Algorithmic Operations Research

Utilizing the Surrogate Dual Bound in Capacity Planning with Economies of Scale

Han-suk Sohn et Dennis L. Bricker

Volume 3, numéro 1, winter 2008

URI : https://id.erudit.org/iderudit/aor3_1art07

Aller au sommaire du numéro

Éditeur(s)

Preeminent Academic Facets Inc.

ISSN

1718-3235 (numérique)

Découvrir la revue

Citer cet article

Sohn, H.-s. & Bricker, D. L. (2008). Utilizing the Surrogate Dual Bound in Capacity Planning with Economies of Scale. *Algorithmic Operations Research*, *3*(1), 67–78.

Résumé de l'article

Minimizing a nondecreasing separable concave cost function over a polyhedral set arises in capacity planning problems where economies of scale and fixed costs are significant, as well as production planning when a learning effect results in decreasing marginal costs. This is an NP-hard combinatorial problem in which the extreme points of the polyhedral set must be enumerated, each of them a local optimum. Branch-and-bound methods have been frequently used to solve these problems. Although it has been shown that in general the bound provided by the surrogate dual is tighter than that of the Lagrangian dual, the latter has generally been preferred because of the apparent computational intractability of the surrogate dual problem. In this paper we describe a branch-and-bound algorithm that exploits the superior surrogate dual bound in a branch-and-bound algorithm without explicitly solving the dual problem. This is accomplished by determining the feasibility of a set of linear inequalities.

Algorithmic Operations Research Vol.3 (2008) 67–78

Utilizing the Surrogate Dual Bound in Capacity Planning with Economies of Scale

Han-suk Sohn

Dept. of Industrial Engineering, New Mexico State University, Las Cruces, NM 88003, USA

Dennis L. Bricker

Dept. of Mechanical and Industrial Engineering, The University of Iowa, Iowa City, IA 52242, USA

Abstract

Minimizing a nondecreasing separable concave cost function over a polyhedral set arises in capacity planning problems where economies of scale and fixed costs are significant, as well as production planning when a learning effect results in decreasing marginal costs. This is an NP-hard combinatorial problem in which the extreme points of the polyhedral set must be enumerated, each of them a local optimum. Branch-and-bound methods have been frequently used to solve these problems. Although it has been shown that in general the bound provided by the surrogate dual is tighter than that of the Lagrangian dual, the latter has generally been preferred because of the apparent computational intractability of the surrogate dual problem. In this paper we describe a branch-and-bound algorithm that exploits the superior surrogate dual bound in a branch-and-bound algorithm without explicitly solving the dual problem. This is accomplished by determining the feasibility of a set of linear inequalities.

Key words: Surrogate Dual, Capacity Planning, Economies of Scale, Nonconvex, Lagrangian Dual, Branch-and-Bound

1. Introduction

Capacity planning is the first, and perhaps the most fundamental, problem of production management. The algorithm proposed and demonstrated here is intended to assist the engineer responsible for the design of a production facility or system of production facilities. Taking into full account the fixed costs of construction and installation, economies of scale, and a matrix specifying inputs and outputs per unit of capacity, the algorithm selects from a set of available process units a subset of units and their capacities which minimize total capital investment costs and meets any requirements on system output and limits on system input.

$$\text{Min} \sum_{j=1}^{n} f_j\left(x_j\right) \tag{1}$$

$$\text{s.t.} \sum_{j=1}^{n} a_{ij}x_j \geq b_i, i = 1, \ldots m \tag{2}$$

$$x_j \geq 0, j = 1, \ldots n \tag{3}$$

where x_j is the capacity of process j

a_{ij} is the input-output coefficient of process j (if positive, a_{ij} is the output of product i per unit capacity of process j; if negative, the input of product i per unit capacity of process j).

b_i is the required net output of product i if positive, or if negative, (the negative of) the limit on resource i (labor, raw material, space, etc.)

f_j is the investment cost function of process j:

$$f_j\left(x_j\right) = \begin{cases} 0 \text{ , if } x_j = 0 \\ \phi_j + \alpha_j x_j^{\beta_j} \text{ , if } x_j > 0 \end{cases} \tag{4}$$

where ϕ_j is a nonnegative fixed cost of installing process j and the log-linear function $\alpha_j x_j^{\beta_j}$ is the variable cost ($0 \leq \beta_j \leq 1$). A similar model arises in production planning when a learning effect results in decreasing marginal costs. (See, for example, Hax and Majluf [14] and Yelle [27].) This problem subsumes the linear fixed-charge problem (when β=1) or the pure fixed charge problem (β=0).

This problem is essentially a combinatorial one, in that its solution must be one of the finite number (viz., $\binom{m+n}{m}$) of basic solutions of the system of linear constraints. That is, the solution is determined by the

Email: Han-suk Sohn [hsohn@nmsu.edu], Dennis L. Bricker [dennis-bricker@uiowa.edu].

choice of processes to be included in the production facility, together with a choice of those input/output requirements to be slack. These choices determine a basis and thereby a basic solution with its process capacities specified.

2. The Nonconvex Problem

The capacity planning problem described above belongs to a more general class of optimal resource allocation problems:

$$P: \text{ Find } \Phi = \text{Min}\{f(x) : Ax \geq b, x \in C\} \quad (5)$$

where $A \in R^{m\times n}$, $b \in R^m$, C is a closed cone within R^n_+ (the nonnegative orthant), and the cost function f : R^n_+ has the properties:

i) f is explicitly quasiconcave on C, i.e., f is quasiconcave and for all $x_1 \in C$, $x_2 \in C$, and $x_0 = \lambda x_1 + (1-\lambda)x_2$ where $0 < \lambda < 1$, $f(x_1) > f(x_2)$ implies that $f(x_0) > f(x_2) = \text{Min}\{f(x_1), f(x_2)\}$. (In particular, concavity of f implies explicit quasiconcavity.)

ii) f is lower semicontinuous on C, i.e., the level set $\{x \in C : f(x) \leq k\}$ is closed for all k.

iii) f is isotone on C, i.e., $x_1 \leq x_2$ implies that $f(x_1) \leq f(x_2)$.

(1) f is homogeneous, i.e., $f(0) = 0$.

The Cauchy-Weierstrass existence theorem, given properties (ii) and (iii), assures us that problem P, if feasible, assumes its minimum value in the feasible region. If P is infeasible, then by adopting the convention that the minimum of the empty set is +∞, we obtain a well-defined minimum. After discussing the more general case, we shall in later sections present results for the case in which C is a closed polyhedral cone, assuring us that the minimum of f in problem P is attained at a vertex of the feasible region.

