March 2006 (Revised February 2010) **Report LIDS - 2692**

Extended Monotropic Programming and Duality[1]

by

Dimitri P. Bertsekas[2]

Abstract

We consider the problem

$$\begin{array}{ll} \text{minimize} & \displaystyle\sum_{i=1}^{m} f_i(x_i) \\ \text{subject to} & x \in S, \end{array}$$

where x_i are multidimensional subvectors of x, f_i are convex functions, and S is a subspace. Monotropic programming, extensively studied by Rockafellar, is the special case where the subvectors x_i are the scalar components of x. We show a strong duality result that parallels Rockafellar's result for monotropic programming, and contains other known and new results as special cases. The proof is based on the use of ϵ-subdifferentials and the ϵ-descent method, which is used here as an analytical vehicle.

[1] Research supported by NSF Grant ECS-0218328. Many thanks are due to Paul Tseng and Janey Yu for helpful comments.

[2] Dept. of Electrical Engineering and Computer Science, M.I.T., Cambridge, Mass., 02139.

1. INTRODUCTION

In this paper, we analyze a class of convex optimization problems, using the tools and terminology of convex analysis, e.g., [Roc70], [BNO03]. In particular, we study the problem

$$\begin{aligned} &\text{minimize} \quad \sum_{i=1}^{m} f_i(x_i) \\ &\text{subject to} \quad x \in S, \end{aligned} \tag{1.1}$$

where $x = (x_1, \ldots, x_m)$ with $x_i \in \Re^{n_i}$, $i = 1, \ldots, m$, each $f_i : \Re^{n_i} \mapsto (-\infty, \infty]$, $i = 1, \ldots, m$, is a closed proper convex function, and S is a subspace of $\Re^{n_1 + \cdots + n_m}$.

We refer to this as an *extended monotropic programming problem*. The special case of problem (1.1) where each component x_i is one-dimensional (i.e., $n_i = 1$) is the monotropic programming problem, introduced and extensively analyzed by Rockafellar in his book [Roc84].

Note that problems involving general linear constraints and an additive convex cost function can be converted to extended monotropic programming problems. In particular, the problem

$$\begin{aligned} &\text{minimize} \quad \sum_{i=1}^{m} f_i(x_i) \\ &\text{subject to} \quad Ax = b, \end{aligned} \tag{1.2}$$

where A is a given matrix and b is a given vector, is equivalent to

$$\begin{aligned} &\text{minimize} \quad \sum_{i=1}^{m} f_i(x_i) \\ &\text{subject to} \quad Ax - z = 0, \qquad z = b, \end{aligned}$$

where z is a vector of artificial variables. This is an extended monotropic programming problem, where the constraint subspace is

$$S = \big\{(x, z) \mid Ax - z = 0\big\},$$

and the indicator function of the set $\big\{(x, z) \mid z = b\big\}$ is added to the cost function. When the functions f_i have the form

$$f_i(x_i) = x_i' Q_i x_i + c_i' x_i + \delta_{X_i}(x_i),$$

where Q_i is a positive semidefinite symmetric matrix, c_i is a vector, and $\delta_{X_i}(\cdot)$ is the indicator function of the nonnegative orthant, problem (1.2) reduces to a convex quadratic programming problem. In the special case where $Q_i = 0$, it reduces to a linear programming problem.

Note also that while the subvectors $x_1, \ldots, x_m$ appear independently in the cost function

$$\sum_{i=1}^{m} f_i(x_i),$$

they may be coupled through the subspace constraint. For example, consider a cost function of the form

$$f(x) = h(x_1, \ldots, x_m) + \sum_{i=1}^{m} f_i(x_i), \tag{1.3}$$

where h is a closed proper convex function of all the components x_i. Then, by introducing an auxiliary vector $z \in \Re^{n_1+\cdots+n_m}$, the problem of minimizing $f(x)$ subject to $x \in S$ can be transformed to the problem

$$\begin{aligned} &\text{minimize} \quad h(z) + \sum_{i=1}^{m} f_i(x_i) \\ &\text{subject to} \quad (x, z) \in \overline{S}, \end{aligned}$$

where $\overline{S}$ is the subspace of $\Re^{2(n_1+\cdots+n_m)}$

$$\overline{S} = \big\{(x, x) \mid x \in S\big\}.$$

This problem is of the form (1.1).

Another problem that can be converted to the extended monotropic programming format (1.1) is

$$\begin{aligned} &\text{minimize} \quad \sum_{i=1}^{m} f_i(x) \\ &\text{subject to} \quad x \in S, \end{aligned} \tag{1.4}$$

where $f_i : \Re^n \mapsto (-\infty, \infty]$ are closed proper convex functions, and S is a subspace of $\Re^n$. This can be done by introducing m copies of x, i.e., auxiliary vectors $z_i \in \Re^n$ that are constrained to be equal, and write the problem as

$$\begin{aligned} &\text{minimize} \quad \sum_{i=1}^{m} f_i(z_i) \\ &\text{subject to} \quad (z_1, \ldots, z_m) \in \overline{S}, \end{aligned}$$

where $\overline{S}$ is the subspace

$$\overline{S} = \big\{(x, \ldots, x) \mid x \in S\big\}.$$

The special case of problem (1.4) where $m = 1$ is the generic convex cost problem with linear constraints,

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad x \in S, \end{aligned}$$

where $f : \Re^n \mapsto (-\infty, \infty]$ is a closed proper convex function, and S is a subspace of $\Re^n$ [cf. the earlier discussion regarding problem (1.2)].

