Mathematical Programming 63 (1994) 193–212
North-Holland

Global minimization by reducing the duality gap

Aharon Ben-Tal, Gideon Eiger and Vladimir Gershovitz
Faculty of Industrial Engineering and Management, Technion – Israel Institute of Technology, Haifa, Israel

Received 13 April 1992
Revised manuscript received 5 April 1993

We derive a general principle demonstrating that by partitioning the feasible set, the duality gap, existing between a nonconvex program and its lagrangian dual, can be reduced, and in important special cases, even eliminated. The principle can be implemented in a Branch and Bound algorithm which computes an approximate global solution and a corresponding lower bound on the global optimal value. The algorithm involves decomposition and a nonsmooth local search. Numerical results for applying the algorithm to the pooling problem in oil refineries are given.

Key words: Global optimization, nonconvex programming, duality gap, branch and bound method, decomposition, nonsmooth optimization, pooling problem.

Introduction

Many engineering design problems give rise to nonlinear programs of the form

$$(\mathrm{P}_Q) \qquad \min_{x\in\mathbb{R}^n,\, q\in\mathbb{R}^p} \{f_0(q,x) \colon q\in Q,\, x\in X(q)\}$$

where Q is a nonempty subset of $\mathbb{R}^p$ and $X(q)$ is given by

$$X(q)=\{x\in\mathbb{R}^n\colon f_i(q,x)\leqslant 0,\ i=1,\ \dots,\ m\}.$$

In most applications, the objective function f_0 and the constraint functions $\{f_i\colon i\geqslant 1\}$ are *not* jointly convex. On the other hand, they possess "partial convexity" in the sense that for every fixed $q\in Q$, the function $f_0(q,\ \cdot\ )$ is a convex function, and $X(q)$ is a convex set. Moreover, the dimension of q is much smaller than the dimension of x ($p \ll n$). Thus a natural approach to solve (P_Q) is by decomposition:

$$(\mathrm{P}_Q) \quad \min_{q\in Q}\left\{\varphi(q) =: \min_{x\in X(q)} f_0(q,x)\right\}.$$

The inner problem is then a *convex program*, while the outer problem is a nonconvex

Correspondence to: Prof. Aharon Ben-Tal, Faculty of Industrial Engineering and Management, Technion – Israel Institute of Technology, Haifa 32000, Israel.

Research supported by Shell Laboratorium, Amsterdam, and GIF – The German–Israel Foundation for Scientific Research and Development.

SSDI 0025-5610(93)E0038-G

program, but of small dimension. An extra difficulty is caused by the fact that $\varphi(\cdot)$ may *well* be a *nonsmooth* function (e.g. if the inner problem is not uniquely solved for certain q's). While the nonsmoothness can be dealt with satisfactorily (see e.g. [3]) due to the availability of good algorithms (such as the Bundle Trust method of Schramm and Zowe [11], or the R-algorithm of Shor [12]), it is the nonconvexity of the outer problem

$$\min_{q \in Q} \varphi(q) \tag{1}$$

which poses the main challenge.

An important issue in global optimization is how to estimate the quality of a candidate solution $\bar{q}$ (obtained, perhaps, by multi-start local search, or any other heuristic). Such an estimate can be obtained by finding a tight lower bound L for problem (1); the quality of $\bar{q}$ is determined then by how small is the gap $\varphi(\bar{q}) - L$.

The most general way to obtain lower bounds on the value of (P_Q) is by *duality*. A dual problem (D_Q) is constructed in terms of the Lagrangian

$$L(q, x; y) =: f_0(x, q) + \sum_{i=1}^{m} y_i f_i(q, x) \tag{2}$$

by

$$(\mathrm{D}_Q) \quad \max_{y \geqslant 0} \left\{ h(y) =: \min_{q \in Q, x} L(q, x, y) \right\},$$

rendering a lower bound due to the *weak duality* relation:

$$\min(\mathrm{P}_Q) \geqslant \max(\mathrm{D}_Q).$$

Unfortunately, strict inequality usually holds, due to the nonconvexity of (P_Q), and the duality gap

$$\min(\mathrm{P}_Q) - \max(\mathrm{D}_Q) \tag{3}$$

may be large, in which event the duality approach is not useful.

In this paper we show a general way to reduce the duality gap, by partitioning the set Q. Let $\{Q_i: i \in I\}$ be such a partitioning, i.e.

$$Q = \bigcup_{i \in I} Q_i;$$

let (P_{Q_i}) denote the primal problem (P_Q) with $Q = Q_i$ and let (D_{Q_i}) be the corresponding dual problem. We prove in Section 1 that

$$\min(\mathrm{P}_Q) \geqslant \min_{i \in I}\{\max(\mathrm{D}_{Q_i})\} \geqslant \max(\mathrm{D}_Q).$$

The right-hand side inequality is typically strict, in which case the duality gap (3) is strictly reduced by using the improved lower bound

$$L = \min_{i \in I}\{\max(\mathrm{D}_{Q_i})\}. \tag{4}$$

Later in the paper, we concentrate mainly on a special case of problem (P_Q), which is linear in x for fixed q:

$$\min_{x\in\mathbb{R}^n, q\in\mathbb{R}^p} \{c^{\mathrm{T}}x: A(q)x \leqslant b, x \geqslant 0, q\in Q\}. \tag{5}$$

Here Q is a polytope in $\mathbb{R}^p$, and for each $q\in Q$, $A(q)$ is an $m\times n$ real matrix. Such models arise in *optical design of water distribution networks* (see e.g. [1, 8]), *the pooling/blending problem in oil refinery* (see e.g. [4, 5]) and in fact, any other *bilinear programming problems.*

For problem (5), we show in Section 2 that for any prescribed tolerance $\varepsilon > 0$, if the partition $\{Q_i: i\in I\}$ is fine enough, then

$$0 < \min(\mathrm{P}_Q) - \min_{i\in I}\{\max(\mathrm{D}_{Q_i})\} < \varepsilon. \tag{6}$$

In Section 4, we outline branch and bound type algorithms, combined with a (nonsmooth) local search algorithm for problem (1), which finds a candidate solution $\bar{q}$, and which realizes the error estimate (6). In Section 5 we apply the algorithm to the Pooling Problem and give numerical results.

Of course, to compute the lower bound L in (4), the dual problem (D_{Q_i}) has to be solved for every $i\in I$, therefore, this must be an easy problem. For the primal problem (5), the dual problem is a *semi-infinite linear programming problem*

$$\max\{-b^{\mathrm{T}}y: A(q)^{\mathrm{T}}y + c \geqslant 0\ \forall q\in Q\} \tag{7}$$

clearly not an easy problem in general. However, in some important situations, which include the applications mentioned above (and in particular the bilinear case), we show in Section 3 that (7) reduces to a usual linear programming problem.