A large number of resource allocation problems in which economies of scale are significant are included in this general formulation, including many in logistics, such as facility location, production planning, transportation planning, as well as plant capacity planning.

The problem of optimally allocating resources when economies of scale are prevalent is fraught with pitfalls, in that mathematical programming algorithms will quite often converge to local but not global solutions. McCormick [20] surveyed methods proposed at that time for obtaining globally optimal solutions to nonconvex problems, and concluded:

"The branch-and-bound approach, relying on the use of underestimating convex functions, seems the most reasonable approach at this time. The efficiency it offers depends upon how quickly the regions which do not contain the global solution are eliminated."

Since that time, several other classes of algorithms for problem P have been proposed, including extreme point ranking, concavity cut reduction, and outer approximation. Benson [2] states that

"In recent years, probably the most popular technique used in algorithms for problem P is branch-and-bound."

In addition to the deterministic algorithms which have been mentioned above, many stochastic algorithms have been proposed for global optimization. (See Bomze [4].) We will restrict the discussion to the class of branch-and-bound algorithms, of which our algorithm is a member.

In general, branching consists of partitioning the set C (more generally not a cone, and usually a hyperrectangle) into disjoint subsets $\{C_k : k = 1, 2, ..., K\}$ and the associated subproblems

$$P_k: \text{ Find } \Phi_k = \text{Min}\{f(x) : Ax \geq b, x \in C_k\} \quad (6)$$

are either solved, or else fathomed by demonstrating that Φ_k exceeds an incumbent value, i.e., a value known to be attainable.

Falk [8] developed much of the theoretical basis for this approach, assuming C is compact, while Falk and Soland [11] and Jones and Soland [17] describe implementations when f is a piecewise-linear separable function and the C_k's are closed hyperrectangles. A lower bound on Φ_k is obtained by replacing f by its convex envelope over the feasible region. This approach has been applied to more structured problems in time-cost tradeoffs in project scheduling [10], in transportation planning [22], and in facility location [23].

Another branch and bound approach (cf. Carillo [7] and Horst [16]) assumes that C is a simplex, and branches by creating simplex-partitions $\{C_k\}$. Lower bounds are again obtained by use of the convex envelope of f taken over C_k, which is represented in terms of the vertices of C_k.

Making use of the fact that, when f is concave, the global optimizer of P also optimizes the convex envelope of f taken over the feasible region, Falk and Hoffman [9] presented an algorithm which represents both the feasible region (assumed to be polyhedral and compact) and the convex envelope of f taken over the feasible region, in terms of its extreme points. The problem is initially relaxed by selecting a subset of the constraints, and a master problem is formed using the extreme points of the feasible region of this relaxation. A column generation scheme is then used to refine this (outer) approximation.

By seeking local optima and generating cuts which eliminate these local optima, cutting plane algorithms are also able to find the global optimum. (cf. Cabot [6], Taha [24], and Tui [25].)

Falk [8] considered the problem P assuming only that f is lower semicontinuous and C compact, and demonstrated that the solution of the bounding problem obtained by replacing $f : C$ by its convex envelope taken over the convex hull of C is equal to the solution of the Lagrangian dual of the original problem. Thus the strength of the bound which is obtained is directly related to the size of the Lagrangian duality gap. Greenberg and Pierskalla ([12], cf. also [13]) demonstrated that the surrogate dual problem, to be described below, has in general a smaller duality gap, thereby providing a more effective bound. The surrogate dual is lesser known and less frequently utilized, however, because of its apparent computational intractability.

If C is not compact, but is a closed cone in the nonnegative orthant and f has the properties assumed in P, then for the purpose of bounding a solution of our problem P in a branch and bound algorithm, one need not solve the surrogate dual problem in order to make use of its superior bound in eliminating regions not containing the global optimizer [5]. Instead, the effectiveness of the surrogate dual as a bound may be tested by a search for a feasible solution of a system of linear inequalities. Success in this search guarantees that the surrogate dual will eliminate the region being considered. Since solving a linear system of inequalities (a computationally tractable problem) will take advantage of the superior performance of the intractable surrogate dual as a bounding problem, a new class of algorithms may be designed which provides alternatives to existing algorithms.

3. Surrogate Duality

Surrogate constraints were first used in implicit enumeration algorithms for zero-one integer linear programming problems. A surrogate constraint is a convex combination of an original set of inequality constraints. The constraint thus obtained is implied by the original set of constraints, so that fathoming tests may be performed on this derivative constraint as a substitute for the original constraints.

The theory of surrogate duality was developed by Greenberg and Pierskalla [12] for more general mathematical programming problems. Suppose that the original primal problem is

$$P' : \text{ Find } \Phi = \text{Min}\left\{f(x) : g(x) \geq b, x \in X\right\}, \quad (7)$$

where $f : R^n \to R^l,\ g : R^n \to R^m$, and X is a closed subset of R^n. Given a vector of surrogate multipliers $u \in R^m_+$(the nonnegative orthant), a surrogate constraint $ug(x) \geq ub$ is defined. The surrogate problem is therefore

$$S(u) = \text{Min}\left\{f(x) : ug(x) \geq ub, x \in X\right\} \quad (8)$$

Since the optimal solution of (7) is feasible in the surrogate problem (8), $S(u)$ clearly provides a lower bound for the solution of (7). The surrogate dual problem is to find the greatest lower bound provided by the family of surrogate problems, i.e.,

$$\hat{S} = \underset{u \geq 0}{\text{Max}}\ S(u) \quad (9)$$

The duality gap of the surrogate dual is less than or equal to that of the more familiar Lagrangian dual, i.e.,

$$\hat{L} \leq \hat{S} \leq \Phi \quad (10)$$

where

$$\hat{L} = \underset{\lambda \geq 0}{\text{Max}}\ L(\lambda)$$

and

$$L(\lambda) = \underset{x \in X}{\text{Min}}\ f(x) - \lambda\left[g(x) - b\right]$$

Even though there may exist a gap between the optimal values of the primal and dual problems when the objective function or feasible region is not convex, the surrogate dual (9) can provide a lower bound for use in fathoming subproblems in a branch and bound algorithm.

Because of the economies of scale and lack of any upper bounds on x (other than those which might be included in the relaxed constraints $Ax \geq b$), the optimal vector of activity levels satisfying the single surrogate constraint $uAx \geq ub$ need have at most one positive activity level. That is, the minimum of an explicitly quasiconcave function occurs at an extreme point of the feasible region:

Theorem 1 (Martos [19]). $\phi(x)$ is quasiconcave in the convex set $X \subset R^n$, if and only if for each nonempty polyhedron $Y^\Delta \subset X$ any global vertex-minimum point of $\phi(x)$in Y^Δ is a global minimum point in Y^Δ.