It can thus be seen that the extended monotropic programming problem contains as special cases broad classes of important optimization problems. These problems share a powerful and symmetric duality theory that we will develop in this paper. In Section 2, we formulate the dual problem, and prepare for the proof of our strong duality result. This result shows that for a feasible problem, strong duality holds if the set

$$\big\{(z_1+\lambda_1,\ldots,z_m+\lambda_m) \mid (z_1,\ldots,z_m)\in S^\perp,\ \lambda_i\in\partial_\epsilon f_i(x_i)\big\}$$

is closed for all feasible $x=(x_1,\ldots,x_m)$ and $\epsilon>0$, where $\partial_\epsilon f_i(x_i)$ is the ϵ-subdifferential of f_i at x_i.[1] While this is an unusual constraint qualification, it can be translated into readily verifiable conditions by using standard results that address the preservation of closedness of the vector sum of closed convex sets.

To prepare the ground for the proof of our duality result, we discuss in Section 3 the ϵ-descent method, introduced by Bertsekas and Mitter in [BeM71], [BeM73] as a general algorithm for convex nondifferentiable optimization. We use a variant of the method (also given in [BeM71]), which involves projection on an outer approximation of the ϵ-subdifferential. In Section 4, we use the ϵ-descent method to prove our strong duality result. This line of proof is unusual, but a closely related line of proof was used by Rockafellar [1981], [1984] to prove strong duality in the special case of monotropic programming. Rockafellar used a variant of the ϵ-descent method that involves descent along elementary vectors of the subspace S. We modified his argument in order to apply it to extended monotropic programming, both because elementary vectors are not useful in our context, and also because of the need for a constraint qualification that takes the form of closedness of a vector sum of ϵ-subdifferentials. In Section 4, we also discuss various special cases where our result may be applied. As an example, we show that strong duality holds for broad classes of multicommodity network flow problems, and for cost functions of the form (1.3), where h is a real-valued function. It seems hard to extend our results to problems with nonlinear constraints. In particular, a notable result, due to Tseng [Tse05], which asserts the absence of a duality gap in separable convex problems with nonlinear constraints, does not seem to be easily extendable to nonseparable problems using our methodology.

In this paper, all vectors are finite dimensional, and are viewed as column vectors. A prime denotes transposition, so $x'y$ is the inner product of two vectors x and y. We adopt throughout the standard norm, $\|x\| = \sqrt{x'x}$. We use standard terminology, facts, and notation from convex analysis (see e.g., [Roc70], [BNO03]). In summary, for a function $f:\Re^n\mapsto(-\infty,\infty]$, the effective domain $\big\{x \mid f(x)<\infty\big\}$ is denoted by $\mathrm{dom}(f)$, the epigraph $\big\{(x,w)\mid f(x)\le w\big\}$ is denoted by $\mathrm{epi}(f)$, and the closure of f [the function whose epigraph is the closure of $\mathrm{epi}(f)$] is denoted by $\mathrm{cl}\, f$. We say that f is closed if $\mathrm{epi}(f)$ is closed and we say that it is proper if $\mathrm{epi}(f)$ is nonempty and does not contain a vertical line. The conjugate function of a closed proper convex function f is the closed proper convex function $g:\Re^n\mapsto(-\infty,\infty]$ given by

$$g(\lambda)=\sup_{x\in\Re^n}\big\{\lambda'x-f(x)\big\},\qquad \lambda\in\Re^n.$$

[1] The original version of the report had a flaw: the functions f_i were assumed lower semicontinuous within their domain, rather than closed. However, the stronger closedness assumption is needed for the support function formula (3.2), as per [Roc70], p. 220, and hence is essential for our results.

A basic fact for our purposes is that the conjugate of g is f. Furthermore, from the definition of the conjugate, we have Fenchel's inequality

$$f(x) + g(\lambda) \geq \lambda' x, \qquad \forall\ (x, \lambda) \in \Re^{2n},$$

which holds as an equality if and only if λ belongs to the subdifferential $\partial f(x)$ of f at x.

2. THE DUAL PROBLEM

To derive the appropriate dual problem, we introduce auxiliary vectors $z_i \in \Re^{n_i}$ and we convert the extended monotropic programming problem (1.1) to the equivalent form

$$\begin{aligned} &\text{minimize} && \sum_{i=1}^{m} f_i(z_i) \\ &\text{subject to} && z_i = x_i, \quad i = 1, \ldots, m, \qquad x \in S. \end{aligned}$$

We then assign a multiplier vector $\lambda_i \in \Re^{n_i}$ to the equality constraint $z_i = x_i$, thereby obtaining the Lagrangian function

$$L(x, z, \lambda) = \sum_{i=1}^{m} f_i(z_i) + \lambda_i'(x_i - z_i), \tag{2.1}$$

where $\lambda = (\lambda_1, \ldots, \lambda_m)$. The dual function is

$$\begin{aligned} q(\lambda) &= \inf_{x \in S,\, z_i \in \Re^{n_i}} L(x, z, \lambda) \\ &= \inf_{x \in S} \lambda' x + \sum_{i=1}^{m} \inf_{z_i \in \Re^{n_i}} \big\{ f_i(z_i) - \lambda_i' z_i \big\} \\ &= \begin{cases} \sum_{i=1}^{m} q_i(\lambda_i) & \text{if } \lambda \in S^\perp, \\ -\infty & \text{otherwise,} \end{cases} \end{aligned}$$

where

$$q_i(\lambda_i) = \inf_{z_i \in \Re} \big\{ f_i(z_i) - \lambda_i' z_i \big\}, \qquad i = 1, \ldots, m, \tag{2.2}$$

and $S^\perp$ is the orthogonal subspace of S.