A related approach to the global optimization of problem (P_Q) can be found in the papers by Floudas and Visweswaran [6]. Visweswaran and Floudas [13] and Floudas and Visweswaran [7]. In these works, a decomposition approach is also used, and dual bounds are *estimated* via certain *relaxation* of the dual problem. The approach involves also some specific partitioning of the feasible set.

1. Reducing the duality gap

Let $\{Q_i: i\in I\}$ be a partition of Q, i.e.

$$\bigcup_{i\in I} Q_i = Q.$$

Corresponding to the ith part of Q, we have the primal problem

$$(\mathrm{P}_{Q_i}) \quad \min_{x,q}\{f_0(q, x): q\in Q_i, x\in X(q)\}$$

and the dual problem

$$(\mathrm{D}_{Q_i}) \quad \max_{y\geqslant 0}\left\{h_i(y) =: \min_{q\in Q_i,\, x} L(q, x, y)\right\}$$

where L is the Lagrangian (2).

We state now a simple but fundamental result showing that a reduction of the duality gap (3) can be achieved by partitioning.

Lemma 1.

$$\min(\mathrm{P}_Q) \geqslant \min_{i\in I}\{\max(\mathrm{D}_{Q_i})\} \geqslant \max(\mathrm{D}_Q). \tag{8}$$

Proof.

$$\min(\mathrm{P}_Q) = \min_{i\in I}\{\min(\mathrm{P}_{Q_i})\} \geqslant \min_{i\in I}\{\max(\mathrm{D}_{Q_i})\},$$

by weak duality for the pair (P_{Q_i}), (D_{Q_i}); hence the left-hand side inequality in (8) holds.

Now,

$$\max(\mathrm{D}_{Q_i}) = \max_{y\geqslant 0} h_i(y),$$

but, for every y,

$$h_i(y) = \min_{q\in Q_i, x} L(q, x, y) \geqslant \min_{q\in Q, x} L(q, x, y)$$

since $Q_i \subset Q$. Hence, for every $i\in I$,

$$\max(\mathrm{D}_{Q_i}) \geqslant \max_{y\geqslant 0}\ \min_{q\in Q, x} L(q, x, y) = \max(\mathrm{D}_Q)$$

proving that the right-hand side inequality in (8) holds. □

Example 1. The following is a reformulation of a well-known test problem in global optimization, the so-called *Haverly Pooling Problem – Case 1* (see e.g. [4; 5, Section 6] and explanation in Section 5, this paper).

$$(\mathrm{H}_Q) \quad \min_{q\in\mathbb{R}^2, x\in\mathbb{R}^6} \quad (-x_1 - x_2)$$

$$\begin{aligned}
& -x_1 - (6q_1 + 16q_2 - 9)x_3 - x_5 = 0,\\
& -x_2 - (6q_1 + 16q_2 - 15)x_4 + 5x_6 = 0,\\
& x_3 + x_5 \leqslant 100,\\
& x_4 + x_6 \leqslant 200,\\
& (3q_1 + q_2 - 2.5)x_3 - 0.5x_5 \leqslant 0,\\
& (3q_1 + q_2 - 1.5)x_4 + 0.5x_6 \leqslant 0,\\
& x_i \geqslant 0, \quad i = 1, 2, \ldots, 6,\\
& q \in Q =: \{(q_1, q_2) : q_1 + q_2 = 1,\ q_1 \geqslant 0,\ q_2 \geqslant 0\}.
\end{aligned} \tag{9}$$

The global minimum of (H_Q) is known [5, p. 59]:

$$\min(\mathrm{H}_Q) = -400.$$

Since (H_Q) is of the special form (5), its dual is of the form (7),which here becomes

$$(\mathrm{DH}_Q) \quad \max_{y \in \mathbb{R}^6} (-100y_3 - 200y_4)$$

$$\begin{aligned} & -y_1 - 1 \geqslant 0, \\ & -y_2 - 1 \geqslant 0, \end{aligned} \qquad (10)$$

$$\begin{aligned} & -(6q_1 + 16q_2 - 9)y_1 + y_3 + (3q_1 + q_2 - 2.5)y_5 \geqslant 0, \quad \forall q \in Q, \\ & -(6q_1 + 16q_2 - 15)y_2 + y_4 + (3q_1 + q_2 - 1.5)y_6 \geqslant 0, \quad \forall q \in Q, \end{aligned}$$

$$\begin{aligned} & -y_1 + y_3 - 0.5y_5 \geqslant 0, \\ & 5y_2 + y_4 + 0.5y_6 \geqslant 0, \end{aligned} \qquad (11)$$

$$y_3 \geqslant 0, \quad y_4 \geqslant 0, \quad y_5 \geqslant 0, \quad y_6 \geqslant 0.$$

In (10), (11), it suffices to consider instead of all $q \in Q$, just the extreme points of the unit simplex Q (see Section 3) and thus (DH_Q) is a linear program, and its optimal value is computed to be

$$\max(\mathrm{DH}_Q) = -500.$$

The duality gap is then

$$\min(\mathrm{H}_Q) - \max(\mathrm{DH}_Q) = (-400) - (-500) = 100.$$

We now consider a natural partition of the unit simplex (9):

$$Q = Q_1 \cup Q_2$$

where

$$Q_1 = \{(q_1, q_2) \colon q_1 + q_2 = 1,\ 0 \leqslant q_1 \leqslant 0.5,\ q_2 \geqslant 0\},$$

$$Q_2 = \{(q_1, q_2) \colon q_1 + q_2 = 1,\ 0.5 \leqslant q_1 \leqslant 1,\ q_2 \geqslant 0\}.$$

We obtain the following optimal values for the corresponding dual problems (for more details see Section 4):

$$\max(\mathrm{DH}_{Q_1}) = -400, \qquad \max(\mathrm{DH}_{Q_2}) = -100.$$

The dual bound L, for this partition, is then

$$L = \min_{i=1,2} \{\max(\mathrm{DH}_{Q_i})\} = \min(-400,\ -100) = -400$$

and the duality gap is reduced to zero,

$$\min(\mathrm{H}_Q) - L = 0.$$

2. Closing the duality gap

The duality gap

$$\min(\mathrm{P}_Q) - \min_{i\in I}\{\max(\mathrm{D}_{Q_i})\} \tag{12}$$

may be further reduced by repartitioning the sets Q_i $(i\in I)$. Can this process be continued until, for a partition which is fine enough, the gap becomes arbitrarily small? First, we have to quantify the concept of a "fine partition". Thus, let $\{Q_i: i\in I\}$ be a given partition, and for each $i\in I$, let t_i and q_i be the radius and center, respectively, of the smallest ball containing Q_i. We call

$$t=t(\{Q_i\})=\max\{t_i: i\in I\}$$

the *radius* of the partition $\{Q_i: i\in I\}$.