Thus, considering problem P and its surrogate dual, letting A^j denote column $j$of the matrix A, and e_j the jth unit coordinate vector, the extreme points of this feasible region, other than perhaps the origin, are each of the form $\left({ub}/{uA^j}\right) e_j$, where $\left({ub}/{uA^j}\right) \geq 0$. If $ub \leq 0$, then $x = 0$ is both feasible and optimal, while if $ub > 0$ and $uA \leq 0$, then the surrogate problem is infeasible. Define the function $f_j\left(t\right) = f\left(te_j\right)$.

Our interest lies in whether $\hat{S}$exceeds some quantity V, e.g., an incumbent value.

Elsewhere we have proved the following result which states that the condition $\hat{S} > V$ may be tested by a search for a feasible solution of the set of linear inequalities:

Theorem 2 ([Bricker [5]). Let $f : R_+^n$ be an isotone, lower semicontinuous, explicitly quasiconcave and homogeneous function, $A \in R^{m\times n}$, and $b \in R^m$. Then the surrogate dual solution of the primal problem

$$\text{Min}\left\{f\left(x\right) : Ax \geq b, x \geq 0\right\}$$

exceeds a finite scalar value V if and only if one of the following conditions is satisfied:

(a) $b \leq 0$ and $V > 0$.

(b) the primal problem is infeasible.

(c) there exists a solution of the linear system

$$\begin{cases} u\left[b - f_j^{-1}\left(V\right) A^j\right] \geq 0, j \in J_1 \\ uA^j \leq 0, j \in J_2 \\ ub > 0, u \geq 0 \end{cases}$$

where $f_j\left(t\right) \equiv f\left(te_j\right)$, e_j being the jth unit coordinate vector, and

$$J_1 = \left\{j : f_j\left(t\right) \geq V \text{ for some } t \geq 0\right\}$$

$$J_2 = \left\{j : f_j\left(t\right) < V \text{ for all } t \geq 0\right\}$$

This feasibility test may be interpreted geometrically as a search for a hyperplane with nonnegative normal vector, separating the vector b from each of the points $\left\{f_j^{-1}\left(V\right) A^j : j \in J_1\right\}$, the origin, and the rays $\left\{A^j : j \in J_2\right\}$. This test can also be given an economic interpretation: if V is a capital investment budget for a production facility, then $f_j^{-1}\left(V\right)$ is the capacity of process j if the budget were allocated exclusively to that process. If this capacity is finite, $f_j^{-1}\left(V\right) A^j$ is then the vector of outputs of process j. The test above then amounts to searching for a set of inputs$u_1, u_2, \ldots u_m$, one for each output, with the property that the combined value of the required outputs $b_1, b_2, \ldots b_m$ exceeds the combined value of the outputs of each activity when operated at capacity. If the budget were to permit an unlimited level of activity j then$u$ is to be selected so that the combined values of the outputs of a unit level of activity, i.e., uA^j, must be negative, i.e., the value of inputs (negative components of A^j) must exceed the value of outputs (positive components of A^j). If such a set of values $u \geq 0$ can be found, then we can conclude that $V < \hat{S}$.

4. Example I

Consider the simple problem

$\Phi = \text{Min } f_1(x_1) + f_2(x_2) + f_3(x_3)$

s.t. $1.25x_1 + 3x_2 + 5x_3 \geq 15$ (11)

$4x_1 + 2x_2 + x_3 \geq 11$

$x_1 \geq 0$, $x_2 \geq 0$, $x_3 \geq 0$

where the functions $f_j\ (j = 1, 2, 3)$ are of the form (4), with ϕ, α, and β as specified in Table 1.

Table 1

j	x	y	z
ϕ_j	xux	yux	xu5
α_j	xu4	xu5	xuy5
β_j	wu6	wu8	wu7

Characteristics of example

Variables x_1, x_2, and x_3 might represent the levels of activity of three proposed production facilities, while 15 and 11 are the required outputs of two products. The operating cost of facility j includes a fixed charge ϕ_j for opening the facility plus a production cost $\alpha_j x_j^{\beta_j}$ which exhibits economies of scale.

The inverse function f_j^{-1} represents the activity level of facility as a function of the operating budget allocated to it. Thus, for an operating budget V, facility j may operate at a level

$$f_j^{-1}\left(V\right) = \begin{cases} \left(\frac{V - \phi_j}{\alpha_j}\right)^{1/\beta_j} \text{ if } V \geq \phi_j \\ 0 \text{ if } 0 \leq V < \phi_j \end{cases}$$

Since clearly $\Phi < +\infty$ and $b > 0$, the test of the validity of the inequality

$$\hat{S} > V$$

is, according to Theorem 2, the test of the feasibility of the linear system

$$\begin{cases} \left[15 - 1.25f_1^{-1}\left(V\right)\right] u_1 + \left[11 - 4f_1^{-1}\left(V\right)\right] u_2 \geq 0 \\ \left[15 - 3.00f_2^{-1}\left(V\right)\right] u_1 + \left[11 - 2f_2^{-1}\left(V\right)\right] u_2 \geq 0 \\ \left[15 - 5.00f_3^{-1}\left(V\right)\right] u_1 + \left[11 - f_3^{-1}\left(V\right)\right] u_2 \geq 0 \end{cases} \quad (12)$$

Suppose we apply the surrogate test above, "arbitrarily" using V=4.8. Then $f_j^{-1}(4.8)$is 5.05192, 2.08493, and 4.00213 for j=1, 2, 3, respectively. To determine whether the surrogate dual value exceeds 4.8, we need to test the feasibility of

$$\begin{cases} 8.68510\, u_1 - 9.20767\, u_2 \geq 0 \\ 8.74522\, u_1 + 6.83015\, u_2 \geq 0 \\ -5.01066\, u_1 + 6.99787\, u_2 \geq 0 \\ u_1 \geq 0, u_2 \geq 0 \end{cases} \tag{13}$$

The system of inequalities (13) is feasible (e.g., u_1 = 0.52, u_2 = 0.48), implying that the surrogate dual exceeds 4.8. (In fact, an enumeration of the basic solutions will determine that the optimal solution is found at $x = (2.13333,\ 0,\ 2.46667)$ which has cost of 7.15751. The surrogate duality gap for this problem is therefore at least o7ux575xtg4u8pv7ux575x or approximately 33%.

