Asia-Pacific Journal of Operational Research
Vol. 21, No. 1 (2004) 9–33

A CONVEX SUBMODEL WITH APPLICATION TO SYSTEM DESIGN

JAVIER SALMERÓN
Operations Research Department, Naval Postgraduate School
Monterey, CA, 93943, USA
jsalmero@nps.navy.mil

ÁNGEL MARÍN
Departamento de Matemática Aplicada y Estadística
E.T.S.I. Aeronáuticos, Universidad Politécnica de Madrid
28040 Madrid, Spain
amarin@dmae.upm.es

Received July 2001
Revised May 2003

In this paper, we present an algorithm to solve a particular convex model explicitly. The model may massively arise when, for example, Benders decomposition or Lagrangean relaxation-decomposition is applied to solve large design problems in facility location and capacity expansion. To attain the optimal solution of the model, we analyze its Karush–Kuhn–Tucker optimality conditions and develop a constructive algorithm that provides the optimal primal and dual solutions. This approach yields better performance than other convex optimization techniques.

Keywords: Convex programming; decomposition methods; Karush–Kuhn–Tucker optimality conditions.

1. Introduction

In this paper, we develop an algorithm to derive an explicit solution to the following model — which we refer to as "subproblem" (SP):

$$\text{SP:} \qquad \min_{z_1,\ldots,z_I} \quad \sum_{i=1}^{I} q_i z_i + W(\tau)$$

$$\text{s.t.} \qquad \begin{cases} \tau = \sum_{i=1}^{I} p_i z_i, \\ \tau \in [l, m], \\ z_i \in [0, b_i], \quad \forall i = 1, \ldots, I, \end{cases}$$

where $q_i \geq 0$, $b_i > 0$, $p_i \geq 0, \forall i = 1, \ldots, I, 0 \leq l < m \leq +\infty$, and $W(.)$ is a suitable convex function whose properties will be established later in this paper.

Remark: We shall assume $p_i = 1, \forall i$ without loss of generality. This can be accomplished by substituting z_i in the SP model by $\tilde{z}_i = p_i z_i$.

This work will show that an explicit solution to both the primal and the dual of model SP can be obtained with very little computational effort. This, in turn, helps speed up convergence of decomposition techniques that massively employ SP as subproblem.

In fact, our interest in this model emerged as a result of applying several decomposition techniques to large-scale capacity expansion problems in the electric field (Marín and Salmerón, 1998, 2001).

Other capacity expansion and system design models involving similar structures may also benefit from the results presented in this work. These models are typically characterized by strategic decisions (e.g., time, location, and capacity for new facilities) and operating decisions (related to the use of that capacity, e.g., in order to meet the system demand over time), sometimes in an uncertainty context. Relevant design problems in capacity expansion involving structures similar to SP arise in models by Bloom (1983), Sherali and Staschus (1990), Bean *et al.* (1992), Malcom and Zenios (1994), Álvarez *et al.* (1994), Hobbs (1995), and Laguna (1998), among others. In these problems, the so-called *strategic* variables (or related constraints) are separated from the *operating* variables by employing Benders decomposition (Geoffrion, 1972), or Lagrangean relaxation and Lagrangean decomposition (Guignard and Kim, 1987).

To clarify the exposition, let us consider a facility location problem with candidate facilities indexed by $i = 1, \ldots, I$, and a time scale indexed by $t = 1, \ldots, T$. A decision variable x_{it} determines whether facility i is considered for expansion in period t. Accordingly, operating variables, z_{it}, refer to the level of utilization of the existing capacity. If d_t is the estimated demand in period t, a typical set of demand constraints is:

$$\sum_{i=1}^{I} p_{it} z_{it} = d_t, \quad \forall t = 1, \ldots, T,$$

which appears in the structure of our subproblem SP. On the other hand, one expects to be able to operate facility i in period t according to the level of investment up to period t. Thus, $z_{it} \leq f(x_{i1}, \ldots, x_{it})$. Notice that, for a given value of x, we may compute $f(x)$ and the expression above becomes $z_{it} \leq b_{it}$, as in SP.

In the presence of demand uncertainty, in addition to linear expansion and production costs, we may incorporate non-linear terms in the objective function. These terms seek to penalize the adequacy of our decision due to demand variability (e.g., Malcom and Zenios, 1994). One way to proceed is through convex penalty functions of the form $W_t(\tau_t)$ where $\tau_t = \sum_{i=1}^{I} p_{it} z_{it}$. Bounds on τ_t, i.e., $\tau_t \in [l_t, m_t]$ (for convenient l_t, m_t depending on the probability distribution of the demand in period t) may be considered. This idea can be applied to many other facility location problems, e.g., Cornuejols *et al.* (1990), Magnanti and Wong (1990), etc., originally formulated under the assumption of deterministic demand.

In other facility location and capacity acquisition combined problems (e.g., Jacobsen, 1990; Jack *et al.*, 1992; Daskin *et al.*, 1992; Drezner, 1995; Verter and Cemal, 1995; Current *et al.*, 1998), model SP appears as the result of successive decompositions.