The hope that the duality gap can be made arbitrarily small is based on the *partial convexity* assumption mentioned in the Introduction:

(A1) $\forall q\in Q$, the functions $f_i(q,\cdot)$, $i=0, 1, \ldots, m$, are convex.

We also need a regularity condition:

(A2) $\forall q\in Q$, the feasible set $X(q)$ satisfies Slater's condition (see e.g. [10]).

Let the radius of partition reduce to zero, i.e. $t=0$ and the set Q_i is reduced to the singleton $\{q_i\}$. Then

$$\min(\mathrm{P}_Q) \geqslant \min_i \max(\mathrm{D}_{\{q_i\}}) = \min_i \min(\mathrm{P}_{\{q_i\}}) = \min(\mathrm{P}_Q).$$

The first equality holds, since by assumptions (A1), (A2), $(\mathrm{P}_{\{q_i\}})$ is a regular convex program, and hence, strong duality holds between it and its dual $(\mathrm{D}_{\{q_i\}})$. It follows that when $t=0$,

$$\min(\mathrm{P}_Q) = \min\{\max(\mathrm{D}_{Q_i})\}$$

and the duality gap is eliminated.

Of course, partitions with radius $t=0$ should be taken only in a "limiting sense", but, if one can show that the duality gap (12) as a function of the partition's radius t, is continuous at $t=0$, then one is assured that for any prescribed $\varepsilon>0$, there is a fine enough partition $\{Q_i: i\in I\}$ for which the duality gap (12) is at most ε. The rest of this section is devoted to this continuity question, and we restrict ourselves to the special case of problem (5). The primal problem is then "partially linear":

$$(\mathrm{PL}_Q) \quad \min_{q\in\mathbb{R}^p, x\in\mathbb{R}^n} \{c^{\mathrm{T}}x: A(q)x\leqslant b,\ x\geqslant 0,\ q\in Q\}$$

and we assume that:

(A3) Q is a polytope in $\mathbb{R}^p$, and for each $q \in Q$, $A(q)$ is an $m \times n$ matrix, whose ij-element $A_{ij}(q)$ is a continuous function on Q.

Note that for problem (PL_Q), Assumption (A1) holds and (A2) is superfluous. A dual problem for problem (PL_Q) is given by

$$\max_{y \geqslant 0} \min_{q \in Q, x \geqslant 0} \{c^{\mathrm{T}}x + y^{\mathrm{T}}(A(q)x - b)\}$$

which reduces to a semi-infinite linear program:

$$(\mathrm{DPL}_Q) \quad \max_{y \geqslant 0}\{-b^{\mathrm{T}}y \colon A(q)^{\mathrm{T}}(y) + c \geqslant 0 \quad \forall q \in Q\}.$$

Let $\{Q_i \colon i \in I\}$ be a partition of Q with radius t, and let the center of the circumscribing ball of Q_i be q_i. We need to study the continuity at $t = 0$ of the function

$$d_i(t) =: \max_{y \geqslant 0}\{-b^{\mathrm{T}}y, A(q)^{\mathrm{T}}(y) + c \geqslant 0 \ \forall q \in B_{q_i}(t)\} \tag{13}$$

where

$$B_{q_i}(t) =: \{q \in \mathbb{R}^p \colon \|q - q_i\| \leqslant t\}$$

is the ball of radius t with center at q_i. For fixed $q \in B_{q_i}(t)$, the jth constraint in (13) is

$$a_j(q)^{\mathrm{T}}y + c_j \geqslant 0$$

where $a_j(q)$ is the jth column of the matrix $A(q)$. The continuum of constraints in (13) can be equivalently expressed by the following n constraints

$$G_j^i(y, t) \geqslant 0, j = 1, \ldots, n$$

where the function $G_j^i \colon \mathbb{R}^m \times [0, \infty) \to \mathbb{R}$ is defined by

$$G_j^i(y, t) =: \min_q\{a_j(q)^{\mathrm{T}}y + c_j \colon q \in B_{q_i}(t)\}. \tag{14}$$

The function $d_i(t)$ is (13) is now given as an optimal value function of a mathematical programming problem:

$$d_i(t) = \max_{y \geqslant 0}\{-b^{\mathrm{T}}y \colon G_j^i(y, t) \geqslant 0, j = 1, \ldots, n\}. \tag{15}$$

To study the continuity properties of the functions $d_i(t)$, $i = 1, \ldots, m$, it suffices to consider the following function:

$$\psi(\lambda) =: \min_{y \geqslant 0}\{b^{\mathrm{T}}y \colon g_j(y, \lambda) \leqslant 0, j = 1, \ldots, n\} \tag{16}$$

where

$$g_j(y, \lambda) =: \max_q\{-a_j(q)^{\mathrm{T}}y - c_j \colon \|q - \bar{q}\| \leqslant \lambda\}. \tag{17}$$

Note that $g_j(\cdot, \lambda)$ is a closed convex function in y ($\forall \lambda \geqslant 0$) since it is the pointwise maximum of affine functions.

In what follows, we shall use various continuity concepts of single and multivalued mappings. The definitions and several results we employ are taken from the book by Bank et al. [2].

We next prove:

Lemma 2. *g_j is lower-semicontinuous (l.s.c.) on $\mathbb{R}^m \times \{0\}$.*

Proof. Let $y \in \mathbb{R}^m$ be fixed and consider sequences $y_n \in \mathbb{R}^m$, $\lambda_n \geqslant 0$ converging to y and 0 respectively.