5. Implementation

In what follows, we will consider separable concave investment cost functions

$$f(x) = \sum_{j=1}^{n} f_j(x_j)$$

where $f_j(x_j)$ satisfies the form (4) for a given fixed cost $\phi_j > 0$, and parameters $\alpha_j \geq 0$ and $0 < \beta_j \leq 1$. The *elasticity* $\beta_j$accounts for economies of scale. (For example, if the elasticity $\beta_j$is 0.9, then increasing the capacity x_j by 10% increases the variable portion of the investment cost f_j by approximately 9%) We will refer to the variables x_j as *structural* variables indexed by $J = \{1, 2, ..., n\}$,and the surplus variables in the inequalities *Ax* $\geq$ b as *logical* variables indexed by $I = \{1, 2, ..., m\}$. As explained above, since we are enumerating the bases of a set of linear inequalities, we will construct an enumeration tree with subproblems defined by a specification of those variables which have been forced into the basis as well as those which have been excluded from the basis. Partition the structural variables by

$$J = J^+ \cup J^o \cup J^f$$

where J^+ and J^o denote the indices of the structural variables which have been forced into and excluded from the basis, respectively, and J^fdenotes the indices of the structural variables whose status has not yet been determined (free variables). Likewise, partition the set of logical variables $I$by

$$I = I^+ \cup I^o \cup I^f$$

where I^+, I^o, and I^f correspond in meaning to J^+, J^o, and J^f. Then the subproblem at a node of the enumeration tree is

$$\sum_{j \in J^+} \phi_j + \text{Min} \sum_{j \notin J^o} \bar{f}_j(x_j)$$

$$\text{s.t.} \sum_{j \notin J^o} a_{ij} x_j > b_i, i \in I^+ \tag{14}$$

$$\sum_{j \notin J^o} a_{ij} x_j = b_i, i \in I^o$$

$$\sum_{j \notin J^o} a_{ij} x_j \geq b_i, i \in I^f$$

$$x_j \geq 0, j \in J^+ \cup J^f, x_j = 0 \text{ for } x_j \in J^o.$$

where

$$\bar{f}_j(x_j) = \begin{cases} \alpha_j x_j^{\beta_j} \text{ if } j \in J^+ \\ \phi_j + \alpha_j x_j^{\beta_j} \text{ if } j \in J^f \end{cases}$$

If the incumbent value is V, then the current subproblem may be fathomed, provided that the surrogate dual value exceeds V. This condition is satisfied if there exists a nonnegative solution $u \in R^m$to the inequalities

$$\begin{cases} u\left[b - f_j^{-1}(V)\, A^j\right] \geq 0,\ j \notin J^o \\ ub > 0 \\ u \geq 0, i \notin I^o \end{cases}$$

that is,

$$\sum_{i=1}^{m} \left(b_i - \left(\frac{V}{\alpha_j}\right)^{1/\beta_j} a_{ij} \right) u_i \geq 0,\ j \in J^+$$

$$\sum_{i=1}^{m} \left(b_i - \left(\frac{V - \phi_j}{\alpha_j}\right)^{1/\beta_j} a_{ij} \right) u_i \geq 0,\ j \in J^f \tag{15}$$

$$\sum_{i=1}^{m} b_i u_i > 0,$$

$$u_j \geq 0 \text{ for } i \in I^+ \cup I^f, u_i = 0 \text{ for } i \in I^o$$

The search for a feasible solution of (15) may be performed by one of several alternatives, e.g., (i) an adaptation of the *relaxation* algorithm of Agmon [1] and of Motzkin and Schoenberg [21], a method often used in a search for optimal multipliers when solving Lagrangian dual problems (where it is known as the subgradient algorithm [15]), *(ii)* the *ellipsoid* algorithm of Khachian [18] (cf. also Bland et al. [3]), *or (iii) a Phase-One Simplex LP algorithm*. The first two are iterative methods which converge to a feasible solution (the ellipsoid method in a number of iterations polynomial in the length of the problem data string) if such a solution exists. In practice, we will terminate the selected algorithm after a specified number of iterations, abandoning the attempt to fathom the node of the enumeration tree.

Fathoming process.

At each node of the branch-and-bound tree, we use the following procedure as fathoming process:

Step 1: fixed costs of variables which have been forced into the basis are summed.

Step 2a: if it is greater than the incumbent, the node is fathomed.

Step 2b: Otherwise, an attempt is made to fathom by the surrogate bound. If that fails, then try obtaining a feasible solution to subproblem and possibly fathom by infeasibility or update incumbent. This attempt is abandoned if not successful within a specified number of iterations.

Step 3: If both of these two attempts to fathom fail, separate subproblem into two subproblems and fathom each of them.

Linear programming approximation. If a node cannot be fathomed, before branching we compute an upper bound on the optimum by finding a feasible solution to the problem (14). This is accomplished by solving the optimization problem (14) with the objective replaced by a linear approximation cx, i.e.,

$$\sum_{j\in J^+} \phi_j + \text{Min} \sum_{j\in J^+\cup J^f} c_j x_j \qquad (16)$$

wherec_j is chosen to be either

(i) marginal costs at $\frac{ub}{uA^j}$ (where $u$is the multiplier vector upon termination of the relaxation algorithm), i.e., $c_j = \frac{df}{dx}\left(\frac{ub}{uA^j}\right) = \alpha_j\beta_j\left(\frac{ub}{uA^j}\right)^{\beta_j-1}$, or

(ii) the average unit cost

$$c_j = f_j\left(\frac{ub}{uA^j}\right)\Big/\frac{ub}{uA^j}$$

(See Figure 1.) In either case, the value of the objective function of (16), evaluated at the solution of the LP, provides an upper bound on the optimum which, if less than V, becomes the new incumbent. (If the LP is determined to be infeasible, then of course the node is fathomed.)

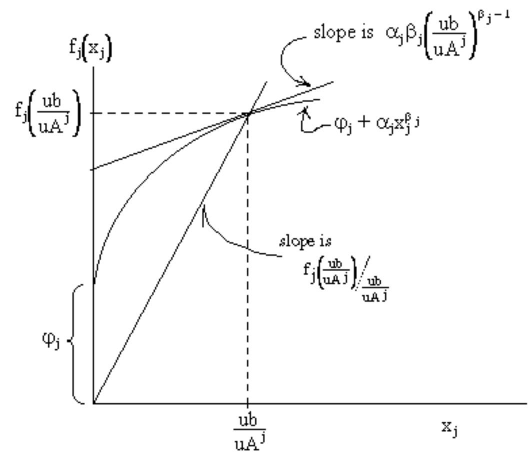

Fig. 1. Linear approximations of objective function

Branching process.