We next establish the generic formulation of the design model that motivates our work. This model will be referred to as Design Model (DM). DM is formulated as follows:

$$\text{DM:} \qquad \min_{x,z} \quad f(x,z) = C_x \cdot x + C_z \cdot z + R(z)$$
$$\text{s.t.} \quad \begin{cases} Bx + Az \leq b^0, \\ x \in \mathcal{X}, \\ z \in \mathcal{Z}, \end{cases} \tag{1.1}$$

where $R(z) = \sum_{t=1}^{T} W_t(z)$, and $\mathcal{X}$ and $\mathcal{Z}$ are the sets of independent constraints for the strategic and operating decisions (vectors x and z, respectively). $\mathcal{X}$ is assumed to be a non-empty, compact set. For example, $\mathcal{X}$ may consist of alternative expansion plans defined through integer variables, a set of resource constraints (e.g., yearly budget), etc. We make the additional assumption that $\mathcal{Z}$ is defined by the following constraints:

$$\mathcal{Z} \equiv \left\{ \begin{array}{ll} l_t \leq \sum_{i=1}^{I} p_{it} z_{it} \leq m_t, & \forall t = 1, \ldots, T \\ z_t \geq 0, & \forall t = 1, \ldots, T \end{array} \right\} \tag{1.2}$$

C_x and C_z are linear costs associated with x and z. A and B are matrices representing coupling constraints that usually limit the operating capacity as a function of the available capacity after an expansion decision has been made. b^0 is the vector of initial capacity. We shall assume $p_{it} = 1$ in the DM (substituting z_{it} by $\tilde{z}_{it} = p_{it} z_{it}$, if needed).

Applying Lagrangian relaxation (e.g., Fisher, 1981) to model DM, it is immediate to verify that by relaxing $Bx + Az \leq b^0$ the resulting model (after incorporating this constraint as a penalty in the objective function), can be decomposed into T submodels like SP. The same occurs using Lagrangean decomposition (Guignard and Kim, 1987), by splitting z and then dualizing the splitting constraints. In these two types of decomposition, the resulting SPs have $b_i = +\infty$. Likewise, when A is diagonal and Generalized Benders' Decomposition is used (considering x as "complicated variables," e.g., Geoffrion, 1972), the resulting subproblem is separable into T subproblems of type SP. In this case, $b_i < +\infty$.

Solving the dual and primal of SP efficiently is crucial to implement any of the above-mentioned techniques. This motivates our exhaustive analysis of the SP submodel in this work, so as to help reduce the computational burden of massively solving instances of SP. In particular, we propose solving its Karush–Kuhn–Tucker optimality conditions (KKTOCs) (Bazaraa and Shetty, 1993, pp. 150–172).

Examples of applied models amenable to solution by their KKTOCs are not very frequent. Our analysis to solve KKTOCs for SP is similar to that of Ferland *et al.* (1991), who exploit KKTOCs to determine optimal multipliers for the cuts introduced in Benders' decomposition master problem. They generate the optimal multipliers (which may not be unique) by analyzing different cases. Explicit use of KKTOCs can also be found in Murphy and Wang (1993), who show how to reconstruct the dual solution of a capacity expansion model by using a network model employing the information provided by the KKTOCs.

Our paper is organized as follows: In Section 2, we present additional details on applications of model SP. Section 3 establishes the properties of the penalty function, $W(\tau)$, and describes the methodology to explicitly solve SP, including a small numerical example. Section 4 outlines our computational experience. Section 5 explains our final conclusions. Appendix A has been included at the end of the paper containing proof of the results used throughout the paper.

2. Examples

We present a few examples where SP may arise as the subproblem of large-size models.

2.1. *Example 1: Capacity expansion under uncertainty*

Let us consider DM in (1.1). We establish the following capacity expansion problem (for exposition's clarity assume a single period): Demand is given by an absolutely continuous random variable (r.v.) D. F_D and f_D are its cumulative distribution and density functions, respectively. We first define l and m for our SP: if the range of D is bounded, we take $m \mid F(m) = 1$. Otherwise, we assume $m \mid F(m) = 1-\epsilon$ for some $\epsilon \ll 1$. The value for l is problem-dependent. We may let l take any representative value for r.v. D, e.g., the expected value $l = E(D)$.

Our risk function $W(.)$ for the problem represents the mathematical expectation of a suitable convex loss function $P_D(\tau, d)$ with respect to a probability measure (induced by the distribution of D). This is a well-known approach to the concept of Risk (e.g., De Groot, 1970). Here, $P_D(\tau, d)$ represents the loss if D takes the value d and our response is $\tau = \sum_i z_i < d$. We also assume $P_D(\tau, d) = 0$ if $\tau \geq d$. Therefore,

$$W(\tau) = E\left\{P_D(\tau, d)\right\} = \int_{\Re} P_D(\tau, s)\, \mathrm{d}F_D(s) = \int_{\Re+} P_D(\tau, s) f_D(s)\, \mathrm{d}s, \quad \forall \tau \geq 0.$$

In particular, for a convenient function $g(.)$ we may define P_D as follows:

$$P_D(\tau, d) = \begin{cases} g(d-\tau), & \text{if } d > \tau, \\ 0, & \text{if } 0 \leq d \leq \tau, \end{cases} \tag{2.1}$$

$g(\xi)$ must verify the following properties for $\xi = d - \tau > 0$: being non-negative, increasing and convex in τ for a fixed d. Some general boundary conditions are necessary as well. For example, we may take $P_D(\tau, d) = k(d - \tau)^n$ for $k > 0$ and $n \geq 1$. It is easy to verify that, with these premises, $W(\tau)$ verifies the necessary hypotheses that will be described later in Section 3.1.