Now,

$$\begin{aligned} g_j(y_n, \lambda_n) &= \sup\{-a_j(q)^{\mathrm{T}} y_n - c_j \colon \|q-\bar{q}\| \leqslant \lambda\} \\ &\geqslant -a_j(\bar{q})^{\mathrm{T}} y_n - c_j, \end{aligned}$$

hence

$$\begin{aligned} \lim_{n\to\infty} g_j(y_n, \lambda_n) &\geqslant \lim_{n\to\infty} (-a_j(\bar{q})^{\mathrm{T}} y_n - c_j) \\ &= -a_j(\bar{q})^{\mathrm{T}} y - c_j = g_j(y, 0) \end{aligned}$$

proving the claimed l.s.c. property. $\square$

Let us denote the feasible set in (16) by

$$M(\lambda) =: \{y \in \mathbb{R}^m \colon g_j(y, \lambda) \leqslant 0, j = 1, \ldots, n\}.$$

In particular,

$$M(0) = \{y \colon a_j(\bar{q})^{\mathrm{T}} y + c_j \geqslant 0, j = 1, \ldots, n\}.$$

We further denote

$$M^\circ(\lambda) = \{y \in \mathbb{R}^m \colon g_j(y, \lambda) < 0, j = 1, \ldots, n\}.$$

In particular,

$$M^\circ(0) = \{y \in \mathbb{R}^n \colon a_j(\bar{q}) y + c_j > 0, j = 1, \ldots, n\}.$$

The nonnegative orthant is denoted by Γ,

$$\Gamma = \{y \in \mathbb{R}^m \colon y \geqslant 0\}.$$

Lemma 3. *Assume that $M^\circ(0) \neq \emptyset$, then g_j is upper-semicontinuous (u.s.c.) on $M^\circ(0) \times \{0\}$.*

Proof. Since $g_j(\cdot, \lambda)$ is convex ($\forall \lambda \geqslant 0$) it is continuous on the open set $M^\circ(0)$ and so

$$\limsup_{y_n \to y_0, \lambda_n \to 0^+} g_j(y_n, \lambda_n = \limsup_{\lambda_n \to 0^+} g_j(y_0, \lambda_n) \quad \forall y_0 \in M^\circ(0). \tag{18}$$

We show next that

$$g_j(y_0, \cdot) \text{ is u.s.c. at } \lambda = 0. \tag{19}$$

Set

$$\rho(\lambda) =: g_j(y_0, \lambda) = \sup\{-a_j(q)^{\mathrm{T}} y_0 - c_j \colon \|q - \bar{q}\| \leqslant \lambda\}. \tag{20}$$

By [2, Theorem 4.2.2 (1)] ρ is u.s.c. at $\lambda = 0$ since the objective function is (20) is continuous and the feasible set is clearly l.s.c.-B in λ. By (18) and (19),

$$\limsup g_j(y_n, \lambda_n) \leqslant g_j(y_0, 0) \quad \forall y_0 \in M^\circ(0)$$

proving that g_j is u.s.c. on $M^\circ(0) \times \{0\}$. □

Lemma 4. *The multivalued mapping $M(\lambda) \cap \Gamma$ is closed at $\lambda = 0$.*

Proof. Γ is a closed set and the functions g_j are l.s.c. on $\mathbb{R}^m \times \{0\}$ by Lemma 1. Hence by [2, Theorem 3.1.1], $M \cap \Gamma$ is closed at $\lambda = 0$. □

Lemma 5. *If $M^\circ(0) \cap \Gamma \neq \emptyset$ then $M(\lambda) \cap \Gamma$ is l.s.c.-B at $\lambda = 0$.*

Proof. By Lemma 2, g_j are u.s.c. on $M^\circ(0) \times \{0\}$. Moreover, Γ is convex, $g_j(\cdot, 0)$ are convex and by assumption $M^\circ(0) \cap \Gamma \neq \emptyset$. It follows from [2, Theorem 3.1.6] that $M \cap \Gamma$ is l.s.c.-B at $\lambda = 0$. □

Combining the above results, we have the desired continuity of the function $\psi(\lambda)$ defined in (16).

Theorem 1. *If the following regularity condition holds:*

$$\exists y \geqslant 0 \quad \textit{such that} \quad a_j(\bar{q})y + c_j > 0 \quad \forall j = 1, \ldots, n, \tag{21}$$

then $\psi(\lambda)$ is continuous at $\lambda = 0$.

Proof. Note that according to our notation

$$\psi(\lambda) = \min\{b^{\mathrm{T}} y \colon y \in M(\lambda) \cap \Gamma\}.$$

We showed in Lemma 5 that

(i) $M(\lambda) \cap \Gamma$ is l.s.c.-B at $\lambda = 0$.

Also, the following is clear:

(ii) $M(0)$ is a closed convex set (in fact a polyhedral set)

(iii) $M(\lambda) \cap \Gamma$ is a convex set $\forall \lambda \geqslant 0$ (by the convexity of $g(\cdot, \lambda)$)

(iv) $M^\circ(0) \cap \Gamma \neq \emptyset$ (This is precisely the regularity condition (21).)

(v) $M \cap \Gamma$ is closed at $\lambda = 0$ (by Lemma 4).

By [2, Theorem 4.3.3] the conditions (i)–(v) suffice to ensure the continuity of ψ at $\lambda = 0$. □

Comparing the functions $d_i(t)$ in (15) to $\psi(\lambda)$ in (16), we see from Theorem 1, that to ensure continuity of $d_i(t)$ at $t=0$ for all $i\in I$, it suffices to impose the following "Slater condition":

(A4) $\forall q\in Q$, there exists $y\geqslant 0$ such that

$$A(q)^{\mathrm{T}}y+c>0.$$

We state now the main result showing that the duality gap can be made arbitrarily small if the partition is fine enough.

Theorem 2. *Consider the problem* (PL_Q) *and suppose assumption* (A3) *and* (A4) *hold. Then, for every* $\varepsilon>0$, *there exists a partition* $\{Q_i^{\varepsilon}: i\in I\}$ *with radius* $t_{\varepsilon}>0$, *such that the duality gap*

$$G(t_\varepsilon)=\min(\mathrm{PL}_Q)-\min_{i\in I}\max(\mathrm{DPL}_{Q_i^{\varepsilon}}) \tag{22}$$

is smaller than ε, *i.e.*

$$0\leqslant G(t_\varepsilon)<\varepsilon.$$

Proof. For any partition $\{Q_i: i\in I\}$ with radius $t>0$,

$$Q_i\subset B_{q_i}(t)\quad \forall i\in I,$$

and so

$$\max(\mathrm{DLP}_{Q_i})\geqslant d_i(t)\quad \forall i\in I.$$

Hence, the duality gap $G(t)$ based on this partition satisfies

$$G(t)\leqslant \min(\mathrm{PL}_Q)-\min_{i\in I}\{d_i(t)\}. \tag{23}$$

Denote the right-hand side of (23) by $H(t)$. Under the assumptions (A3) and (A4) it follows from Theorem 1 that