Note that since q_i can be written as

$$q_i(\lambda_i) = -\sup_{z_i \in \Re} \big\{ \lambda_i' z_i - f_i(z_i) \big\},$$

it follows that $-q_i$ is the conjugate of f_i, so $-q_i$ is a closed proper convex function. The dual problem is

$$\begin{aligned} &\text{maximize} && \sum_{i=1}^{m} q_i(\lambda_i) \\ &\text{subject to} && \lambda \in S^\perp. \end{aligned} \tag{2.3}$$

Thus, with a change of sign to convert maximization to minimization, the dual problem has the same form as the primal. In fact, assuming that the functions f_i are closed, when the dual problem is dualized, it yields the primal problem, and the duality is fully symmetric.

Since the extended monotropic programming problem can be viewed as a special case of a convex programming problem with linear equality constraints, it is possible to obtain optimality conditions as a special case of classical conditions, which state that (x, λ) is a pair of primal and dual optimal solutions if and only if x is primal feasible, λ is dual feasible, and x minimizes the Lagrangian function (see e.g., [BNO03], Prop. 6.2.5). The Lagrangian minimization condition is in turn true if and only if x_i attains the infimum in the equation

$$q_i(\lambda_i) = \inf_{z_i \in \Re^{n_i}} \big\{ f_i(z_i) - \lambda_i' z_i \big\}, \qquad i = 1, \ldots, m,$$

or equivalently, by Fenchel's inequality,

$$\lambda_i \in \partial f_i(x_i), \qquad i = 1, \ldots, m.$$

We thus obtain the following.

Proposition 2.1: Let f^* be the optimal value of problem (1.1) and assume that $-\infty < f^* < \infty$. The vectors x^* and λ^* are optimal primal and dual solutions, respectively, and the optimal primal and dual costs are equal if and only if

$$x^* \in S, \qquad \lambda^* \in S^\perp, \qquad \lambda_i^* \in \partial f_i(x_i^*), \quad i = 1, \ldots, m.$$

3. THE ϵ-DESCENT METHOD

Given a closed proper convex function $f : \Re^n \mapsto (-\infty, \infty]$ and a scalar $\epsilon > 0$, we say that a vector λ is an *ϵ-subgradient* of f at a point $x \in \mathrm{dom}(f)$ if

$$f(z) \geq f(x) + (z - x)'\lambda - \epsilon, \qquad \forall\, z \in \Re^n. \tag{3.1}$$

The *ϵ-subdifferential*, denoted $\partial_\epsilon f(x)$, is the set of all ϵ-subgradients of f at x, and by convention, $\partial_\epsilon f(x) = \emptyset$ for $x \notin \mathrm{dom}(f)$.

The properties of the ϵ-subdifferential have been discussed extensively; see e.g., Hiriart-Urruty and Lemarechal [HiL93], and Hiriart-Urruty et al. [HMS95]. Let us provide a brief discussion of some of the properties that are useful for our purposes:

(1) For any $x \in \text{dom}(f)$ and $\epsilon > 0$, $\partial_\epsilon f(x)$ is nonempty and closed, and it is also compact if x is in the interior of $\text{dom}(f)$.

(2) The support function of $\partial_\epsilon f(x)$ is given by the formula ([Roc70], p. 220)

$$\sup_{\lambda \in \partial_\epsilon f(x)} y'\lambda = \inf_{\alpha > 0} \frac{f(x + \alpha y) - f(x) + \epsilon}{\alpha}, \qquad y \in \Re^n. \tag{3.2}$$

(3) We say that a direction y is an *ϵ-descent direction at $x \in dom(f)$* if

$$\inf_{\alpha > 0} f(x + \alpha y) < f(x) - \epsilon.$$

By Eq. (3.2) it follows that

$$y \text{ is an } \epsilon\text{-descent direction} \quad \text{if and only if} \quad \sup_{\lambda \in \partial_\epsilon f(x)} y'\lambda < 0.$$

In particular, if $0 \notin \partial_\epsilon f(x)$ and $\overline{\lambda}$ is the projection of the origin on $\partial_\epsilon f(x)$, the vector $-\overline{\lambda}$ is an ϵ-descent direction.

The ϵ-descent method is based on observation (3) above. It starts at some $x_0 \in \text{dom}(f)$ and generates a sequence $\{x_k\} \subset \text{dom}(f)$. The kth iteration is

$$x_{k+1} = x_k + \alpha_k y_k, \tag{3.3}$$

y_k is an ϵ-descent direction (if one can be found) and α_k is a positive stepsize that reduces the cost function by more than ϵ, i.e.,

$$f(x_k + \alpha_k y_k) < f(x_k) - \epsilon.$$

The iteration can be implemented by finding the projection of the origin on $\partial_\epsilon f(x_k)$,

$$\lambda_k = \arg \min_{\lambda \in \partial_\epsilon f(x_k)} \|\lambda\|.$$

If $\lambda_k \neq 0$, then by observation (3) above, $-\lambda_k$ is an ϵ-descent direction, and can be used as the direction y_k in the iteration (3.3).

We will use a variant of this implementation where $\partial_\epsilon f(x_k)$ is approximated by a closed set $A(x_k)$ such that

$$\partial_\epsilon f(x_k) \subset A(x_k) \subset \partial_{\gamma\epsilon} f(x_k),$$

where γ is a scalar with $\gamma > 1$. In this variant, the direction used in iteration (3.3) is $y_k = -\lambda_k$, where

$$\lambda_k = \arg \min_{\lambda \in A(x_k)} \|\lambda\|$$

is the projection of the origin on $A(x_k)$. If $\lambda_k = 0$ [equivalently $0 \in A(x_k)$], the method stops, and it follows that x_k is within $\gamma\epsilon$ of being optimal. If $\lambda_k \neq 0$, it follows that by suitable choice of the stepsize α_k, we can move along the direction $y_k = -\lambda_k$ to decrease the cost function by more than ϵ. Thus for a fixed $\epsilon >$ and assuming that f is bounded below, the method is guaranteed to terminate in a finite number of iterations with a $\gamma\epsilon$-optimal solution.