2.2. *Example 2: Multi-resource systems*

Let us consider a system that may work under different conditions, or by employing different resources, $i = 1, \ldots, I$. Suppose that the system operates properly when the total (combined) amount of resources is $\tau = m$, where $\tau = \sum_{i=1}^{I} p_i z_i$ and p_i is the weight for the contribution of z_i units of resource i to the correct functioning of the system. Assume that the system can also be operated if $\tau \in [l, m)$, but in this case the operator "pays" an extra-cost $W(\tau)$. Here, i represents the different components that contribute to the system operation (e.g., fuel types for a thermal generating unit).

In a related example, suppose that i represents different types of maintenance (e.g., intensive, medium, and low). Each maintenance operation of type i improves the system performance (e.g., its safety or production rate) according to a ratio p_i. If we let z_1, z_2, and z_3 be the number of maintenance operations per unit of time, and establish a certain level, m, of preventive maintenance for a safe and optimal performance, we might estimate $W(\tau)$ for $\tau = p_1 z_1 + p_2 z_2 + p_3 z_3 < m$ as the expected cost of the corrective maintenance. Usually, $W(\tau) = W(m - \tau)$.

Strategic levels in these example problems arise when the available resources of type i depend on previous decisions, e.g, "How much fuel of type i must be procured?" and "How many qualified workers to perform maintenance type i should be available?"

2.3. *Example 3: Piecewise linearization*

SP can be partially adapted to some special cases of piecewise linearization of convex functions (Fourer, 1988; Kontogiorgis, 2000).

Suppose that z is a decision variable (z_t for more than one period or resource). Let us assume that the part of the objective function related to z in DM is formulated as follows:

$$\text{Minimize} \quad Q(z) + W(z), \tag{2.2}$$

where $W(z)$ is a suitable convex function and $Q(z)$ is an increasing convex function that can be approached by a piecewise linear function: $Q(z) \simeq \sum_{i=1}^{I} q_i z_i$, where I is the number of segments considered for the approach: $z = \sum_{i=1}^{I} z_i$, and the slopes q_i are increasing: $q_1 \leq \cdots \leq q_I$. Also,

$$0 \leq z_i \leq k_i, \tag{2.3}$$

where k_i is the length of the ith segment.

The model might consider lower and upper bounds (l, m, respectively) on z, leading to the new constraint:

$$l \leq \sum_{i=1}^{I} z_i \leq m. \tag{2.4}$$

It is well known that (2.2) can be approached by:

$$\text{Minimize } \sum_{i=1}^{I} q_i z_i + W(z), \quad \text{subject to (2.3) and (2.4),} \tag{2.5}$$

which defines a problem of type SP.

Similarly to previous examples, strategic decision variables x may exist, in which case (2.3) can be replaced by: $0 \leq z_i \leq k_i(\widehat{x})$ for each fixed $x = \widehat{x}$.

2.4. *Example 4: Transportation problem*

Let us consider the standard formulation of the transportation problem (TP) or market equilibrium model, in which a certain commodity is produced by I supply markets (producers) and is consumed by J demand markets (consumers):

$$\begin{aligned} \text{TP:} \qquad & \min_{x_{ij} \geq 0} \quad \sum_{i,j} c_{ij} x_{ij}, \\ & \text{subject to} \quad \sum_{j} x_{ij} = O_i, \ \forall i \quad \text{and} \quad \sum_{i} x_{ij} = D_j, \ \forall j. \end{aligned} \tag{2.6}$$

Dafermos and Nagurney (1989) propose a progressive equilibration algorithm by successively balancing each supply market or each demand market. To do this, they establish supply prices, π_i, and demand prices, ρ_j, and certain equilibrium conditions. This can also be viewed as solving the dual problem associated with the relaxation of producers or markets. For example, by relaxing the producer constraints, we obtain:

$$\max_{\pi_i \in \Re} \min_{x_{ij} \geq 0} \sum_{i,j} c_{ij} x_{ij} + \sum_{i} \pi_i \left(\sum_{j} x_{ij} - O_i \right), \quad \text{subject to} \quad \sum_{i} x_{ij} = D_j, \ \forall j. \tag{2.7}$$

Besides, if the market demand is uncertain, the first constraint in model (2.7) can also be moved into the objective function:

$$\max_{\pi_i \in \Re} \min_{x_{ij} \geq 0} \sum_{i,j} c_{ij} x_{ij} + \sum_{i} \pi_i \left(\sum_{j} x_{ij} - O_i \right) + \sum_{j} W_j \left(\tau_j^D \right),$$

where $\tau_j^D = \sum_i x_{ij}$ and $W_j(.)$ is a penalty. The inner minimization (for a fixed π) is separable into J submodels of type SP. Other constraints, such as bounds on the flow x_{ij} can also be incorporated.

3. Solving SP

In this section, we establish sufficient conditions on $W(.)$ that allow us develop a solving procedure for SP based on its KKTOCs.

