$. This is one version of the characterisation theorem for best L^1-approximation (see for example Singer, 1970). In the case where $(X, Y) = (L^\infty, L^1)$ and $e \equiv 1$, let us denote the set $\{t \in T \mid \sum_i \bar{\lambda}_i a_i(t) > c(t)\}$ by $Z_>$, and similarly for $Z_=$ and $Z_<$. Then if $\mu(Z_=) = 0$ the complementary slackness conditions define $\bar{x}_1, \bar{x}_2$ essentially uniquely to be $\chi_{Z_>}$ and $\chi_{Z_<}$ respectively, so the optimality condition simplifies to $\int_{Z_>} a_i = \int_{Z_<} a_i$, each $i = 1, \ldots, n$ (see Singer, 1970). An alternative approach to this result is to observe that $\mu(Z_=) = 0$ is the condition for the objective function in $(L^1\mathrm{P})$ to be differentiable at $\bar{\lambda}$, and apply Theorem 8.5. Notice in particular that if $T = [\alpha, \beta] \subset \mathbb{R}$, with μ Lebesgue measure and $a_1, \ldots, a_n$, c pseudo-Haar on $[\alpha, \beta]$, then we will always have $\mu(Z_=) = 0$.

If $[0, e]$ is $\sigma(X, Y)$-compact then there exists an extreme point optimal solution of $(\mathrm{D}L^1\mathrm{P})$. If furthermore the principal ideal X_e has no atoms then this extreme point will satisfy $x_1 \wedge x_2 = 0$, by Theorem 8.12. When $(X, Y) = (L^\infty, L^1)$ with (T, μ) nonatomic and $e \equiv 1$ it follows that there is an optimal solution of $(\mathrm{D}L^1\mathrm{P})$ of the form $\bar{x}_1 = \chi_{T_1}$, $\bar{x}_2 = \chi_{T_1^c}$, for some $T_1 \subset T$.

9. Semi-infinite transportation problems

In this final section we shall examine how our previous results can be applied to problems generalizing the classical transportation problem. In these examples the set F defined in (8.1) arises naturally from the constraints: the nonnegative x_i's represent the supply strategies associated with each of m supply points, and the constraint $\sum_{i=1}^m x_i = e$ reflects the requirement that a total demand distribution represented by e has to be supplied from the m supply points. When the distribution strategy is subject to linear transportation costs and the total supply at each of the m supply points is given, the resulting problem is a semi-infinite transportation problem (see Kortanek and Yamasaki, 1982), which is a special case of the bounded linear problem considered in the previous section. When the underlying space X is finite-dimensional the problem reduces to the classical transportation problem. If, more generally, the supply distributions are subject to certain convex production costs, we obtain a 'generalized market area problem' (see Lowe and Hurter, 1976, and Todd 1978). We shall see how the results of the above authors can be rederived in a more general setting using our duality theorems.

Let us first consider the linear case, the semi-infinite transportation problem. We shall adopt the following notation:

$$\begin{aligned}
&(X, Y) \text{ a dual pair,}\\
&S_X \subset X,\ S_Y \subset Y \quad \text{convex cones partially ordering } X \text{ and } Y,\\
&y_1, \ldots, y_n, \quad a_1, \ldots, a_n \in Y,\\
&b \in \mathbb{R}^n, \quad e \in S_X.
\end{aligned} \tag{9.1}$$

The primal semi-infinite transportation problem (cf. Kortanek and Yamasaki, 1982) is then

$$\begin{array}{lll}
\text{(STP)} & \inf & \sum_{i=1}^{n} \langle x_i, y_i \rangle \\
 & \text{subject to} & \langle x_i, a_i \rangle = b_i, \ i = 1, \ldots, n, \\
 & & \sum_{i=1}^{n} x_i = e, \\
 & & x_i \geqslant 0, \ x_i \in X, \quad i = 1, \ldots, n.
\end{array}$$

This is a special case of (BLP) with $m = n$, $P = \{0\}$ and $(A_i x)_j = \langle x_i, a_i \rangle$ for $i = j$, and 0 for $i \neq j$. The dual problem becomes

$$\begin{array}{lll}
(\text{DSTP}_1) & \sup & b^{\mathrm{T}} \lambda - \langle e, y \rangle \\
 & \text{subject to} & \lambda_i a_i - y \leqslant y_i, \ i = 1, \ldots, n, \\
 & & \lambda \in \mathbb{R}^n, \quad y \in Y.
\end{array}$$

This is exactly the primal-dual pair considered in Kortanek and Yamasaki (1982).