$$H(t)\ \text{ is continuous at }\ t=0. \tag{24}$$

If $t=0$, then

$$Q_i=\{q_i\}=B_{q_i}(0)$$

and so equality holds in (23), i.e. $G(0)=H(0)$. Moreover, as previously discussed, $G(0)=0$, hence

$$G(0)=H(0)=0. \tag{25}$$

It follows from (24) and (25), that for every $\varepsilon>0$, we can find a partition with a small enough radius $t_\varepsilon>0$ such that

$$H(t_\varepsilon) < \varepsilon$$

and so, by (23),

$$G(t_\varepsilon) < \varepsilon. \qquad \square$$

3. Solving the dual problem

To generate a lower bound for the optimal value $\min(\mathrm{PL}_Q)$, based on a partition $\{Q_i: i \in I\}$, a dual problem has to be solved for each $i \in I$:

$$(\mathrm{DPL}_{Q_i}) \quad \max_{y \geq 0}\{-b^{\mathrm{T}}y: a_j(q)^{\mathrm{T}}y + c_j \geq 0 \ \forall q \in Q_i, j = 1, \ldots, n\}.$$

This semi-infinite linear program can be greatly simplified in many important situations covered by the following result.

Theorem 3. *Under assumption* (A3), *if, for each* $j = 1, \ldots, n$, *the* j*th constraint function*

$$v_j(q) = a_j(q)^{\mathrm{T}}y + c_j$$

is quasi-concave and Q_i *is a polytope, then* (DPL_{Q_i}) *is equivalent to the following finitely constrained linear program*

$$\max_{y \geq 0}\{-b^{\mathrm{T}}y: a_j(q^k)^{\mathrm{T}}y + c_j \geq 0, j = 1, \ldots, n, k = 1, \ldots, K\}$$

where $\{q^1, q^2, \ldots, q^K\}$ *are the extreme points of* Q_i.

Proof. The j constraint of (DPL_{Q_i}) can be written as

$$\min_{q \in Q_i} v_j(q) \geq 0.$$

Since $v_j(q)$ is quasi-concave, the minimum is attained at an extreme point of Q_i. The latter fact is well known for concave function; we give a proof for the quasi-concave case in Lemma 6 below. $\square$

Lemma 6. *Let* $S \subset \mathbb{R}^n$ *be a compact convex set and* $v: \mathbb{R}^n \to R$ *be a quasi-concave function on* S. *Then*

$$\min_{q \in S} v(q) = \min_{q \in \operatorname{ext} S} v(q)$$

where ext S *denotes the set of extreme points of* S.

Proof.

$$
\begin{aligned}
\min_{q\in \text{ext } S} v(q) &\geqslant \min_{q\in S} v(q) \\
&= \min_{q\in \text{conv}(\text{ext } S)} v(q) \\
&\quad \text{(by the Krein–Milman Theorem [10, Section 18])} \\
&\geqslant v(\bar{q}) \quad (\text{where } \bar{q}\in \arg\min\{v(q)\colon q\in \text{conv}(\text{ext } S)\}) \\
&= v\left(\sum_{i=1}^{n+1} \lambda_i q^i\right) \quad (\text{for some } \lambda_i \geqslant 0,\ \Sigma\lambda_i = 1 \text{ and} \\
&\qquad\qquad q^i \in \text{ext } S, \text{ by Caratheodory Theorem}) \\
&\geqslant \min_{i=1,\ldots,n+1} v(q^i) \quad (\text{by the quasi-concavity of } v\text{; see [9]}) \\
&\geqslant \min_{q\in \text{ext } S} v(q). \qquad \square
\end{aligned}
$$

Example 2. For bilinearly constrained problems, the functions $v_j(q)$ are all linear in q, hence quasi-concave. Moreover, for the pooling/blending problem Q is the unit simplex so the extreme points are the unit vectors.

Example 3. For the Water Distribution Network problem [1, 3, 8], the function $v_j(q)$ is of the form

$$v_j(q) = k_i \,\text{sign}(q_j)\,|q_j|^{1.852}\cdot w_j(y) + c_j$$

where $k_i > 0$, and $w_j(y)$ is a linear function of y.

It is easily seen that v_j depends only on the single variable q_j and it is monotone (increasing or decreasing, depending on the sign of $w_j(y)$) in this variable. Thus, $v_j(\cdot)$ is quasi-linear and so quasi-concave. Moreover, for each $j = 1, \ldots, n$, the continuum of constraint $v_j(q) \geqslant 0$, $q\in Q$ can be replaced by just two constraints

$$v_j(q_j^{\max}) \geqslant 0, \qquad v_j(q_j^{\min}) \geqslant 0,$$

where

$$q_j^{\max} = \max\{q_j\colon q\in Q\}, \qquad q_j^{\min} = \min\{q_j\colon q\in Q\},$$

since Q is typically a "box", the numbers $q_j^{\max}$, $q_j^{\min}$ are trivially found.

Remark. Simple finitely-constrained dual problems can be obtained for problems more general than the partial linear problem (PL_Q), e.g. problems where also the objective function may be bilinear in the (x, q) variables.

4. An algorithm for finding the global minimum of problem (PL_Q)

We describe here a "Branch and Bound" type algorithm, based on the Partition Principle (Lemma 1) which produces a candidate primal global solution for problem (PL_Q) whose optimal value is at most ε larger than the true global value.

The algorithm starts with an initial estimate $f_{\min}$ of $\min(PL_Q)$ (found e.g. by a local search), it then computes the dual bound $g_Q = \max(DPL_Q)$.

If $f_{\min} - g_Q \leqslant \varepsilon$ then the algorithm stops; otherwise Q is partitioned into sub-polytopes $\{Q_i\}$. The dual problem (PL_{Q_i}) is solved for each i, and the part (denoted Q_{i*}) producing $\min_i g_{Q_i}$ is recorded. A new estimate $f_{\min}$ is computed by a local search from the center of Q_{i*}. If $f_{\min} - g_{Q_{i*}} > \varepsilon$, then we branch by partitioning Q_{i*} etc. After l successive branching levels, we have a current polytope Q^l and the algorithm continues as follows:

Step 1. Partition Q^l into $\{Q_i^l\colon i=1, \ldots, k\}$, and compute $g_i^l = \max(DPL_{Q_i^l})$ for $i = 1, \ldots, k$.

Step 2. Find $i_l \in \arg\min_i g_i^l$, compute a new estimate of the primal value $f^{\,l}_{\min}$ by (locally) minimizing $\varphi(q)$, starting from the center of Q_{i_l}.