3.1. *Hypothesis on the risk function* $W(.)$

Consider the SP model $\left(\text{recall that we assume } p_i = 1, \forall i, \text{ and } \tau = \sum_{i=1}^{I} z_i\right)$. We require $W(\tau)$ to satisfy the following *hypotheses*:

1. $W(.)$ is defined over the non-negative real line, $\Re^+$.
2. $W(\tau) = 0,\ \forall \tau \geq m$.
3. $W(\tau)$ is strictly decreasing, $\forall \tau \in (l, m)$.
4. $W(\tau)$ is strictly convex, $\forall \tau \in (l, m)$.
5. $W(\tau) \in \mathcal{C}^2[(l, m)]$ and $\lim_{\tau \to m}(dW/d\tau) = 0$.

(*Remark*: Some weaker assumptions than those in Hypothesis 4 and 5 ensure that the solution to the KKTOCs is optimal. In particular, it suffices that $W(\tau)$ is convex, and that $\lim_{\tau \to m}(dW/d\tau) = k \leq 0$. This will be shown in Appendix A.)

Hypotheses 1 and 2 are logical conditions to accommodate constraints $z_i \geq 0$ and $\tau \leq m$ in the context of our SP. Hypothesis 3 guarantees that the penalty decreases as τ approaches the maximum m. Hypothesis 4 requires convexity in order to guarantee optimality for the solution of the KKTOCs. Hypothesis 5 guarantees that the penalty function changes are "smooth" and bounded for large values of τ.

3.2. *Solving SP*

After incorporating dual variables, the problem to be solved can be stated as follows:

$$\text{SP:} \quad \min_{z_1,\ldots,z_I} \quad \sum_{i=1}^{I} q_i z_i + W\left(\sum_{i=1}^{I} z_i\right)$$

$$\text{s.t.} \quad \begin{cases} z_i \leq b_i, & \forall i = 1,\ldots,I\ (\mu_i), \\ \sum_{i=1}^{I} z_i \leq m, & (v), \\ \sum_{i=1}^{I} z_i \geq l, & (w), \\ z_i \geq 0, & \forall i = 1,\ldots,I\ (u_i). \end{cases} \tag{3.1}$$

We assume that SP has a feasible solution: The necessary and sufficient condition is $\sum_{i=1}^{I} b_i \geq l$. Without loss of generality we will assume that the decision variables z_i are arranged according to their coefficients in the objective function:

$$q_1 \leq q_2 \leq \cdots \leq q_I.$$

Notation remark: Henceforth, we shall denote $G_i(\tau) = \partial W(\tau)/\partial z_i, \forall i = 1,\ldots,I$, where $\tau = \sum_{i=1}^{I} z_i$. Taking into account that $G_1(\tau) = \cdots = G_I(\tau)$ we shall denote $G(\tau) = G_i(\tau), \forall i = 1,\ldots,I$.

The KKTOCs for SP (e.g., Bazaraa and Shetty, 1993), state that if z locally solves (3.1) then there exist scalars $u_i \geq 0, \forall i = 1, \dots, I; \mu_i \geq 0, \forall i = 1, \dots, I; v \geq 0; w \geq 0$ such that:

$$\text{KKTOC} \begin{cases} G(\tau) + q_i + \mu_i + v - w - u_i = 0, & \forall i = 1, \dots, I, \\ \mu_i(z_i - b_i) = 0, & \forall i = 1, \dots, I, \\ v(\tau - m) = 0, & \\ w(\tau - l) = 0, & \\ u_i z_i = 0, & \forall i = 1, \dots, I. \end{cases} \tag{3.2}$$

We shall attempt to determine a set of z_i, u_i, μ_i, v, w verifying the KKTOCs. Since SP is a convex problem, finding $u_i \geq 0, \forall i = 1, \dots, I, \mu_i \geq 0, \forall i = 1, \dots, I, v \geq 0$ and $w \geq 0$ that solve KKTOCs is a sufficient condition for optimality.

Next, we proceed to obtain explicitly the optimal solution to SP:

Definition 3.1. Let us consider the above-defined SP:

(a) We define the cumulative-resource vector as follows: $B = (B_1, \dots, B_I)$, where $B_r = \sum_{i=1}^{r} b_i$.
(b) We define the marginal index of SP as the largest index $r \in \{1, \dots, I\}$ such that $-q_r \geq G(B_r)$ (if it exists).

Remark 3.1. If $r < I$, it is clear that $-q_{r+1} < G(B_{r+1})$ holds.

Remark 3.2. $-q_r - G(B_r)$ decreases as r grows since q_r, B_r, and the function G are increasing. The meaning of the comparison between q_r and $G(B_r)$ aims to determine whether the linear cost q_r of the rth most-expensive resource offsets the risk $W(.)$ (viewed as a penalty for failing to satisfy the original constraint). This idea is clarified in the proof of Theorem 4 in the Appendix A.

Definition 3.2. We define the inverse function of $G(\tau)$ as $G^{-1}(Q') = p \in \Re \mid G(p) = Q'$.

Note that $G^{-1}(.)$ is well defined (by hypotheses 3 and 4). Since, by hypothesis 5, $\lim_{\tau \to m} G(\tau) = 0$, the domain of $G^{-1}(.)$ will be $Q' | G(0) \leq Q' < 0$. If $m < +\infty$, then $G(m) = 0$ and $Q' = 0$ will also be part of the domain of $G^{-1}(.)$. Without loss of generality, we extend the domain of $G^{-1}(.)$ by letting $G^{-1}(Q') = 0, \forall Q' < G(0)$.