We now focus on the case where f is the sum of functions,

$$f(x) = f_1(x) + \cdots + f_m(x).$$

The following proposition shows that we may use as approximation the closure of the vector sum of the ϵ-subdifferentials:

$$A(x) = \mathrm{cl}\big(\partial_\epsilon f_1(x) + \cdots + \partial_\epsilon f_m(x)\big).$$

This case, and a corresponding ϵ-descent algorithm, were discussed in [BeM71] under the assumption that the functions f_i are real-valued, in which case the sets $\partial_\epsilon f_i(x)$ are compact and the closure operation is unnecessary in the preceding equation. Other versions of the following result are also known; see Hiriart-Urruty et al. [HMS95], Th. 3.2.

Proposition 3.1: Let f be the sum of m closed proper convex functions $f_i : \Re^n \mapsto (-\infty, \infty]$, $i = 1, \ldots, m$,

$$f(x) = f_1(x) + \cdots + f_m(x),$$

and let ϵ be a positive scalar. For a vector $x \in \mathrm{dom}(f)$

$$\partial_\epsilon f(x) \subset \mathrm{cl}\big(\partial_\epsilon f_1(x) + \cdots + \partial_\epsilon f_m(x)\big) \subset \partial_{m\epsilon} f(x). \tag{3.4}$$

Proof: Let $\lambda_i \in \partial_\epsilon f_i(x)$ for $i = 1, \ldots, m$. Then we have

$$f_i(z) \geq f_i(x) + \lambda_i'(z - x) - \epsilon, \qquad \forall\ z \in \Re^n,\ i = 1, \ldots, m,$$

and by adding over all i, we obtain

$$f(z) \geq f(x) + (\lambda_1 + \cdots + \lambda_m)'(z - x) - m\epsilon, \qquad \forall\ z \in \Re^n.$$

Hence $\lambda_1 + \cdots + \lambda_m \in \partial_{m\epsilon} f(x)$, and it follows that

$$\partial_\epsilon f_1(x) + \cdots + \partial_\epsilon f_m(x) \subset \partial_{m\epsilon} f(x).$$

Since $\partial_{m\epsilon} f(x)$ is closed, this proves the right-hand side of Eq. (3.4).

To prove the left-hand side of Eq. (3.4), assume to arrive at a contradiction, that there exists a $\lambda \in \partial_\epsilon f(x)$ such that

$$\lambda \notin \mathrm{cl}\big(\partial_\epsilon f_1(x) + \cdots + \partial_\epsilon f_m(x)\big).$$

Then there exists a hyperplane strictly separating λ from the set $\mathrm{cl}\big(\partial_\epsilon f_1(x) + \cdots + \partial_\epsilon f_m(x)\big)$, i.e., there exist a vector y and a scalar b such that

$$y'(\lambda_1 + \cdots + \lambda_m) < b < y'\lambda, \qquad \forall\ \lambda_1 \in \partial_\epsilon f_1(x), \ldots, \lambda_m \in \partial_\epsilon f_m(x).$$

From this we obtain

$$\sup_{\lambda_1 \in \partial_\epsilon f_1(x)} y'\lambda_1 + \cdots + \sup_{\lambda_m \in \partial_\epsilon f_m(x)} y'\lambda_m < y'\lambda,$$

so that by Eq. (3.2),

$$\inf_{\alpha > 0} \frac{f_1(x + \alpha y) - f_1(x) + \epsilon}{\alpha} + \cdots + \inf_{\alpha > 0} \frac{f_m(x + \alpha y) - f_m(x) + \epsilon}{\alpha} < y'\lambda.$$

It follows that there exist positive scalars $\alpha_1, \ldots, \alpha_m$ such that

$$\frac{f_1(x + \alpha_1 y) - f_1(x) + \epsilon}{\alpha_1} + \cdots + \frac{f_m(x + \alpha_m y) - f_m(x) + \epsilon}{\alpha_m} < y'\lambda. \tag{3.5}$$

Let

$$\overline{\alpha} = \min\{\alpha_1, \ldots, \alpha_m\}.$$

By the convexity of f_i, the ratio $\big(f_i(x + \alpha y) - f_i(x)\big)/\alpha$ is monotonically nondecreasing in α. Thus, since $\alpha_i \geq \overline{\alpha}$, we have

$$\frac{f_i(x + \alpha_i y) - f_i(x)}{\alpha_i} \geq \frac{f_i(x + \overline{\alpha} y) - f_i(x)}{\overline{\alpha}}, \qquad i = 1, \ldots, m,$$

and from Eq. (3.5) and the definition of $\overline{\alpha}$ we obtain

$$\begin{aligned} y'\lambda &> \frac{f_1(x + \alpha_1 y) - f_1(x) + \epsilon}{\alpha_1} + \cdots + \frac{f_m(x + \alpha_m y) - f_m(x) + \epsilon}{\alpha_m} \\ &\geq \frac{f_1(x + \overline{\alpha} y) - f_1(x) + \epsilon}{\overline{\alpha}} + \cdots + \frac{f_m(x + \overline{\alpha} y) - f_m(x) + \epsilon}{\overline{\alpha}} \\ &= \frac{f(x + \overline{\alpha} y) - f(x) + \epsilon}{\overline{\alpha}} \\ &\geq \inf_{\alpha > 0} \frac{f(x + \alpha y) - f(x) + \epsilon}{\alpha}. \end{aligned}$$

Since $\lambda \in \partial_\epsilon f(x)$, this contradicts Eq. (3.2), and proves the left-hand side of Eq. (3.4). **Q.E.D.**