Branching must be done from nodes which cannot be fathomed. This is accomplished by selecting a variable and constructing two descendent nodes—one in which the variable is forced into the basis, and the other in which the variable is excluded from the basis. Logical variables will be selected only when no free structural variables remain. The choice of structural variable may be made based upon one of the following criteria:

(1) select the structural variable with largest cost coefficient α.
(2) select the structural variable with largest fixed cost φ.
(3) select the structural variable with largest cost at ub/uA^j.

If no free structural variables remain, the logical variable corresponding to the largest surrogate multiplier will be selected for branching.

Test Problems.

Random problems were generated according to the following scheme:

(1) integer constraint coefficients are uniformly distributed in a specified interval.
(2) a specified density δ is achieved by replacing randomly-selected constraint coefficients with zero (while avoiding the creation of columns with zero density).
(3) fixed costs (φ), cost coefficients (α) and exponents (β) are uniformly distributed in specified intervals.

(4) the right-hand-side vector b was assigned by fixing a specified fraction ρ of the variables at a specified value ξ and evaluating the linear constraint functions at this point.

6. Example II.

A problem with 5 constraints and 20 structural variables was randomly generated with the following characteristics:

- constraint coefficients are uniform in [-2, +10]
- density 75%
- right-hand-sides b obtained by evaluating the linear functions after fixing
- 25% of the variables at the value 5
- fixed costs φ uniform in [2,5]
- cost coefficient α uniform in [0.5, 2.0]
- cost exponent β uniform in [0.5, 1]

The cost parameters and constraint parameters are shown in Tables 2 and 3, respectively.

Table 2

j	$\varphi[j]$	$\alpha[j]$	$\beta[j]$
1	3	1.98	0.55
2	5	1.87	0.69
3	4	1.65	0.64
4	4	0.52	0.86
5	4	0.18	0.78
6	5	0.10	0.74
7	3	0.53	0.80
8	5	1.93	0.67
9	5	0.16	0.92
10	4	0.42	0.70
11	3	0.27	0.64
12	5	0.43	0.59
13	5	1.49	0.83
14	5	1.04	0.56
15	4	0.13	0.93
16	3	0.67	0.79
17	4	0.92	0.97
18	4	0.46	0.94
19	4	0.66	0.85
20	4	1.22	0.58

Cost parameters for the 5×50 example problem

Up to 50 iterations of the relaxation algorithm will be used here for testing the surrogate dual bound. The LP, which is solved at nodes at which 3 or more variables have been either forced into or out of the basis, uses as its objective the marginal costs at the point with $x_j = \frac{\sum_i u_i b_i}{\sum_i u_i a_{ij}}$ where u is the most recently computed surrogate multiplier vector.

Branch-&-Bound Algorithm

Current Parameters

The current values of parameters to be used in the branch-

and-bound algorithm are:

TAU = reflection factor = 1.5

EPS1 = tolerance for fathoming = 0.0001

Max_Relax = max # relaxation iterations = 50

Do_LP_level = level of tree at which LP is attempted = 3

Rule for choosing structural variable for branch based upon:

Maximum Cost of F(Ub/UA)

LP objective computed by

marginal cost at X=Ub/UA

****

A total of 497 nodes of 53130 (0.935 %) were examined,

of which

432 nodes were fathomed by surrogate dual bound,

4 nodes were fathomed by excessive fixed costs,

3 nodes were fathomed by infeasibility of LP.

The total number of basic solutions (both feasible and infeasible) of the constraints is $\binom{20}{5}$= 53130. In this example, 497 nodes of the enumeration tree were generated, which is less than 1% of the number of basic solutions. The overwhelming majority of these nodes were fathomed by the surrogate dual. The optimal solution (shown in Table 4) has two positive structural variables, contributing approximately 44% and 56%, respectively, to the total cost.

Table 4

j	x_j	Cost	$F\%$
5	11.666667	5.2231818	44.27
10	13.333333	6.5745801	55.72
Surplus in Constraints ($Ax - b$)			
i	S[i]		
1	1.6666667		
4	60.0		
5	50.0		

Optimal solution of the 5×20 example problem with optimal cost 11.7977618611

Table 3

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20		b
-1	3	-1	0	-1	0	0	6	0	4	0	5	0	3	2	6	6	5	4	5	$\geq$	40
2	1	-1	4	0	0	2	3	1	3	4	3	2	1	3	2	5	0	4	6	$\geq$	40
5	6	4	0	3	4	0	2	0	0	0	0	0	0	0	0	5	3	1	4	$\geq$	35
0	6	0	1	4	0	0	0	5	4	0	0	1	0	1	0	0	-1	0	4	$\geq$	40
1	0	0	6	5	3	3	2	0	2	5	5	0	4	0	4	0	-1	-1	0	$\geq$	80

Constraint parameters for 5×20 example problem.

7. Computational Experience and Analysis of results

Although the algorithm does not require the explicit computation of the surrogate dual value, in the experiments below the effort was expended to compute and report this dual value for the root node of the enumeration tree, i.e., the original problem.

The proposed algorithm was implemented in the APL language using APL+Win 3.6. To compare the performance of the surrogate dual bound with that of the Lagrangian dual bound, one hundred problems with 50 constraints and 100 structural variables were randomly generated with the same characteristics as Example II. Table 5 summarizes the computational results for each of the one hundred problems. Note that the surrogate dual provided a tighter lower bound than the Lagrangian dual. The smallest gap between the surrogate dual bound and the Lagrangian dual bound (as a percentage of the surrogate dual bound) is 0.369%, the mean is 21.885%, and the largest gap is 50%. Note also that, compared to the Lagrangian dual bound (LR), while the surrogate value (SD) takes much longer to compute the simple test of the inequalities (ST) is equally effective or even better in bounding within a branch-and-bound algorithm. On the average, the surrogate dual bound takes 20 times longer computational times than the Lagrangian dual bound. However, the simple test of the inequalities takes 85% less time than the Lagrangian dual bound.