Algorithm. Lemmas 1–2 and Theorems 1–4 (see Appendix A) provide an explicit way to obtain the optimal solution to SP by solving its KKTOC (3.2). Based on these results, we next explain the main steps of our algorithm to solve SP, which we call *SP algorithm*:

INPUT: q_i (in ascending order) and b_i, $\forall\, i = 1, \dots, I$. l, m, and implicit or explicit forms of the risk function $W(\tau)$, its derivative $G(\tau)$ and $G^{-1}(.)$

Step 0. Feasibility check: The necessary and sufficient condition is $B_I \geq l$. If SP in infeasible, STOP.

Step 1. Check (according to Definition 3.1(b)) whether a marginal index for SP exists, i.e., find the largest r such that $-q_r \geq G(B_r)$. If r does not exist, go to Step 2. Otherwise: If $r = I$, go to Step 3, and if $r < I$, go to Step 4.

Step 2. Compute $G^{-1}(-q_1)$. Use Theorem 2 (Appendix A) to obtain the optimal solution. STOP.

Step 3. Find the optimal solution by applying Theorem 3 (Appendix A). STOP.

Step 4. Compute $q_{r+1} + G(B_r)$ and use Theorem 4 (Appendix A). STOP.

OUTPUT: Primal and dual solution to SP: z_i, μ_i and u_i, $\forall\, i = 1, \ldots, I$; v and w.

3.3. *A numerical example*

We next show how to numerically implement our algorithm to solve a small instance of SP with $I = 10$ decision variables. This example is referenced as "sample index $s = 1000$" in our computational results (Sections 4.1 and 4.3). Data for this problem are as follows:

Vectors: $q_i = 10i$, $b_i = 25$, $\forall\, i = 1, \ldots, 10$.

Random variable: Our uncertain demand, D, is distributed as a symmetric triangular r.v. over the interval (100, 200). We take $l = 125$, $m = 200$.

Penalty and derived functions: The loss function considered is $P_D(\tau, d) = k(\tau - d)$ for $\tau = \sum_{i=1}^{10} z_i$. We take $k = s = 1000$. Thus, our risk function becomes: $W(\tau) = E_D\{P_D(\tau, d)\} = ((200 - \tau)^3/15)$, if $\tau \leq 200$, and $W(\tau) = 0$, if $\tau > 200$. Accordingly, $G(\tau) = (-(200 - \tau)^2/5)$, if $\tau \leq 200$, and $G(\tau) = 0$, if $\tau > 200$, and $G^{-1}(Q) = 200 - \sqrt{-5Q}$ for $Q \leq 0$.

Our instance of SP is as follows:

$$\text{SP:} \qquad \min_{z_1,\ldots,z_{10}} \quad 10z_1 + 20z_2 + \cdots + 100z_{10} + \frac{1}{15}(200 - z_1 - \cdots - z_{10})^3$$

$$\text{s.t.} \quad \begin{cases} z_i \leq 25, & \forall i\ (\mu_i), \\ z_i \geq 0, & \forall i\ (u_i), \\ z_1 + \cdots + z_{10} \leq 200, & (v), \\ z_1 + \cdots + z_{10} \geq 125, & (w). \end{cases}$$

We may apply the SP-algorithm from Section 3.2.

Marginal index calculation: We first obtain the cumulative resource vector: $B = (25, 50, \ldots, 250)$, and note that

$$-q_{10} = -100 < G(B_{10}) = G(250) = 0,$$

$$-q_9 = -90 < G(B_9) = G(225) = 0,$$

$$-q_8 = -80 < G(B_8) = G(200) = 0,$$

$$-q_7 = -70 \geq G(B_7) = G(175) = -125,$$

Thus, the marginal index is $r = 7$.

Case evaluation: Since $r = 7 < I = 10$ we apply Theorem 4 in Appendix A. To do so, we first compute $q_8 + G(B_7) = -45 < 0$. This falls into Case 3, which gives us the optimal solution for $\tau^* = \tilde{B}_7 = G^{-1}(-q_8) = G^{-1}(-80) = 180$.

Primal solution: We derive primal variables as follows:

$$z_1^* = 25, \ldots, z_7^* = 25, \quad z_8^* = 5, \quad z_9^* = 0, \quad z_{10}^* = 0,$$

whose cost is: 7400 (linear) $+(20^3/15)$ (penalty) $= 7933.33$ (total).