The potential lack of closure of the set $\partial_\epsilon f_1(x) + \cdots + \partial_\epsilon f_m(x)$ indicates a practical difficulty in implementing the method. In particular, in order to find an ϵ-descent direction one will ordinarily minimize $\|\lambda_1 + \cdots + \lambda_m\|$ over $\lambda_i \in \partial_\epsilon f_i(x)$, $i = 1, \ldots, m$, but an optimal solution to this problem may not exist. Thus, it may be difficult to check algorithmically whether

$$0 \in \mathrm{cl}\big(\partial_\epsilon f_1(x) + \cdots + \partial_\epsilon f_m(x)\big),$$

which is the test for $m\epsilon$-optimality of x. We will see in the next section that the lack of closure of the set $\partial_\epsilon f_1(x) + \cdots + \partial_\epsilon f_m(x)$ may be the cause of a duality gap in the extended monotropic programming context.

4. STRONG DUALITY THEOREM

We are now ready to prove the main result of the paper. Let f^* and q^* be the optimal values of the primal and dual problems (1.1) and (2.3), respectively, and note that by weak duality, we have $q^* \leq f^*$. Let us introduce the functions $\overline{f}_i : \Re^{n_1+\cdots+n_m} \mapsto (-\infty, \infty]$ of the vector $x = (x_1, \ldots, x_m)$, defined by

$$\overline{f}_i(x) = f_i(x_i), \qquad i = 1, \ldots, m.$$

Note that the ϵ-subdifferentials of $\overline{f}_i$ and f_i are related by

$$\partial_\epsilon \overline{f}_i(x) = \big\{(0, \ldots, 0, \lambda_i, 0, \ldots, 0) \mid \lambda_i \in \partial_\epsilon f_i(x_i)\big\}, \quad i = 1, \ldots, m, \tag{4.1}$$

where the nonzero element in $(0, \ldots, 0, \lambda_i, 0, \ldots, 0)$ is in the ith position. The following proposition gives conditions for strong duality.

Proposition 4.1: Assume that the extended monotropic programming problem (1.1) is feasible, and that for all feasible solutions x, the set

$$T(x, \epsilon) = S^\perp + \partial_\epsilon \overline{f}_1(x) + \cdots + \partial_\epsilon \overline{f}_m(x)$$

is closed for all $\epsilon > 0$. Then $q^* = f^*$.

Proof: If $f^* = -\infty$, then $q^* = f^*$ by weak duality, so we may assume that $f^* > -\infty$. Let F denote the feasible region of the primal problem:

$$F = S \cap \big(\cap_{i=1}^m \mathrm{dom}(\overline{f}_i)\big).$$

We apply the ϵ-descent method based on outer approximation of the subdifferential (cf. Section 3) to the minimization of the function

$$f(x) = \delta_S(x) + \sum_{i=1}^{m} \overline{f}_i(x) = \delta_S(x) + \sum_{i=1}^{m} f_i(x_i),$$

where δ_S is the indicator function of S. In this method, we start with a vector $x^0 \in F$, and we generate a sequence $\{x^k\} \subset F$. At the kth iteration, given the current iterate x^k, we find the vector of minimum norm w^k on the set $T(x^k, \epsilon)$ (which is closed by assumption). If $w^k = 0$ the method stops, verifying that $0 \in \partial_{m\epsilon} f(x^k)$ (cf. Prop. 3.1). If $w^k \neq 0$, we generate a vector $x^{k+1} \in F$ of the form $x^{k+1} = x^k - \alpha^k w^k$, such that

$$f(x^{k+1}) < f(x^k) - \epsilon;$$

such a vector is guaranteed to exist, since $0 \notin T(x^k, \epsilon)$ and hence $0 \notin \partial_\epsilon f(x^k)$ by Prop. 3.1. Since $f(x^k) \geq f^*$ and we have assumed that $f^* > -\infty$, the method must stop at some iteration with a vector $x = (x_1, \ldots, x_m)$ such that $0 \in T(x, \epsilon)$. Thus some vector in $\partial_\epsilon \overline{f}_1(x) + \cdots + \partial_\epsilon \overline{f}_m(x)$ must belong to $S^\perp$. In view of Eq. (4.1), it follows that there must exist vectors

$$\lambda_i \in \partial_\epsilon f_i(x_i), \qquad i = 1, \ldots, m,$$

such that

$$\lambda = (\lambda_1, \ldots, \lambda_m) \in S^\perp.$$

From the definition of an ϵ-subgradient, we have [cf. Eqs. (2.2) and (3.1)]

$$f_i(x_i) \leq q_i(\lambda_i) + \lambda_i' x_i + \epsilon, \qquad i = 1, \ldots, m,$$

and by adding over i and using the fact $x \in S$ and $\lambda \in S^\perp$, we obtain

$$\sum_{i=1}^{m} f_i(x_i) \leq \sum_{i=1}^{m} q_i(\lambda_i) + m\epsilon.$$

Since x is primal feasible and λ is dual feasible, it follows that

$$f^* \leq q^* + m\epsilon.$$

Taking the limit as $\epsilon \to 0$, we obtain $f^* = q^*$. **Q.E.D.**

Some Special Cases

We now delineate some special cases where the assumptions of the preceding proposition are satisfied. We first note that in view of Eq. (4.1), the set $\partial_\epsilon \overline{f}_i(x)$ is compact if $\partial_\epsilon f_i(x_i)$ is compact, and it is polyhedral if $\partial_\epsilon f_i(x_i)$ is polyhedral. Since the vector sum of a compact set and a polyhedral set is closed ([Roc70], Th. 20.3, [BMO03], p. 68), it follows that if *each of the sets $\partial_\epsilon f_i(x_i)$ is either compact or polyhedral, then $T(x,\epsilon)$ is closed, and by Prop. 4.1, we have $q^* = f^*$.* Furthermore, the set $\partial_\epsilon f_i(x_i)$ is compact if $x_i \in \text{int}\big(\text{dom}(f_i)\big)$ (as in the case where f_i is real-valued), and it is polyhedral if f_i is polyhedral.[1] There are some other interesting special cases where $\partial_\epsilon f_i(x_i)$ is polyhedral, as we now describe.