Following Theorem 8.4, the constraint qualification requires a feasible $(\hat{x}_1, \ldots, \hat{x}_m)$ for (STP), satisfying (8.3) (or, if (X, Y) is a dual lattice pair with $e \in \operatorname{qri} X_+$, with each $\hat{x}_i \in \operatorname{qri} X_+$). Either of the following assumptions (cf. Kortanek and Yamasaki, 1982), is sufficient to ensure this.

Assumption 9.1. $\langle e, a_i \rangle > 0$, $b_i > 0$, each $i = 1, \ldots, n$, and $\sum_{i=1}^{n} (b_i / \langle e, a_i \rangle) = 1$.

Assumption 9.2. $a_i = a$, $b_i > 0$, each $i = 1, \ldots, n$, and $\sum_{i=1}^{n} b_i = \langle e, a \rangle$.

Clearly Assumption 9.2 implies Assumption 9.1, which in turn implies that the point defined by $\hat{x}_i = (b_i / \langle e, a_i \rangle) e$, each $i = 1, \ldots, m$, is feasible, and this satisfies the constraint qualification by Corollary 3.13. The most usual version of the semi-infinite transportation problem has $X = M(T)$, $Y = C(T)$, with T a compact Hausdorff space, and $a_i \equiv 1$, each $i = 1, \ldots, n$. Assumption 9.2 is then simply the requirement that each supply point has a strictly positive supply and that the total supply is equal to the total demand.

Theorem 9.3. *With the notation of* (9.1), *suppose either of Assumptions* 9.1 *and* 9.2 *holds, and that* $((X, S_X), (Y, S_Y), \langle\cdot,\cdot\rangle)$ *is a pseudo-Riesz pair. Then the values of* (STP) *and* (DSTP_1) *are equal, and if* $(\bar{x}_1, \ldots, \bar{x}_n)$ *and* $(\bar{\lambda}; \bar{y})$ *are feasible for* (STP) *and* (DSTP_1) *respectively then they are optimal if and only if*

$$\bar{x}_i, \bar{y}+y_i-\lambda_i a_i\rangle = 0, \quad \text{each } i = 1, \ldots, n. \tag{9.2}$$

Proof. Theorem 8.4. □

This duality theorem is proved in Kortanek and Yamasaki (1982) in particular for the special case where X is a reflexive Banach lattice with $Y = X^*$, with the lattice orderings (cf. Theorem 4 in the above paper). This case follows from Theorem 9.3 by Examples 8.3. They also consider the case when $\tau\text{-}(Y, X)-\operatorname{int} S_Y \neq \emptyset$ (cf. Theorem 2 in the above paper); since S_Y must then be generating this case also follows from Theorem 9.3 by Theorem 8.1.

Theorem 9.4. *With the notation of* (9.1), *if* $[0, e]$ *is* $\sigma(X, Y)$*-compact and* (STP) *is consistent then there exists an optimal extreme point for* (STP).

Proof. Apply Theorem 8.4. □

Theorem 9.5. *With the notation of* (9.1), *if* (X, S_X) *is an Archimedean vector lattice and the principal ideal* X_e *has no atoms then a feasible* $(\bar{x}_1, \ldots, \bar{x}_n)$ *for* (STP) *is an extreme point of the feasible region if and only if each* $\bar{x}_i$ *is a characteristic element of* $[0, e]$ (*with* $x_i \wedge x_j = 0$ *for* $i \neq j$).