Dual solution: Similarly, we calculate dual variables:

$$\begin{aligned}
&\mu_1^* = q_8 - q_1 = 70, \ldots, \quad \mu_7^* = q_8 - q_7 = 10, \quad \mu_8^* = 0, \quad \mu_9^* = 0,\\
&\mu_{10}^* = 0, \quad u_1^* = 0, \ldots, \quad u_8^* = 0, \quad \mu_9^* = q_9 - q_8 = 10,\\
&\mu_{10}^* = q_{10} - q_8 = 20, \quad v = 0, \quad w = 0.
\end{aligned}$$

3.4. *Special case: all $b_i = +\infty$*

Consider the SP model where upper bounds on all individual primary variables are relaxed (i.e., $b_i = +\infty$, $\forall\, i$). The new model (which we call SP_∞) is as follows:

$$\text{SP}_\infty: \qquad \min_{z_1,\ldots,z_I} \quad \sum_{i=1}^{I} q_i z_i + W\left(\sum_{i=1}^{I} z_i\right)$$
$$\text{s.t.} \qquad \begin{cases} \sum_{i=1}^{I} z_i \le m & (v),\\ \sum_{i=1}^{I} z_i \ge l & (w),\\ \qquad z_i \ge 0, & \forall i = 1,\ldots,I;\ (u_i). \end{cases} \tag{3.3}$$

Since SP_∞ is a particular case of SP, it is clear that we may use the above-mentioned SP-algorithm to solve it (e.g., assuming ficticious bounds $b_i = M$, $\forall\, i$ for a large M with which we may carry out the computations).

It is clear that only the single variable with the least cost (i.e., z_1) will take a non-zero value. Thus, it is worth developing a special algorithm (see justification in Appendix A, Theorem 5) that provides that solution directly. The SP_∞ algorithm is as follows:

INPUT: $q_1 = \min\{q_i \mid i = 1, \ldots, I\}$. (*Remark*: The other q_i are irrelevant to the problem.) l, m, and implicit or explicit forms of the risk function $W(\tau)$, its derivative $G(\tau)$ and $G^{-1}(.)$

Step 1. If $G(l) < -q_1$, the solution is: $z_1 = G^{-1}(-q_1)$, $z_2 = \cdots = z_I = 0$ (primal) and $u_i = q_i - q_1, \forall i$ (dual). STOP.

Step 2. If $G(l) \ge -q_1$, the solution is: $z_1 = l$, $z_2 = \cdots = z_I = 0$ (primal) and $u_i = q_i - q_1, \forall i$ (dual). STOP.

OUTPUT: Primal and dual solution to SP_∞.

4. Computational Experience

4.1. *Description of test cases*

The hardware supporting our tests is an IBM A21p laptop with a 1 GHz Pentium-III processor and 512 Mb of RAM. SP has been used as submodel of large-scale stochastic problems like those reported in Marín and Salmerón (1998, 2001) . In particular, we have taken advantage of SP to solve capacity expansion planning models with a large number of periods, T, and subperiods, S, for each period.

At each iteration of Generalized Benders' Decomposition and Lagrangean Relaxation or Decomposition, a total of $T \cdot S$ subproblems were solved. In those works, demand was modeled using uniform, triangular, normal, and gamma probability distributions.

For the cases mentioned, the calculation of $G^{-1}(.)$ was obtained very efficiently. By employing $P_D(\tau, d) = k(\tau - d)$ as loss function, and $W(\tau) = E_D\{P_D(\tau, d)\}$, it is easy to verify that:

$$\frac{\partial W(\tau)}{\partial \tau} = G(\tau) = -k \int_{\tau}^{\infty} f_D(s)\,\mathrm{d}s = -k\,\Pr\,\{D > \tau\}, \quad \forall i = 1, \ldots, I, \tag{4.1}$$

which is a well-known function for many probability distributions. In this case, the computational burden to calculate $G^{-1}(.)$ is significantly simplified.

In this paper, we explore SP itself (and the techniques to solve it), independent of the application where it is used. For further reference, we next specify the details of the benchmark test-bed used in this work:

$$l = 125, \quad m = 200, \qquad \text{for a random demand } D \equiv \textit{Triangular}\ (100, 200; 150)$$

$$q_i = \frac{100\,i}{I}, \quad b_i = \frac{250}{I}, \qquad \forall i = 1, \ldots, I.$$

Notice that $\sum_{i=1}^{I} b_i = 250 \geq l$. Thus, SP is feasible. In fact, τ can reach its maximum ($m = 200$), if needed. In addition, each b_i is relatively small, and the contribution of a large number of variables is needed even for the smallest feasible value $\tau = l$.

Considering $P(\tau, d) = k(d - \tau)$, the aforementioned triangular distribution yields:

$$W(\tau) = \int_{\Re^+} P_D(\tau, s) f_D(s)\,\mathrm{d}s = \frac{k(200 - \tau)^3}{15{,}000}.$$

(The reader can easily use other probability distributions to create more complex risk functions while still taking advantage of (4.1) to easily apply the SP algorithm.)

In our different tests, we try $I = 10$, $I = 100$, $I = 1000$, and $I = 10{,}000$ variables per subproblem. We run a total of $n = 1, \ldots, 10{,}000$ samples for each value of I. At sample s, the risk function above is computed for $k = s$. This choice of k allow us to conduct an extensive testing of our algorithm. Depending on the sample index, $n = 1, \ldots, 10{,}000$, the optimal solution changes: the larger sample index, the larger value of τ, accounting for a lower tendency to accept risk as penalty increases. This tradeoff, in turn, ensures that most of the listed cases in our SP-algorithm (and especially those that require more computation) are covered by multiple samples, so as to assess the algorithm performance without being subordinated to a specific algorithm flow pattern. Solution to some of these cases is listed below for further reference.