Step 3. (Check Branch or Bound).

if $f^{\,l}_{\min} - g^l_{i_l} \leqslant \varepsilon$

then (BOUND): if $l=0$ then stop else go to Step 4

else (BRANCH): go to Step 5.

Step 4. (BOUND) $g^{l-1}_{i_{l-1}} \leftarrow g^l_{i_l}$;

$l \leftarrow l-1$;

go to Step 2.

Step 5. (BRANCH) $l \leftarrow l+1$;

$Q^l \leftarrow Q^l_{i_l}$;

go to Step 1.

Note. The algorithm stops by returning to level 0 with the stopping criteria satisfied. At this point, $g^0_{i_0}$ is a lower bound for problem (PL_Q) which is at most ε smaller than its global value.

Example 4. We demonstrate the algorithm by reexamining the Haverly Pooling Problem (Example 1). We start by obtaining an initial estimate $f_{\min}$ by making a local search from the point $q^\circ = (\frac{1}{2}, \frac{1}{2})$ which is the center of the polytope

$$Q = \{(q_1, q_2)\colon q_1 + q_2 = 1,\ q_1 \geqslant 0,\ q_2 \geqslant 0\}.$$

It turns out that q° itself is a local minimum and so

$$f_{\min} = \varphi(q^\circ) = 0.$$

Recall that here $\varphi(\cdot)$ is the optimal value of a linear problem

$$\varphi(q) =: \min_x \{c^{\mathrm{T}}x\colon A(q)x \leqslant b,\ x \geqslant 0\}.$$

The dual bound was computed in Example 1:

$$\max(\mathrm{DPL}_Q) = -500.$$

We now partition Q as in Example 1: $Q = Q_1 \cup Q_2$ where

$$Q_1 = \{(q_1, q_2)\colon q_1 + q_2 = 1,\ 0 \leqslant q_1 \leqslant 0.5,\ q_2 \geqslant 0\},$$

$$Q_2 = \{(q_1, q_2)\colon q_1 + q_2 = 1,\ 0.5 \leqslant q_1 \leqslant 1,\ q_2 \geqslant 0\}.$$

The dual problem (DPL_{Q_1}) is here the problem (DH_{Q_1}), as in Example 1, Section 1, but in the constraints (10), (11) the only values of $q = (q_1, q_2)$ which need to be considered are

$$(q_1, q_2) = [\tfrac{1}{2}, \tfrac{1}{2}], \qquad (q_1, q_2) = (0, 1).$$

For the second dual problems (DPL_{Q_2}) the values for q are

$$(q_1, q_2) = [\tfrac{1}{2}, \tfrac{1}{2}], \qquad (q_1, q_2) = (1, 0).$$

The optimal solution of problem (DPL_{Q_1}) is

$$y = (-1, -1, 0, 2, 0, 6)$$

and of problem (DPL_{Q_2})

$$y = (-1, -1, 1, 0, 4, 10),$$

and the corresponding dual bounds are

$$\max(\mathrm{DPL}_{Q_1}) = -400, \qquad \max(\mathrm{DPL}_{Q_2}) = -100.$$

The value of the lower bound is now $L = \min(-400, -100) = -400$, and it is obtained for Q_1, so a local search is performed from the starting point $q^1 = (\frac{1}{4}, \frac{3}{4})$ which is the center of Q_1. The corresponding value at this point is

$$\varphi(q^1) = -300.$$

This point is not a local minimum, i.e.

$$0 \notin \partial\varphi(q^1).$$

A Bundle–Trust algorithm was employed and terminated at a local solution

$$q^* = (0, 1)$$

with a corresponding new estimate of the global solution

$$f_{\min} = \varphi(q^*) = -400.$$

Thus $f_{min} = L$ and the global solution is indeed q^*.

5. Application — the pooling problem in an oil refinery

In oil refineries, a two-stage process is used to form final products from given oil components. In the first stage, ''intermediate'' products are obtained by combining the components into special tanks, named ''pools''. In the second stage, these intermediate products are combined to form final products of prescribed quantities and qualities. A certain quantity of an initial component can be combined directly in the second stage. Given prices of the components and the products, the total profit is to be maximized.

The mathematical model of the pooling problem is as follows.

Let the following finite sets represent:

$\{1, 2, \ldots, i, \ldots, I\}$ – components;

$\{1, 2, \ldots, j, \ldots, J\}$ – products;

$\{1, 2, \ldots, l, \ldots, L\}$ – pools;

$\{1, 2, \ldots, k, \ldots, K\}$ – qualities;

the following variables represent:

x_{il} – amount of component i allocated to pool l;

y_{lj} – amount going from pool l to product j;

z_{ij} – amount of component i going directly to product j;

p_{lk} – level of quality k in pool l;

and the following input parameters represent:

A_i – upper bounds for component availabilities;

D_j – upper bounds for product demands;

S_l – upper bounds for pool sizes;

P_{jk} – upper bounds for product qualities;

C_{ik} – level of quality k in component i;

c_i – unit price of component i;

d_j – unit price of product j.

The formulation of the pooling problem in terms of the above list is:

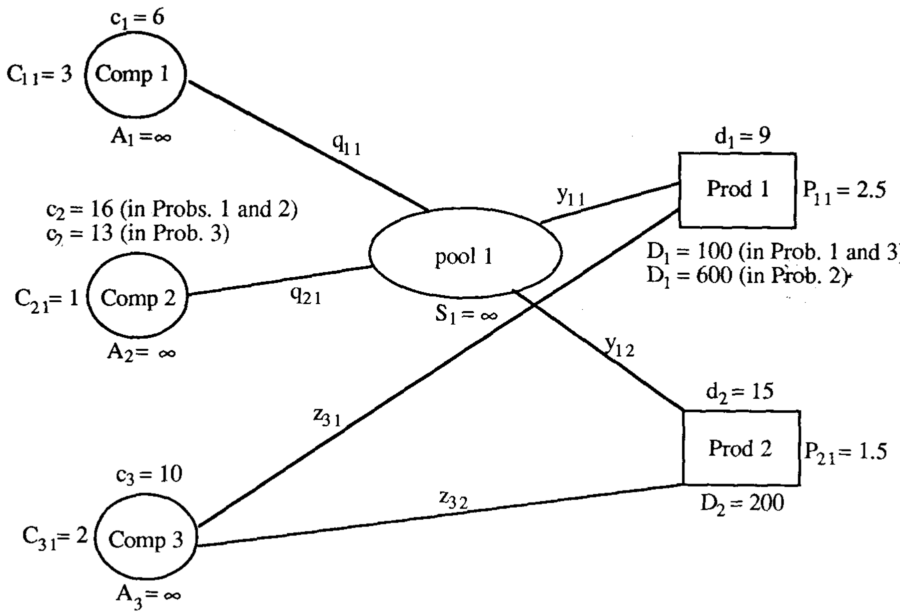

Fig. 1. Problems 1, 2 and 3 (three cases of the Haverly Pooling Problem).