One such special case is when f_i depends on a single scalar component of x, as in the case of a monotropic programming problem. The following definition introduces a more general case.

Definition 4.1: We say that a closed proper convex function $h : \Re^n \mapsto (-\infty,\infty]$ is *essentially one-dimensional* if it has the form

$$h(x) = \overline{h}(a'x),$$

where a is a vector in $\Re^n$ and $\overline{h} : \Re \mapsto (-\infty,\infty]$ is a scalar closed proper convex function.

The following proposition establishes the main associated property for our purposes. A proof may be obtained by using general results on the ϵ-subdifferential of the composition of a convex function and a linear function (see Hiriart-Urruty et al. [HMS95], Th. 7.1). We give here a simpler specialized proof.

Proposition 4.2: Let $h : \Re^n \mapsto (-\infty,\infty]$ be a closed proper convex essentially one-dimensional function. Then for all $x \in \text{dom}(h)$ and $\epsilon > 0$, the ϵ-subdifferential $\partial_\epsilon h(x)$ is nonempty and polyhedral.

Proof: Let $h(x) = \overline{h}(a'x)$, where a is a vector in $\Re^n$ and $\overline{h}$ is a scalar closed proper convex function. If $a = 0$, then h is a constant function, and $\partial_\epsilon h(x)$ is equal to $\{0\}$, a polyhedral set. Thus, we may assume that $a \neq 0$. We note that $\lambda \in \partial_\epsilon h(x)$ if and only if

$$\overline{h}(a'z) \geq \overline{h}(a'x) + (z-x)'\lambda - \epsilon, \qquad \forall\ z \in \Re^n.$$

[1] In our use of the term, a polyhedral set is a nonempty set that is specified by a finite number of affine inequalities (as defined in [BNO03]). A polyhedral function is an extended real-valued function whose epigraph is a polyhedral set.

Writing λ in the form $\lambda = \xi a + v$ with $\xi \in \Re$ and $v \perp a$, we have

$$\overline{h}(a'z) \geq \overline{h}(a'x) + (z - x)'(\xi a + v) - \epsilon, \qquad \forall\ z \in \Re^n,$$

and by taking $z = \gamma a + \delta v$ with $\gamma, \delta \in \Re$ and γ chosen so that $\gamma\|a\|^2 \in \mathrm{dom}(\overline{h})$, we obtain

$$\overline{h}\big(\gamma\|a\|^2\big) \geq \overline{h}(a'x) + (\gamma a + \delta v - x)'\lambda - \epsilon = \overline{h}(a'x) + (\gamma a - x)'\lambda - \epsilon + \delta v'\lambda, \qquad \forall\ \delta \in \Re.$$

Since $v'\lambda = \|v\|^2$ and δ can be arbitrarily large, this relation implies that $v = 0$, so it follows that every $\lambda \in \partial_\epsilon h(x)$ must be a scalar multiple of a. Since $\partial_\epsilon h(x)$ is also a closed convex set, it must be a nonempty closed interval in $\Re^n$, and hence is polyhedral. **Q.E.D.**

Another interesting special case is described in the following definition.

Definition 4.2: We say that a closed proper convex function $h : \Re^n \mapsto (-\infty, \infty]$ is *domain one-dimensional* if the affine hull of $\mathrm{dom}(h)$ is either a single point or a line, i.e.,

$$\mathrm{aff}\big(\mathrm{dom}(h)\big) = \{\gamma a + b \mid \gamma \in \Re\},$$

where a and b are some vectors in $\Re^n$.

The following proposition parallels Prop. 4.2.

Proposition 4.3: Let $h : \Re^n \mapsto (-\infty, \infty]$ be a closed proper convex domain one-dimensional function. Then for all $x \in \mathrm{dom}(h)$ and $\epsilon > 0$, the ϵ-subdifferential $\partial_\epsilon h(x)$ is nonempty and polyhedral.

Proof: Denote by a and b the vectors associated with the domain of h as per Definition 4.2. We note that for $\overline{\gamma} a + b \in \mathrm{dom}(h)$, we have $\lambda \in \partial_\epsilon h(\overline{\gamma} a + b)$ if and only if

$$h(\gamma a + b) \geq h(\overline{\gamma} a + b) + (\gamma - \overline{\gamma})a'\lambda - \epsilon, \qquad \forall\ \gamma \in \Re,$$

or equivalently, if and only if $a'\lambda \in \partial_\epsilon \overline{h}(\overline{\gamma})$, where $\overline{h}$ is the one-dimensional convex function

$$\overline{h}(\gamma) = h(\gamma a + b), \qquad \gamma \in \Re.$$

Thus,

$$\partial_\epsilon h(\overline{\gamma} a + b) = \big\{\lambda \mid a'\lambda \in \partial_\epsilon \overline{h}(\overline{\gamma})\big\}.$$

Since $\partial_\epsilon \overline{h}(\overline{\gamma})$ is a nonempty closed interval ($\overline{h}$ is closed because h is), it follows that $\partial_\epsilon h(\overline{\gamma} a + b)$ is nonempty and polyhedral [if $a = 0$, it is equal to $\Re^n$, and if $a \neq 0$, it is the vector sum of two polyhedral sets: the interval $\big\{\gamma a \mid \gamma \|a\|^2 \in \partial_\epsilon \overline{h}(\overline{\gamma})\big\}$ and the subspace that is orthogonal to a]. **Q.E.D.**

By combining the preceding two propositions with Prop. 4.1, we obtain the following.