A final detail on our implementation is that we assume linear costs already arranged as $q_1 \leq \cdots \leq q_i$ at the time we apply the SP-algorithm. This assumption is realistic in practice since, for example, if Benders' decomposition is applied to model DM, then all the linear costs associated with the subproblem can be sorted out at the beginning and will remain the same for all the subproblems (i.e., only the right-hand side, b_i, will change between iterations). (In the case of employing Lagrangean relaxation-decomposition, linear costs may change from one iteration to another. However, we also notice that as a result of dualizing the coupling constraints, the submodel will be of type SP_∞, which requires finding the minimum linear cost q_1 only.)

4.2. *Alternative solving methods*

We also solve model SP by using MINOS, as implemented in GAMS (Brooke *et al.*, 1998). In this case, the computational time used is that of solving the model only (given by the "model.resusd" variable in GAMS), which excludes generation and other overhead time.

We explore a second alternative based on parametric optimization (PO) (e.g., Kabadi and Aneja, 1997). SP lends itself to resolution by using the following scheme:

$$\min \ F(\tau), \quad \text{subject to } l \leq \tau \leq m,$$

where $F(\tau) = \sum_{i=1}^{I} q_i z_i + W(\tau)$, and z_i is easily derived as:

$$z_i = \begin{cases} b_i, & \text{if } \sum_{j=1}^{i} b_j \leq \tau, \\ \tau - \sum_{j=1}^{i-1} b_j, & \text{if } \sum_{j=1}^{i-1} b_j \leq \tau < \sum_{j=1}^{i} b_j, \\ 0, & \text{if } \sum_{j=1}^{i-1} b_j > \tau. \end{cases}$$

Thus, we can solve SP as a constrained one-dimensional problem with convex objective $F(\tau)$. The expensive step here is computing $F(\tau)$, since it entails determining the (first) variables for which $z_i > 0$, and their related cost. To simplify this task, we employ a dichotomous algorithm "Golden Search" (Bazaraa and Shetty, 1993) to find the optimal τ. This method has the advantage of evaluating $F(\tau)$ just once per iteration.

As opposed to our SP-algorithm, PO requires a tolerance (also known as length of the interval of uncertainty) so as to establish a stopping criterion, which must be specified beforehand. The values we use in our computations are tol $= 0.001$ and tol $= 1$. Since τ^* lies in the interval $[125, 200]$, solution errors are roughly bounded by 0.001 and 1%, respectively. It is important to underline the importance of obtaining optimal solutions to the subproblems, especially if Benders' decomposition is used: Otherwise, the cuts for the master problem might be inaccurate, with the potential risk of not being violated by the previous solution which makes convergence fail.

Table 1. Solution for some of the samples for different values of I.

I	Sample index (s)	Solution, τ	Linear cost, $C \cdot z$	Non-linear cost, $W(\tau)$	Total cost, F
10	1	125.0000	3750.0000	28.1250	3778.1250
	1000	180.0000	7400.0000	533.3333	7933.3335
	10,000	193.6754	8494.0352	168.6547	8662.6895
100	1	125.0000	3187.5000	28.1250	3215.6250
	1000	180.8950	6635.3364	464.8882	7100.2246
	10,000	193.7550	7605.3906	162.3697	7767.7603
1000	1	125.0000	3131.2500	28.1250	3159.3750
	1000	180.9737	6559.3467	459.1675	7018.5142
	10,000	193.7710	7519.1328	161.1221	7680.2549
10,000	1	125.0000	3125.3591	28.1250	3153.4841
	1000	180.9737	6551.7388	459.1675	7010.9063
	10,000	193.7742	7511.4771	160.8735	7672.3506

Table 2. Time to solve a total of $n = 1, \ldots, 10{,}000$ samples for SP.

I	KKT	MINOS	PO, tol = 0.001%	PO, tol = 1.0%
10	0.02	328	0.18	0.09
100	0.13	366	1.27	0.60
1000	1.20	373	12.17	5.68
10000	12.14	3073	119.27	56.63

4.3. *Results*

The solution for some of the samples and for different values of I is shown in Table 1. Table 2 shows the computational time to solve all the samples for a given value of I.

As expected, KKT and PO outperform MINOS. For KKT and PO, solving time exhibits linear growth in the submodel size (number of decision variables, I). KKT is about five times faster than PO when its tolerance level is set to 1%, and about ten times faster for a tolerance of 0.001%. Recall that the solution provided by KKT (our SP-algorithm) is exact.

5. Conclusions

We have presented an algorithm to explicitly solve a "submodel" that arises in decomposition of other large-scale models in system design. The algorithm is more efficient than other applicable state-of-the-art techniques.

This efficiency is based on the fact that the number of arithmetic operations is a linear function of the number of variables in the submodel.

An important benefit of the methodology proposed in this work is that it allows us to work with a variety of risk functions, $W(.)$, under relatively weak assumptions. In particular, the class of risk functions derived from the expectation of a given loss function with respect to generic probability distributions are used in the computational results.

The most important limitation to realize is that solving the proposed submodel via its KKTOCs complicates the extension of the algorithm to other submodels with more complex structures in the subproblem constraints. A future research area may be extending the applicability of this work to a broader range of constraints for the submodel.