Table 5

	GAP (%)	Time Ratio (LR/SD)	Time Ratio (LR/ST)
Min	0.369	0.017	1.832
Mean	21.885	0.050	7.409
Max	50.000	0.316	27.089

Summary of computational results

To test the performance of the implemented branch-and-bound algorithm for problems with a variety of characteristics, smaller test problems were randomly generated. Five levels were used for constraint coefficient density, and two levels each for fixed costs, cost coefficients, and cost exponents, providing forty different combinations. Specifically, these problems possessed the following characteristics:

- #constraints: 5
- #variables: 20
- #basic solutions: 53130
- Constraint coefficients: integers uniform in [-5, 10] with density δ = 1%, 25%, 50%, 75%, or 100%
- Fixed costs: integers uniform in either [0, 5] or [5, 10]
- Cost coefficient α uniform in either [0, 5] or [1, 10]
- Cost exponent β uniform in either [0.1, 0.5] or [0.5, 1]

One hundred problems were generated from each of the forty different combinations. Computational results are recorded in Tables 5 and 6. The first column of these tables represents problem generation parameters, i.e., it consists of three alphabetic letters (L or H indicating the lower or higher interval, respectively, for fixed costs, cost coefficients α, and cost exponents β) and one numerical number (1, 25, 50, 75, or 100, representing the density δ. (For example, problems in the set **LHH5** have fixed costs, cost coefficients, and cost exponents sampled from the intervals [0,5], [1, 10], and [0.5, 1], respectively, and 5% density.) The other statistics in these tables are defined as follows:

NN = number of nodes evaluated

NN% = ratio of NN to total number of basic solutions

NS = number of nodes fathomed by surrogate bound

NS% = fraction of nodes which were fathomed by surrogate dual bound

NF = number of nodes fathomed because sum of fixed cost exceeds incumbent

NF% = fraction of nodes which were fathomed because of excessive fixed costs

NI = number of nodes fathomed because of infeasible LP

NI% = fraction of nodes which were fathomed by infeasible LP

NJ = number of basic structural variables in optimal solution

F% = optimal fixed costs as fraction of total costs

%GAP = surrogate duality gap as percent of optimal cost

Statistical analysis based on both Tables 6 and 7 has been performed to identify the parameter(s) that causes the change in the surrogate duality gap.

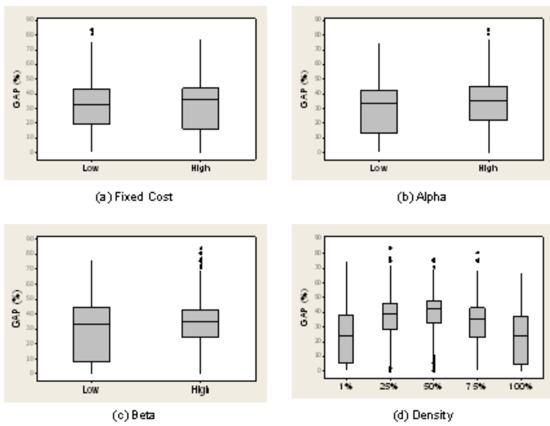

Fig. 2. Box-and-Whisker Plot

Effect of magnitude of fixed costs: In this research we have used two intervals in testing for the effect of fixed costs: uniform in either [0, 5] or [5, 10]. The Tukey's studentized range (HSD) means comparison test (Tukey [26]) has been performed to check if the average solution gap is same with two intervals tested. It is observed that the average surrogate duality gap between two intervals tested is *not* significantly different. Thus, the result suggests that the variation of surrogate duality gap does not depend on the value of the fixed cost in the interval tested. The box-and-whisker plot displayed in Figure 2a provides a graphical view of this result.

Effect of magnitude of cost coefficient (α: Two intervals were used in testing for the effect of the cost coefficient: uniform in either [0, 5] or [1, 10]. The Tukey's studentized range (HSD) means comparison test has been performed to check if the average solution gap is same across the two intervals tested. The analysis shows that the low value of cost coefficient is associated with the variation of the surrogate duality gap that is significantly different from the high value of cost coefficient. Furthermore, we note that problems with smaller cost coefficients have smaller surrogate duality gaps. The box-and-whisker plot also indicates that the surrogate duality gap may be more variable with low values of cost coefficients than with high values (See Figure 2b).

Effect of magnitude of cost exponent (β: Two intervals were used in testing for the effect of the cost exponent: uniform in either [0.1, 0.5] or [0.5, 1]. The Tukey's studentized range (HSD) means comparison test has been performed to check if the average solution gap is same with two intervals tested. We indicate that there is a significant difference in the surrogate duality gap across the value of β : the result suggests that the surrogate duality gap is larger for problems with higher value of cost exponent (See Figure 2c).

Effect of problem density: In this research we have used five categories for problem density: 1%, 25%, 50%, 75%, or 100%. The Tukey's studentized range (HSD) means comparison test has been performed to check whether the average solution gap is the same with five different density values. Test results shows that problem densities of both 1% and 100% are associated with a surrogate duality gap that is significantly different from the other three density types: changing density between 1% and 100% has no significant influence on the variation of the surrogate duality gap and so we have combined these two groups. On the other hand, we note that changing density among these four new groups (1 & 100%, 25%, 50%, and 75%) has a significant influence in accounting for the variation in the surrogate duality gap. We observe that problems with either 1% or 100% density are relatively easier to solve than those with other densities, and problems with 50% density are the most difficult (See Figure 2d).

8. Conclusion

We have presented a surrogate dual for the linearly-constrained capacity planning problem with (separable) fixed charges and continuous economies of scale of the form (4). This problem subsumes the linear fixed-charge problem (β =1) and the pure fixed charge problem (β =0).

Duality theory assures us that the surrogate duality gap is no larger than the Lagrangian duality gap, although the surrogate dual is generally more difficult to solve. We have shown, however, that the ability of the surrogate dual to fathom subproblems in a branch-and-bound algorithm may be determined without directly solving the surrogate dual itself, but that a simple test of the feasibility of a certain linear system of inequalities will suffice.