Proposition 4.4: Assume that the extended monotropic programming problem (1.1) is feasible, and that each function f_i is real-valued, or is polyhedral, or is essentially one-dimensional, or is domain one-dimensional. Then $q^* = f^*$.

Here is an example of a class of problems where strong duality is implied by Prop. 4.4.

Example 4.1: (Multicommodity Network Flow Problems)

Consider a directed graph consisting of a set $\mathcal{N}$ of nodes and a set $\mathcal{A}$ of directed arcs. The flows on the arcs are of K different types (commodities). We denote by $x_{ij}(k)$ the flow of kth type on arc (i,j) $(k = 1,\ldots,K)$. These flows must satisfy conservation of flow and supply/demand constraints of the form

$$\sum_{\{j \mid (i,j)\in\mathcal{A}\}} x_{ij}(k) - \sum_{\{j \mid (j,i)\in\mathcal{A}\}} x_{ji}(k) = s_i(k), \quad \forall\, i \in \mathcal{N},\ k = 1,\ldots,K, \tag{4.2}$$

where $s_i(k)$ is the amount of flow of type k entering the network at node i [$s_i(k) > 0$ indicates supply, and $s_i(k) < 0$ indicates demand]. The supplies/demands $s_i(k)$ are given and for the problem to be feasible, they must satisfy

$$\sum_{i\in\mathcal{N}} s_i(k) = 0, \qquad k = 1,\ldots,K, \tag{4.3}$$

(total supply and total demand of each type should be equal).

The problem is to minimize

$$\sum_{(i,j)\in\mathcal{A}} f_{ij}(x_{ij})$$

subject to the constraints (4.2), where $f_{ij} : \Re \mapsto (-\infty,\infty]$ are closed proper convex functions of the total flow on arc (i,j), i.e., the sum

$$x_{ij} = x_{ij}(1) + \cdots + x_{ij}(K). \tag{4.4}$$

In typical applications in communication and transportation contexts (see e.g, [BeG92], [Ber98]), the function f_{ij} is monotonically increasing, thus representing a penalty for a large amount of total flow x_{ij} on arc (i,j). Furthermore, f_{ij} may embody a capacity constraint, whereby x_{ij} should lie within certain bounds.

We can formulate the problem into the extended monotropic programming format (1.1) by introducing an additional variable $z_i(k)$ for each $i \in \mathcal{N}$ and $k = 1, \ldots, K$, and by converting the conservation of flow constraint (4.2) to the subspace constraint

$$\sum_{\{j \mid (i,j) \in \mathcal{A}\}} x_{ij}(k) - \sum_{\{j \mid (j,i) \in \mathcal{A}\}} x_{ji}(k) - z_i(k) = 0, \qquad \forall\, i \in \mathcal{N},\ k = 1, \ldots, K,$$

while changing the cost function to

$$\sum_{(i,j) \in \mathcal{A}} f_{ij}(x_{ij}) + \sum_{k=1}^{K} \sum_{i \in \mathcal{N}} d_{ik}\big(z_i(k)\big), \tag{4.5}$$

where d_{ik} is the function

$$d_{ik}(z) = \begin{cases} 0 & \text{if } z = s_i(k), \\ \infty & \text{otherwise.} \end{cases}$$

It can be seen, using the definition (4.4) of x_{ij}, that the cost function (4.5) is the sum of closed proper convex functions that are essentially one-dimensional. It follows from Prop. 4.4 that if the optimal value of the problem is finite, there is no duality gap. This conclusion holds also for some more general versions of the problem. For example, the cost function may contain an additional real-valued convex function and/or a polyhedral function that depends on all the arc flows $x_{ij}(k)$. Furthermore, instead of being fixed, the supply/demand amounts may be variable and subject to optimization under the constraint

$$\sum_{i \in \mathcal{N}} z_i(k) = 0, \qquad k = 1, \ldots, K,$$

[cf. Eq. (4.3)], while the functions d_{ik} in the cost (4.5) may be replaced by arbitrary closed proper convex functions of $z_i(k)$.

It turns out that there is a conjugacy relation between essentially one-dimensional functions and domain one-dimensional functions such that the affine hull of their domain is a subspace. This is shown in the following proposition, which establishes a somewhat more general connection, needed for our purposes.

Proposition 4.5:

(a) The conjugate of an essentially one-dimensional function is a domain one-dimensional function such that the affine hull of its domain is a subspace.

(b) The conjugate of a domain one-dimensional function is the sum of an essentially one-dimensional function and a linear function.

Proof: (a) Let $h : \Re^n \mapsto (-\infty, \infty]$ be essentially one-dimensional, so that

$$h(x) = \overline{h}(a'x),$$

where a is a vector in $\Re^n$ and $\overline{h} : \Re \mapsto (-\infty, \infty]$ is a scalar closed proper convex function. If $a = 0$, then h is a constant function, so its conjugate is domain one-dimensional, since its domain is $\{0\}$. We may thus assume that $a \neq 0$. We claim that the conjugate

$$g(\lambda) = \sup_{x \in \Re^n} \big\{\lambda' x - \overline{h}(a'x)\big\}, \tag{4.6}$$

takes infinite values if λ is outside the one-dimensional subspace spanned by a, implying that g is domain one-dimensional with the desired property. Indeed, let λ be of the form $\lambda = \xi a + v$, where ξ is a scalar, and v is a nonzero vector with $v \perp a$. If we take $x = \gamma a + \delta v$ in Eq. (4.6), where γ is such that $\gamma \|a\|^2 \in \mathrm{dom}(\overline{h})$, we obtain

$$\begin{aligned} g(\lambda) &= \sup_{x \in \Re^n} \big\{\lambda' x - \overline{h}(a'x)\big\} \\ &\geq \sup_{\delta \in \Re} \big\{(\xi a + v)'(\gamma a + \delta v) - \overline{h}\big(\gamma \|a\|^2\big)\big\} \\ &= \xi \gamma \|a\|^2 - \overline{h}\big(\gamma \|a\|^2\big) + \sup_{\delta \in \Re} \big\{\delta \|v\|^2\big\}, \end{aligned}$$

so it follows that $g(\lambda) = \infty$.