Proof. Apply Theorem 8.12 and the remarks thereafter. □

Following our discussion after Theorem 8.12, we see that when $(X, Y) = (L^p(T, \mu), L^q(T, \mu))$, $1 \leqslant p \leqslant \infty$, with μ nonatomic, we can restrict attention to feasible $(x_1, \ldots, x_n)$ where the x_i's have disjoint support, i.e. of the form $(e\chi_{T_1}, \ldots, e\chi_{T_n})$, where $T = \bigcup_{i=1}^n T_i$ is a partition of T. A similar argument holds for $(X, Y) = (M(T), C(T))$ when e is nonatomic on T, by Examples 8.9(iv). Thus in these cases (STP) reduces to a problem of set-partitioning, reflecting the fact that optimal distribution strategies arise from assigning to each supply point a distinct area in the underlying demand set, for which it has to supply the whole demand. This observation was made in Corley and Roberts (1972). These authors also discuss the relationship between (STP) and the Neyman–Pearson lemma of statistics. The duality approach to the Neyman–Pearson problem was discussed in Francis and Wright (1969).

Suppose now that (X, Y) is a dual lattice pair. In this case the dual problem becomes

$$\text{(DSTP}_2\text{)}\quad \text{maximize}\quad b^{\mathrm{T}}\lambda - \left\langle e, \bigvee_i (\lambda_i a_i - y_i)\right\rangle$$

$$\text{subject to}\quad \lambda \in \mathbb{R}^n.$$

Define $g_i(\lambda) := \lambda_i a_i - y_i$, and $N_i(\lambda) = N(\bigvee_k g_k(\lambda) - g_i(\lambda))$, for $i = 1, \ldots, n$. Consider the following condition:

$$\left\langle x, \left|\bigvee_{k\neq i} g_k(\lambda) - g_i(\lambda)\right|\right\rangle > 0, \quad \text{for all } 0 \neq x \in [0, e],\ i = 1, \ldots, n. \tag{9.3}$$

As we observed at the end of Section 5, (9.3) can be interpreted in the cases $(X, Y) = (L^p(T, \mu), L^q(T, \mu))$, $1 \leq p \leq \infty$, and $(M(T), C(T))$ as requiring the set $\{\lambda_i a_i(t) - y_i(t) \mid i = 1, \ldots, n\}$ to have a unique largest element μ-a.e. on support(e), and e-a.e., respectively.

In Todd (1978) the case considered is $(X, Y) = (M(T), C(T))$, and $a_i = 1$, $i = 1, \ldots, n$. Assumption 1 in his paper requires

$$e\{t \in T \mid \lambda_i - y_i(t) = \lambda_j - y_j(t)\} = 0,$$

for all $i \neq j$, λ_i, $\lambda_j \neq \mathbb{R}$. This is clearly sufficient to ensure (9.3) holds for all λ.

Theorem 9.6. *With the notation of* (9.1), *suppose that either of Assumptions* 9.1 *and* 9.2 *holds. Suppose also that* (X, Y) *is a dual lattice pair. then the values of* (STP) *and* (DSTP$_2$) *are equal, with attainment in* (DSTP$_2$).

Suppose further that (X, Y) *is a countably regular lattice pair. Then the dual objective function is differentiable at any* λ *for which* (9.3) *holds; at such points we have*

$$X_e = \bigoplus_{i=1}^{n} (X_e \cap N_i(\lambda)), \tag{9.4}$$

and the gradient is given by $(b_i - \langle P_{X_e \cap N_i(\lambda)}(e), a_i\rangle)_{i=1}^n$, *where* $P_{X_e \cap N_i(\lambda)} : X_e \to X_e \cap N_i(\lambda)$, $i = 1, \ldots, n$, *are the natural projections associated with* (9.4).

If furthermore $\bar{\lambda}$ *is optimal for* (DSTP$_2$) *and* (9.3) *holds at* $\bar{\lambda}$ *then the unique optimal solution of the problem* (STP) *is given by* $\bar{x}_i = P_{X_i \cap N_i(\bar{\lambda})}(e)$, $i = 1, \ldots, n$.

Proof. Apply Theorem 8.5. □

In Theorem 5 of Kortanek and Yamasaki (1982) the existence of an optimal solution to (DSTP$_2$) is proved under Assumption 9.2. However, the duality results they give require interiority or local compactness conditions on the cones involved, or X to be a reflexive Banach lattice with $Y = X^*$.