To compare the performance of the surrogate dual bound with the one of the Lagrangian dual bound, one hundred randomly generated problems with 50 constraints and 100 structural variables were tested. In this

Table 6

	NN	NN%	NS	NS%	NF	NF%	NI	NI%	NJ	F%	GAP
LLL1	696.7	1.31	56.95	25.40	28.30	12.50	23.38	2.60	1.41	51.96	15.99
LLL25	7333.0	13.80	319.50	10.89	247.00	3.31	197.50	2.51	1.96	56.38	33.57
LLL50	6970.0	13.12	343.10	11.62	281.40	3.62	222.60	3.05	1.98	52.81	34.05
LLL75	1615.0	3.04	131.40	19.34	86.53	8.52	79.77	3.47	1.70	53.06	22.32
LLL100	365.0	0.69	53.86	26.24	29.40	15.17	25.99	2.42	1.36	57.02	13.57
HLL1	421.7	0.79	27.66	14.90	97.22	27.31	13.81	2.27	1.08	75.31	10.52
HLL25	3031.0	5.71	133.20	8.14	317.00	12.94	59.14	2.29	1.59	71.30	33.98
HLL50	3113.0	5.86	136.10	6.44	440.40	16.25	95.29	3.08	1.87	76.44	39.93
HLL75	964.6	1.82	63.15	11.28	184.20	23.37	38.01	3.46	1.49	73.91	26.64
HLL100	328.9	0.62	32.19	14.53	92.34	28.41	11.81	2.46	1.10	74.10	12.86
LHL1	596.2	1.12	61.59	29.75	16.07	9.78	27.27	2.42	1.57	51.19	27.92
LHL25	12720.0	23.93	472.70	11.10	57.59	1.81	376.90	2.21	1.98	41.10	37.76
LHL50	12120.0	22.81	460.30	13.61	60.71	1.85	497.40	3.17	2.19	45.94	43.26
LHL75	3425.0	6.45	295.30	22.10	47.95	4.85	236.70	4.28	1.95	48.58	39.02
LHL100	771.7	1.45	94.18	32.21	20.54	7.48	60.62	2.57	1.58	48.98	27.16
HHL1	644.9	1.21	42.57	24.57	43.05	16.97	14.74	2.14	1.28	70.09	20.13
HHL25	6156.0	11.59	405.70	10.77	252.20	5.18	129.00	2.10	1.76	61.60	39.31
HHL50	4016.0	7.56	253.70	13.46	228.30	7.33	95.44	2.71	1.80	64.45	40.50
HHL75	867.6	1.63	89.31	21.67	113.40	11.78	38.33	3.25	1.58	68.33	32.86
HHL100	230.9	0.43	37.37	27.39	34.23	15.61	8.57	2.29	1.18	66.10	20.05
LLH1	3651.0	6.87	284.80	29.02	21.45	5.72	279.10	3.08	1.88	44.86	29.46
LLH25	15470.0	29.12	379.90	13.24	46.19	1.20	458.00	2.14	2.25	36.68	33.13
LLH50	16680.0	31.39	436.60	13.60	22.04	0.97	606.40	2.86	2.42	42.06	38.35
LLH75	6771.0	12.74	248.90	26.19	2.36	2.79	440.20	3.13	2.05	41.41	32.95
LLH100	1955.0	3.68	159.60	32.47	7.92	10.09	169.80	2.43	1.84	49.01	27.29
HLH1	1908.0	3.59	140.00	23.34	126.30	12.48	104.60	3.32	1.51	59.22	26.53
HLH25	6514.0	12.26	382.90	11.20	99.25	3.40	172.90	2.20	1.93	57.25	39.35
HLH50	7347.0	13.83	510.10	15.41	188.70	2.98	269.40	3.03	2.08	57.69	42.96
HLH75	1105.0	2.08	156.60	24.37	55.64	7.04	66.28	3.49	1.77	60.42	34.37
HLH100	310.1	0.58	48.22	30.67	27.46	13.31	16.58	2.28	1.54	65.27	26.97
LHH1	5934.0	11.17	241.10	32.71	1.52	5.54	425.80	2.06	1.81	33.93	26.32
LHH25	23850.0	44.90	441.40	10.39	0.12	0.24	787.60	2.27	2.60	29.56	36.32
LHH50	21490.0	40.44	366.60	14.69	0.15	0.40	928.00	2.50	2.48	27.66	34.59
LHH75	14720.0	27.71	274.90	23.92	2.36	1.74	886.10	3.04	2.24	33.63	33.07
LHH100	8929.0	16.81	150.80	34.68	0.55	5.44	675.90	2.48	1.92	33.27	28.28
HHH1	3343.0	6.29	214.70	30.99	24.78	6.70	240.90	2.98	1.69	48.31	28.01
HHH25	13640.0	25.68	536.90	12.40	74.44	1.67	459.20	2.14	2.16	45.18	37.65
HHH50	10960.0	20.63	513.30	16.32	111.40	1.33	435.70	2.93	2.27	49.51	39.31
HHH75	3970.0	7.47	286.60	24.60	31.76	4.17	240.20	3.41	1.96	49.56	35.92
HHH100	1884.0	3.55	167.90	32.96	15.40	8.66	159.90	2.49	1.66	52.12	28.43

Mean values

Table 7

	NN	NN%	NS	NS%	NF	NF%	NI	NI%	NJ	F%	GAP
LLL1	113	0.21	28	22.42	3	7.32	2	1.27	1	48.26	12.38
LLL25	2461	4.63	180	8.85	4	0.74	52	2.49	2	57.02	38.20
LLL50	2313	4.35	211	8.50	6	0.89	81	2.70	2	50.08	39.60
LLL75	417	0.78	55	17.88	4	4.35	11	2.99	2	49.77	28.66
LLL100	85	0.16	29	22.68	3	11.43	1	1.30	1	57.04	6.11
HLL1	205	0.39	19	10.89	56	29.58	5	2.21	1	74.69	3.85
HLL25	1639	3.09	87	5.40	230	12.66	34	2.02	2	71.43	39.11
HLL50	2271	4.27	80	4.01	319	17.31	76	3.03	2	77.06	43.65
HLL75	507	0.95	50	8.33	119	23.91	16	3.16	1	74.26	31.85
HLL100	261	0.49	19	11.11	63	31.68	6	2.56	1	77.54	5.71
LHL1	55	0.10	19	30.50	2	4.65	0	0.00	2	51.62	29.31
LHL25	4261	8.02	192	7.07	1	0.01	32	2.10	2	38.11	42.10
LHL50	1911	3.60	179	11.67	1	0.00	46	3.20	2	43.60	45.57
LHL75	533	1.00	94	20.83	1	0.27	18	4.78	2	49.81	42.97
LHL100	47	0.09	18	30.47	1	1.42	0	0.00	2	47.75	30.03
HHL1	89	0.17	24	22.22	4	16.23	2	1.77	1	72.83	13.72
HHL25	3521	6.63	170	7.83	24	1.31	45	1.67	2	57.95	45.24
HHL50	1479	2.78	132	11.84	10	2.43	36	2.70	2	64.33	44.16
HHL75	377	0.71	62	22.22	5	4.91	10	3.05	2	71.11	39.27
HHL100	101	0.19	27	28.21	4	13.91	2	2.22	1	65.77	17.79
LLH1	85	0.16	29	27.71	0	0.00	1	1.41	2	45.73	29.31
LLH25	2177	4.10	102	6.42	0	0.00	36	2.02	2	35.72	33.48
LLH50	3533	6.65	158	6.21	0	0.00	70	2.87	2	40.81	38.07
LLH75	219	0.41	39	22.57	0	0.00	7	1.75	2	38.17	33.21
LLH100	11	0.02	5	33.33	1	1.86	0	0.00	2	48.53	28.32
HLH1	247	0.46	34	22.22	3	5.88	7	2.56	1	58.20	31.91
HLH25	2771	5.22	157	8.28	1	0.07	30	1.96	2	53.30	40.83
HLH50	3117	5.87	246	13.76	1	0.05	57	3.05	2	55.68	42.90
HLH75	313	0.59	37	22.22	2	2.01	6	3.10	2	58.44	37.94
HLH100	59	0.11	19	33.33	2	6.45	1	1.70	2	67.43	32.99
LHH1	9	0.02	5	33.33	0	0.00	0	0.00	2	32.77	27.22
LHH25	6183	11.64	64	4.48	0	0.00	86	2.16	3	27.29	33.92
LHH50	3367	6.34	43	5.86	0	0.00	53	2.43	2	25.59	33.52
LHH75	223	0.42	6	17.04	0	0.00	6	1.33	2	29.53	31.64
LHH100	9	0.02	5	44.44	0	0.00	0	0.00	2	28.95	28.35
HHH1	71	0.13	22	33.33	0	0.00	1	1.40	2	45.54	31.67
HHH25	1941	3.65	113	7.51	0	0.00	32	2.08	2	43.05	37.71
HHH50	1313	2.47	142	12.74	0	0.00	20	2.70	2	46.81	40.32
HHH75	471	0.89	55	22.37	0	0.00	11	2.88	2	48.92	36.83
HHH100	45	0.08	14	33.33	1	2.02	0	0.00	2	47.93	32.21