$$\max \left\{ -\sum_i \sum_l c_i x_{il} + \sum_l \sum_j d_j y_{lj} + \sum_i \sum_j (d_j - c_i) z_{ij} \right\}$$

subject to

$$\sum_l x_{il} + \sum_j z_{ij} \leqslant A_i$$

$$\sum_i x_{il} - \sum_j y_{lj} = 0, \qquad (26)$$

$$\sum_i x_{il} \leqslant S_l,$$

$$\sum_l y_{lj} + \sum_i z_{ij} \leqslant D_j,$$

$$\sum_l (p_{lk} - P_{jk}) y_{lj} + \sum_i (C_{ij} - P_{jk}) z_{ij} \leqslant 0, \qquad 19,,,R,I0$$

$$p_{lk} \geqslant 0, \quad x_{il} \geqslant 0, \quad y_{lj} \geqslant 0, \quad z_{ij} \geqslant 0.$$

It is obvious that the variables *lk* belong to the "box"

$$P = \{p_{lk} : \underline{p}_{lk} \leqslant p_{lk} \leqslant \bar{p}_{lk}\} \qquad (28)$$

where

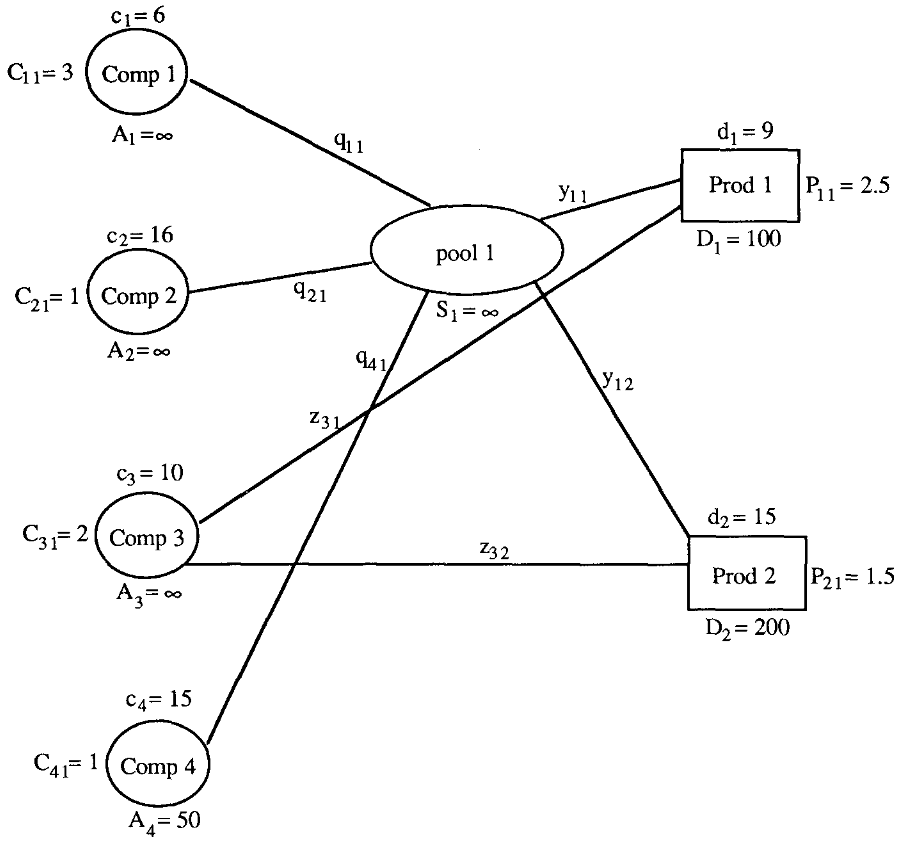

Fig. 2. Problem 4.

$$\underline{p}_{lk} = \min_i C_{ik}, \qquad \bar{p}_{lk} = \max_i C_{ik}.$$

The formulation is referred to as the "p-formulation" of the pooling problem.

Here we use another formulation which can be obtained by introducing new variables q_{il} according to the relations

$$x_{il} = q_{il} \sum_j y_{lj}.$$

Equation (27) gives then

$$p_{lk} = \sum_i C_{ik} q_{il}.$$

It follows from the definition of q_{il} and equations (26) that the variables q_{il} belong to the "simplex"