The generalized market area problem

We now turn to a generalization of the semi-infinite transportation problem known as the generalized market area problem (see Lowe and Hurter, 1976). In this problem, instead of constraining the total production at each supply point by $\langle x_i, a_i\rangle = b_i$ for each i, we impose a convex production cost of the form $k((\langle x_i, a_i\rangle)_{i=1}^n)$. The semi-infinite transportation problem is the special case where $k(\cdot) = \delta(\cdot \mid \{b\})$.

The problem we are interested is thus

$$\text{(GMP)} \quad \inf \quad \sum_{i=1}^{n} \langle x_i, y_i\rangle + k((\langle x_j, a_j\rangle)_{j=1}^n)$$

$$\text{subject to} \quad \sum_{i=1}^{n} x_i = e,$$

$$x_i \geq 0, \ x_i \in X, \quad i = 1, \dots, n.$$

If $(X, S_X), (Y, S_Y), \langle\cdot,\cdot\rangle)$ is a pseudo-Riesz pair then by applying Corollary 4.8 we obtain the dual problem

$$(\text{DGMP}_1) \quad \sup \quad -\langle e, y\rangle - k^*(-\lambda)$$

$$\text{subject to} \quad \lambda_i a_i - y \leq y_i, \ i = 1, \dots, n,$$

$$\lambda \in \mathbb{R}^n, \quad y \in Y.$$

Theorem 9.7. *With the notation of* (9.1), *and* $k : \mathbb{R}^n \to]-\infty, \infty]$, *convex, suppose that* $((X, S_X), (Y, S_Y), \langle\cdot,\cdot\rangle)$ *is a pseudo-Riesz pair. Suppose further that for some feasible* $(\hat{x}_1, \dots, \hat{x}_n)$ *for* (GMP) *we have*

$$\begin{aligned} &(\langle \hat{x}_i, a_i\rangle)_1^n \in \operatorname{ri}(\operatorname{dom} k) \quad (\textit{or simply } \operatorname{dom} k \textit{ if } k \textit{ polyhedral}),\\ &\sigma(X, Y)\text{-cl}\, \mathbb{P}[0, \hat{x}_i] = \sigma(X, Y) - \operatorname{cl} \mathbb{P}[0, e], \quad i = 1, \dots, n. \end{aligned} \tag{9.5}$$

Then the values of (GMP) *and* (DGMP_1) *are equal. In this case feasible* $(\bar{x}_1, \dots, \bar{x}_n)$ *for* (GMP) *and* $(\bar{\lambda}; \bar{y})$ *for* (DGMP_1) *are optimal if and only if*

$$\begin{aligned} &\langle \bar{x}_i, y_i + \bar{y} - \bar{\lambda}_i a_i\rangle = 0, \quad i = 1, \dots, n,\\ &-\bar{\lambda} \in \partial k((\langle \bar{x}_i, a_i\rangle)_1^n). \end{aligned} \tag{9.6}$$

If (GMP) *is consistent, k is closed and* $[0, e]$ *is* $\sigma(X, Y)$*-compact then the value of* (GMP) *is attained.*

Proof. Apply Corollary 4.6 and Theorem 3.12. If the values of (GMP) and (DGMP_1) are equal and $(\bar{x}_1, \dots, \bar{x}_m)$ and $(\bar{\lambda}; \bar{y})$ are respectively feasible for the two problems, then they are optimal if and only if we have

$$\begin{aligned}-\langle e, \bar{y}\rangle - k^*(-\bar{\lambda}) &\leq -\sum_i \langle \bar{x}_i, \bar{y}\rangle + k((\langle \bar{x}_i, a_i\rangle)_1^n) + \sum_i \bar{\lambda}_i \langle \bar{x}_i, a_i\rangle \\ &\leq -\sum_i \langle \bar{x}_i, \bar{\lambda}_i a_i - y_i\rangle + k((\langle \bar{x}_i, a_i\rangle)_1^n) + \sum_i \bar{\lambda}_i \langle \bar{x}_i, a_i\rangle \\ &= \sum_i \langle \bar{x}_i, y_i\rangle + k((\langle \bar{x}_i, a_i\rangle)_1^n) \\ &= -\langle e, \bar{y}\rangle - k^*(-\bar{\lambda}),\end{aligned}$$

and (9.6) follows by Rockafellar (1970, 23.5).