Median values

experiment, the tighter bound was obtained with the surrogate dual for every test problem.

In order to characterize problems of this class having relatively small surrogate duality gaps, a statistical analysis of the surrogate duality gap of four thousand randomly generated problems with various problem parameters (density of linear constraints, fixed costs ϕ, cost coefficients α, and exponents β) was performed. This analysis suggests that these gaps are smaller for problems with either very high or very low density, and for problems with small values of α or β.

References

[1] Agmon S. The Relaxation Method for Linear Inequalities. Canadian Journal of Mathematics 1954; 6: 382-392.

[2] Benson HP. Constrained Concave Minimization: A Unified Critical Survey. Naval Research Logistics 1996; 43: 765-795.

[3] Bland RG, Goldfarb D, Todd MJ. The Ellipsoid Method: a Survey. Operations Research 1981; 29: 1039-1091.

[4] Bomze IM, Csendes T, Horst R, Pardalos PM (Eds.). Developments in Global Optimization (Nonconvex Optimization and Its Applications) Dordrecht, Kluwer Academic Publishers, 1997.

[5] Bricker DL. Bounding a Class of Nonconvex Linearly-Constrained Resource Allocation Problems via the Surrogate Dual. Mathematical Programming 1980; 18: 68-83.

[6] Cabot A. Variations on a Cutting Plane Method for Solving Concave Minimization Problems with Linear Constraints. Naval Research Logistics Quarterly 1974; 21: 265-274.

[7] Carillo M. A Relaxation Algorithm for the Minimization of a Quasiconcave Function on a Convex Polyhedron. Mathematical Programming 1977; 13: 69-80.

[8] Falk JE. Lagrange Multipliers and Nonconvex Programs. SIAM Journal of Control 1969; 7(4): 534-545.

[9] Falk JE, Hoffman KR. A successive underestimation method for concave minimization problems. Mathematics of Operations Research. 1976; 1: 251-259.

Received 18 December 2006; revised 23 May 2007; accepted 29 May 2007

[10] Falk J, Horowitz J. Critical Path Problems with Concave Cost-time Curves. Management Science 1972; 19: 446-455.

[11] Falk JE, Soland R. An Algorithm for Separable Nonconvex Programming Problems. Management Science 1969; 15: 550-562.

[12] Greenberg HJ, Pierskalla WP. Surrogate Mathematical Programming. Operations Research 1970; 18: 924-939.

[13] Greenberg HJ, Pierskalla WP. Quasi-Conjugate Functions and Surrogate Duality. Cahier Centre d'Etude de Recherche Operationelle 1973; 15: 437-448.

[14] Hax AC, Majluf NS. Competitive cost dynamics: the experience curve. Interfaces 1982; 12(5).

[15] Held M, Wolfe P, Crowder HP. Validation of subgradient optimization. Mathematical Programming 1974; 6(1): 62-88.

[16] Horst R. An Algorithm for Nonconvex Programming Problems. Mathematical Programming 1976; 10: 312-321.

[17] Jones A, Soland R. A branch and bound algorithm for multi-level fixed charge problems. Management Science 1969; 16(1): 67-76.

[18] Khachian LG. A Polynomial Algorithm in Linear Programming. Doklady Akademiia Nauk SSSR 1979; 224: 1093-1096.

[19] Martos B. Nonlinear Programming Theory and Methods. Amsterdam, North-Holland, 1975.

[20] McCormick GP. Attempts to Calculate Global Solutions of Problems That May Have Local Minima. Numerical Methods for Nonlinear Optimization. F. A. Lootsma. New York, Academic Press, 1972. p. 209-221.

[21] Motzkin TS, Schoenberg IJ. The Relaxation Method for Linear Inequalities. Canadian Journal of Mathematics 1954; 6: 393-484.

[22] Rech P, Barton LG. A Non-convex Transportation Algorithm. In: Applications of Mathematical Programming Techniques. E. M. L. Beale (Eds.). New York, American Elsevier, 1970.

[23] Soland RM. Optimal Facility Location with Concave Costs. Operations Research 1974; 22: 373-382.

[24] Taha HA. Concave Minimization over a Convex Polyhedron. Naval_Research Logistics Quarterly 1973; 20: 533-548.

[25] Tui H. Concave Programming Under Linear Constraints. Soviet Mathematics 1964; 5: 1437-1440.

[26] Tukey JW. Comparing Individual Means in the Analysis of Variance. Biometrics 1949; 5: 99-114.

[27] Yelle LE. The learning curve: historical review and comprehensive survey. Decision Science 1979; 10(2): 302-328.