(b) Let $h : \Re^n \mapsto (-\infty, \infty]$ be domain one-dimensional, so that

$$\mathrm{aff}\big(\mathrm{dom}(h)\big) = \{\gamma a + b \mid \gamma \in \Re\},$$

for some vectors a and b. If $a = b = 0$, the domain of h is $\{0\}$, so its conjugate is the function taking the constant value $-h(0)$ and is essentially one-dimensional. If $b = 0$ and $a \neq 0$, then the conjugate is

$$g(\lambda) = \sup_{x \in \Re^n} \big\{\lambda' x - h(x)\big\} = \sup_{\gamma \in \Re} \big\{\gamma a' \lambda - h(\gamma a)\big\},$$

so $g(\lambda) = \overline{g}(a'\lambda)$ where $\overline{g}$ is the conjugate of the scalar function $\overline{h}(\gamma) = h(\gamma a)$. Since $\overline{h}$ is closed, convex, and proper, the same is true for $\overline{g}$, and it follows that g is essentially one-dimensional. Finally, consider the case where $b \neq 0$. Then we use a translation argument and write $h(x) = \hat{h}(x - b)$, where $\hat{h}$ is a function such that the affine hull of its domain is the subspace spanned by a. The conjugate of $\hat{h}$ is essentially one-dimensional (by the preceding argument), and the conjugate of h is obtained by adding $b'\lambda$ to it. **Q.E.D.**

We now turn to the dual problem, and derive a duality result that is analogous to the one of Prop. 4.4. We say that a function is *co-finite* if its conjugate is real-valued (see [Roc70], p. 116). If we apply Prop. 4.4 to the dual problem (2.3), we obtain the following.

Proposition 4.6: Assume that the dual extended monotropic programming problem (2.3) is feasible. Assume further that each f_i is co-finite, or is polyhedral, or is essentially one-dimensional, or is domain one-dimensional. Then $q^* = f^*$.

In the special case of a monotropic programming problem, where the functions f_i are essentially one-dimensional (they depend on the single scalar component x_i), Props. 4.4 and 4.6 yield the following.

Proposition 4.7: Consider the monotropic programming problem, where $n_i = 1$ for all i. Assume that either the problem is feasible, or else its dual problem is feasible. Then $q^* = f^*$.

Proof: This is a consequence of Props. 4.4 and 4.6, and the fact that when $n_i = 1$, the functions f_i and q_i are essentially one-dimensional. Applying Prop. 4.4 to the primal problem, shows that $q^* = f^*$ under the hypothesis that the primal problem is feasible. Applying Prop. 4.6 to the dual problem, shows that $q^* = f^*$ under the hypothesis that the dual problem is feasible. **Q.E.D.**

5. REFERENCES

[BNO03] Bertsekas, D. P., with Nedić, A., and Ozdaglar, A. E., 2003. Convex Analysis and Optimization, Athena Scientific, Belmont, MA.

[BeG92] Bertsekas, D. P., and Gallager, R. G., 1992. Data Networks (2nd Edition), Prentice-Hall, Englewood Cliffs, N. J.

[BeM71] Bertsekas, D. P., and Mitter, S. K., 1971. "Steepest Descent for Optimization Problems with Nondifferentiable Cost Functionals," Proc. 5th Annual Princeton Confer. Inform. Sci. Systems, Princeton, N. J., pp. 347-351.

[BeM73] Bertsekas, D. P., and Mitter, S. K., 1973. "A Descent Numerical Method for Optimization Problems with Nondifferentiable Cost Functionals," SIAM J. on Control, Vol. 11, pp. 637-652.

[Ber98] Bertsekas, D. P., 1998. Network Optimization: Continuous and Discrete Models, Athena Scientific, Belmont, MA.

[HMS95] Hiriart-Urruty, J.-B., Moussaoui, M., Seeger, A., and Volle, M., 1995. "Subdifferential Calculus Without Qualification Conditions, Using Approximate Subdifferentials: A Survey," Nonlinear Analysis, Vol. 24, pp. 1727-1754.

[HiL93] Hiriart-Urruty, J.-B., and Lemarechal, C., 1993. Convex Analysis and Minimization Algorithms, Vols. I and II, Springer-Verlag, Berlin and N. Y.

[Roc70] Rockafellar, R. T., 1970. Convex Analysis, Princeton Univ. Press, Princeton, N. J.

[Roc81] Rockafellar, R. T., 1981. "Monotropic Programming: Descent Algorithms and Duality," in Nonlinear Programming 4, by Mangasarian, O. L., Meyer, R. R., and Robinson, S. M. (eds.), Academic Press, N. Y., pp. 327-366.

[Roc84] Rockafellar, R. T., 1984. Network Flows and Monotropic Optimization, Wiley, N. Y.; republished by Athena Scientific, Belmont, MA, 1998.

[Tse05] Tseng, P., 2005. "Some Convex Programs Without a Duality Gap," Report, Mathematics Department, University of Washington, Seattle; revised February 2006; to appear in Mathematical Programming, B.