To see the last assertion, write (GMP) as

$$\inf_{\substack{u\in\mathbb{R}\\ v\in\mathbb{R}^n}} \left\{ u + k(v) \,\middle|\, u = \sum_i \langle x_i, y_i\rangle,\ v_i = \langle x_i, a_i\rangle \forall i, \text{for some } (x_1, \ldots, x_n) \in F \right\},$$

where F is given by (8.1). If $[0, e]$ is $\sigma(X, Y)$-compact then so is F. Thus (GMP) is equivalent to minimizing a closed, proper convex function over a compact subset of $\mathbb{R}^{n+1}$ (since the continuous image of a compact set is compact). Attainment in (GMP) now follows by Rockafellar (1970, 27.3). □

Now suppose that (X, Y) is a dual lattice pair. In this case the dual problem becomes

$$(\text{DGMP}_2) \quad \text{maximize} \quad -\left\langle e, \bigvee_{i=1}^{n} (\lambda_i a_i - y_i)\right\rangle - k^*(-\lambda)$$

$$\text{subject to} \quad \lambda \in \mathbb{R}^n.$$

Theorem 9.8. *With the notation of* (9.1), *and* $k:\mathbb{R}^n \to\,]-\infty, \infty]$, *convex, suppose that* (X, Y) *is a dual lattice pair and that for some feasible* $(\hat{x}_1, \ldots, \hat{x}_n)$ *for* (GMP), (9.5) *holds. Then the values of* (GMP) *and* (DGMP_2) *are equal, with attainment in* (DGMP_2).

Suppose further that (X, Y) *is a countably regular lattice pair. Suppose that k is closed, that* $\bar{\lambda}$ *is optimal for* (DGMP_2), *and that* (9.3) *holds at* $\bar{\lambda}$. *Then* (9.4) *holds and the unique optimal solution of* (GMP) *is given by* $\bar{x}_i = P_{X_e \cap N_i(\bar{\lambda})}(e)$, $i = 1, \ldots, n$.

Proof. The duality result follows from Corollary 4.8. Under the further conditions we know that (9.4) holds at $\bar{\lambda}$, as in Theorem 9.6, and that the function $\langle e, \bigvee_{i=1}^n (\lambda_i a_i - y_i)\rangle$ is differentiable at $\bar{\lambda}$ with gradient $(\langle \bar{x}_i, a_i\rangle)_1^n$. Since $\bar{\lambda}$ is optimal it follows that $0 \in -(\langle \bar{x}_i, a_i\rangle)_1^n + \partial k^*(-\bar{\lambda})$. Since k is closed, this is equivalent to $-\bar{\lambda} \in \partial k((\langle \bar{x}_i, a_i\rangle)_1^n)$, by Rockafellar (1970, 23.5). If we now set $\bar{y} := \bigvee_{i=1}^n (\bar{\lambda}_i a_i - y_i)$, it follows that $(\bar{x}_1, \ldots, \bar{x}_n)$ and $(\bar{\lambda}; \bar{y})$ are feasible for (GMP) and (DGMP_2) respectively, and satisfy the complementary slackness conditions (9.6). By Theorem 9.7 they are therefore optimal. The uniqueness of $(\bar{x}_1, \ldots, \bar{x}_n)$ follows from the fact that for any optimal $(x_1, \ldots, x_n)$ we have $x_i \in X_e \cap N_i(\bar{\lambda})$, by (9.6), but $\sum_i x_i = e$ and $\sum_i (X_e \cap N_i(\bar{\lambda}))$ is a direct sum by (9.4). □

We can now rederive the results in Todd (1978) by specializing to the case $X = M(K)$, $Y = C(K)$, $K \subset \mathbb{R}^m$ compact, e absolutely continuous with respect to Lebesgue measure, and $a_i \equiv 1$, each $i = 1, \ldots, n$. As we already observed, Assumption 1 in Todd's paper is made to ensure that (9.3) holds for any λ. Many of the results on the semi-infinite transportation problem in the first half of this section could now be rederived from the special case $k(\cdot) = \delta(\cdot \mid \{b\})$. It is worth observing that we could generalize much of the theory of Section 8 on the bounded linear problem (BLP) in much the same way as we have extended the semi-infinite transportation problem to the generalized market area problem.