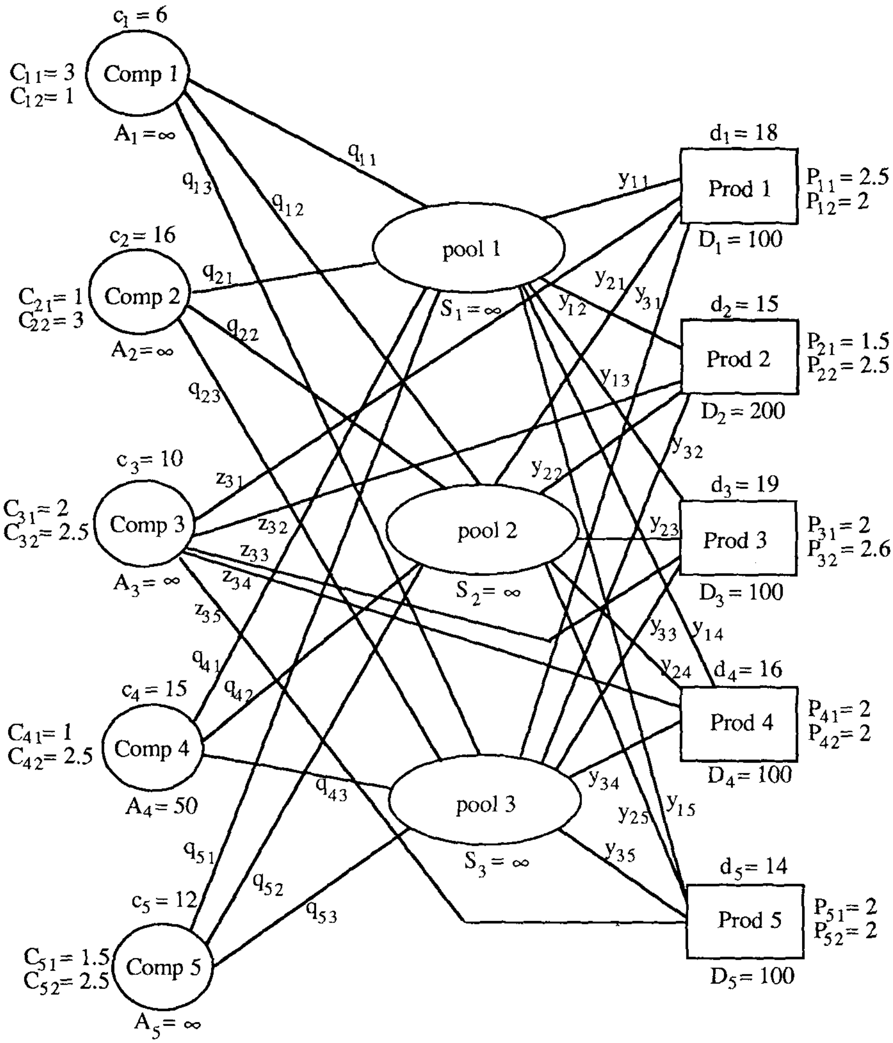

Fig. 3. Problem 5.

$$Q = \left\{ q_{il}:\ q_{il} \geqslant 0,\ \sum_i q_{il} = 1 \right\}. \tag{29}$$

The new "q-formulation" is

$$\min - \sum_j \gamma_j$$

subject to

Table 1

Dimensions and solutions for Problems 1–5

Problem #		1	2	3	4	5
# outer variables q		2	2	2	3	12
# inner variables (γ, y, z)		6	6	6	6	25
# constraints		6	6	6	7	20
# variables in dual		6	6	6	7	20
# constraints in dual		8	8	8	10	70
initial solution	q	$(\frac{1}{3}, \frac{2}{3})$	$(\frac{3}{4}, \frac{1}{4})$	$(0.42, 0.58)$	$(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$	$(0.31, 0, 0, , 0.69, 0, 0, 0, 1, 0.25, 0.25, 0.25, 0.25)$
	objective value	0	−300	0	0	−1500
	lower bound	−500	−1000	−875	−550	−3500
final solution	q	$(0, 1)$	$(\frac{1}{2}, \frac{1}{2})$	$(\frac{1}{4}, \frac{3}{4})$	$(0, \frac{1}{2}, \frac{1}{2})$	$(0, 0, 0, 1, 0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 1, 0, 0, 0)$
	objective value	−400	−600	−750	−450	−3500
	lower bound	−400	−600	−750	−450	−3500
# dual problems solved		3	3	3	25	283

$$-\gamma_j + \sum_l \left(d_j - \sum_i c_i q_{il}\right) y_{lj} + \sum_i (d_j - c_i) z_{ij} = 0,$$

$$\sum_l \sum_j q_{il} y_{lj} + \sum_j z_{ij} \leqslant A_i,$$

$$\sum_j y_{lj} \leqslant S_l,$$

$$\sum_l y_{lj} + \sum_i z_{ij} \leqslant D_j,$$

$$\sum_l \left(\sum_i C_{ik} q_{il} - P_{jk}\right) y_{lj} + \sum_i (C_{ik} - P_{jk}) z_{ij} \leqslant 0,$$

$$q_{il} \geqslant 0, \quad \sum_i q_{il} = 1, \quad y_{lj} \geqslant 0, \quad z_{ij} \geqslant 0.$$

Note that the problem is now in the form of the partially linear problem (PL_Q) of Section 2.

We use the q-formulation rather than the p-formulation because typically in applications the number of extreme points of the "simplex" Q (see (29)) is much smaller than that of the "box" P (see (28)). Consequently, the dual problems corresponding to the q-formulation are of much smaller size.

Figures 1–3 represent graphically five different pooling problems with numerical data and variables corresponding to the q-formulation.

Figure 1 represents three different cases of the Haverly Pooling Problem (see e.g. [5, Section 6]). Figures 2 and 3 represent two additional problems of a larger size. The results of solving problems 1–5 by the algorithm described in Section 4 are reported in Table 1.

References

[1] E. Alperovits and U. Shamir, "Design of optimal water distribution system," *Water Resources Research* 13(6) (1977) 885–900.

[2] B. Bank, J. Guddat, D. Klatte, B. Kummer and K. Tammer, *Nonlinear Parametric Optimization* (Birkhauser, Basel, 1983).

[3] A. Ben-Tal, G. Eiger, J. Outrata and J. Zowe, "A nondifferential approach to decomposable optimization problems with an application to the design of water distribution networks," in: W. Oettlie and D. Pallaschke, eds., *Advances in Optimization, Lecture Notes in Econom. and Math. Syst. No. 382* (Springer, Berlin, 1992).

[4] C.A. Floudas and A. Aggarwal, "A decomposition approach for global optimum search in the pooling problem," *ORSA Journal on Computing* 2(3) (1990) 225–235.

[5] C.A. Floudas and P.M. Pardalos, "A collection of test problems for constrained global optimization algorithms," *Lecture Notes in Comput. Sci. No. 455* (Springer, Berlin, 1987).

[6] C.A. Floudas and V. Visweswaran, "A global optimization algorithm (GOP) for certain classes of nonconvex NLPs: I. Theory," *Computers and Chemical Engineering* 14(12) (1990) 1397–1417.

[7] C.A. Floudas and V. Visweswaran, "A primal-relaxed dual global optimization approach," to appear in: *Journal of Optimization Theory and Applications* (1992).

[8] O. Fujiwara and D.B. Khang, "A two-phase decomposition method for optimal design of looped water distribution networks," *Water Resources Research* 26(4) (1990) 539–549.

[9] O.L. Mangasarian, *Nonlinear Programming* (McGraw-Hill, New Jersey, 1969).

[10] R.T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, NJ, 1972).

[11] H. Schramm and J. Zowe, "A version of the bundle idea for minimizing a nonsmooth function: Conceptual idea, convergence analysis, numerical results," *SIAM Journal on Optimization* 2 (1992).

[12] N.Z. Shor, *Minimization Methods for Nondifferentiable Functions* (Springer, Berlin, 1985).

[13] V. Visweswaran and C.A. Floudas, "A global optimization algorithm (GOP) for certain classes of nonconvex NLPs: II. Application of theory and test problems," *Computers and Chemical Engineering* 14(12) (1990) 1417–1434.