References

E.J. Anderson and P. Nash, *Linear Programming in Infinite-dimensional Spaces* (Wiley, Chichester, 1987).

A. Ben-Tal, J.M. Borwein and M. Teboulle, "A dual approach to multidimensional L_p spectral estimation problems," *SIAM Journal on control and Optimization* 26 (1988) 985–996.

A. Ben-Tal, J.M. Borwein and M. Teboulle, "Spectral estimation via convex programming," to appear in: *Systems and Management Science by Extremal Methods: Research Honoring Abraham Charnes at Age 70* (Kluwer Academic Publishers, Dordrecht, 1992).

J.M. Borwein, "Convex relations in analysis and optimization," in: S. Schaible and W.T. Ziemba, eds., *Generalized Concavity in Optimization and Economics* (Academic Press, New York, 1981a) pp. 335–377.

J.M. Borwein, "A Lagrange multiplier theorem and a sandwich theorem for convex relations," *Mathematica Scandinavica* 48 (1981b) 189–204.

J.M. Borwein, "Automatic continuity and openness of convex relations," *Proceedings of the American Mathematical Society* 99 (1987) 49–55.

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

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

F.H. Clarke, *Optimization and Nonsmooth Analysis* (Wiley, New York, 1983).

H.W. Corley Jr. and S.D. Roberts, "A partitioning problem with applications in regional design," *Operations Research* 20 (1972) 1010–1019.

C. De Boor, "On 'best' interpolation," *Journal of Approximation Theory* 16 (1976) 28–42.

A.L. Dontchev and B.D. Kalchev, "Duality and well-posedness in convex interpolation," *Numerical Functional Analysis and Optimization* 10 (1989) 673–689.

I. Ekeland and R. Temam, *Convex Analysis and Variational Problems* (North-Holland, Amsterdam, 1976).

J. Favard, "Sur l'interpolation," *Journal de Mathematiques Pures et Appliquees* 19 (1940) 281–306.

R.L. Francis and G.F. Wright, "Some duality relationships for the generalized Neyman–Pearson problem," *Journal of Optimization* 4 (1969) 394–412.

K. Glashoff and S.-A. Gustafson, *Linear Optimization and Approximation* (Springer, New York, 1983).

R.B. Holmes, *Geometric Functional Analysis and Applications* (Springer, New York, 1975).

L.D. Irvine, S.P. Marin and P.W. Smith, "Constrained interpolation and smoothing," *Constructive Approximation* 2 (1986) 129–151.

S. Karlin and W.J. Studden, *Tchebycheff Systems: With Applications in Analysis and Statistics* (Wiley, New York, 1966).

K.O. Kortanek and M. Yamasaki, "Semi-infinite transportation problems," *Journal of Mathematical Analysis and Applications* 88 (1982) 555–565.

T.J. Lowe and A.P. Hurter Jr., "The generalized market area problem," *Management Science* 22 (1976) 1105–1115.

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

R.T. Rockafellar, "Integrals which are convex functionals," *Pacific Journal of Mathematics* 24 (1968) 525–539.

R.T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, NJ, 1970).

R.T. Rockafellar, *Conjugate Duality and Optimization* (SIAM, Philadelphia, PA, 1974).

H.H. Schaefer, *Topological Vector Spaces* (Springer, New York, 1971).

H.H. Schaefer, *Banach Lattices and Positive Operators* (Springer, Berlin, 1974).

L.L. Schumaker, *Spline Functions: Basic Theory* (Wiley, New York, 1981).

I. Singer, *Best Approximation in Normed Linear Spaces by Elements of Linear Subspaces* (Springer, Berlin, 1970).

M.J. Todd, "Solving the generalized market area problem," *Management Science* 24 (1978) 1549–1554.