Appendix A

Lemmas 1 and 2 along with Theorems 1–4 provide a constructive proof to attain the optimal solution to SP and its dual. Theorem 5 solves SP_∞. Hypotheses 1–5 for $W(.)$ (see Section 3.1) and Definitions 3.1 and 3.2 (see Section 3.2) are assumed throughout this appendix. The final section in the appendix addresses the validity of the results under weaker assumption on $W(.)$ (relaxation of Hypotheses 4 and 5).

Lemma 1. *The optimal solution to the dual of SP verifies:*

(a) $u_i \mu_i = 0, \ \forall i = 1, \ldots, I.$
(b) $(u_i - u_j) - (\mu_i - \mu_j) = q_i - q_j, \ \forall i, \ j = 1, \ldots, I.$

Proof.

(a) Suppose $u_i \neq 0$. Then, by the last equality in KKTOC, $z_i = 0$. Since $b_i > 0$, the second KKTOC implies $\mu_i = 0$.
(b) By the first KKTOC: $G(\tau) + v - w = u_i - q_i - \mu_i, \ \forall\, i = 1, \ldots, I$. Since $\tau = \sum_{i=1}^{I} z_i$, v and w do not depend on i, the above identity can only be verified for every i if $u_i - q_i - \mu_i = u_j - q_j - \mu_j$, $\forall\, i, j$. □

Corollary 1.

(a) *For each $i, j \in \{1, \ldots, I\}$ such that $q_i \leq q_j$ the inequalities $\mu_i \geq \mu_j$ and $u_i \leq u_j$ hold.*
(b) *If $u_i = u_j > 0 \Rightarrow q_i = q_j$.*
(c) *If $\mu_i = \mu_j > 0 \ \Rightarrow \ q_i = q_j$.*
(d) *If $q_i = q_j$ and $u_i, u_j > 0 \Rightarrow u_i = u_j$. If $q_i = q_j$ and $\mu_i, \mu_j > 0 \Rightarrow \mu_i = \mu_j$.*

Proof.

(a) Suppose $\mu_i > 0$. By Lemma 1(a), $u_i = 0$, and by Lemma 1(b) we have $\mu_i = q_j - q_i + \mu_j - u_j$. Since either μ_j or u_j must be zero, we consider two possible cases:

- If $\mu_j > 0 \Rightarrow u_j = 0$ and then $\mu_i \geq \mu_j$ (because $q_j - q_i \geq 0$).
- If $u_j > 0 \Rightarrow \mu_j = 0$ and then $\mu_i \geq \mu_j$ (because μ_i must be non-negative).

Let us consider the case $\mu_i = 0$. Again, by Lemma 1(a): $\mu_j = q_i - q_j + u_j - u_i$. Now, if $\mu_j > 0$ then $u_j = 0$, but this leads to $\mu_j = q_i - q_j - u_i \le 0 \Rightarrow \mu_j = 0$, which is a contradiction. Consequently, it must be $\mu_j = 0$ and $\mu_i \ge \mu_j$. Using analogous reasoning it may easily be proven that $u_i \le u_j$.

Proof for (b)–(d) follow trivially by Lemma 1 and the fact that $q_1 \le \cdots \le q_I$.

□

Theorem 1. *There exist two indices k, s such that $k \ge 0$, $s \ge 0$, $k \le s \le I$ and the optimal solutions to SP and its dual verify*:

(a)
$$\begin{array}{ccc} u_1 = 0 \cdots u_k = 0 & u_{k+1} = 0 \cdots u_s = 0 & 0 < u_{s+1} \le \cdots \le u_I \\ \mu_1 \ge \cdots \ge \mu_k > 0 & \mu_{k+1} = 0 \cdots \mu_s = 0 & \mu_{s+1} = 0 \cdots \mu_I = 0 \\ z_1 = b_1 \ge 0 \cdots z_k = b_k \ge 0 & z_{k+1} \ge 0 \cdots z_s \ge 0 & z_{s+1} = 0 \cdots z_I = 0 \end{array}$$

(b) *Moreover, if $i < j$ and $z_i < b_i$ then $z_j = 0$.*

Proof.

(a) First, let us consider two indices i, j, $i < j$, such that $q_i < q_j$, If $u_i > 0$, by Lemma 1(a) it follows $\mu_i = 0$. Note that $\mu_j = 0$, because if $\mu_j > 0$ then $u_j = 0$ and $\mu_j = q_i - q_j - u_i < 0$. On the contrary, if $u_j = 0$ and $\mu_j > 0$, then it would be necessary that $u_i = 0$. In fact, if $u_i > 0$ we have proved that $\mu_i = 0$ and therefore $u_i = q_i - q_j - \mu_j < 0$.

Suppose now $u_i = \mu_i = 0$. For the case being considered, it must be $\mu_j = 0$ and $u_j = q_j - q_i > 0$. Otherwise (i.e., if $\mu_j > 0$), $u_j = 0$ and then it would be $\mu_j = q_i - q_j - u_i = q_i - q_j < 0$. Likewise, if $u_j = \mu_j = 0$ then $u_i = 0$; $\mu_i = q_j - q_i$ because $u_i > 0$ implies that $\mu_i = 0$ and then $u_i = q_i - q_j + u_j = q_i - q_j < 0$.

We shall next consider i, j, $i < j$, where $q_i = q_j$. Then, if $u_i > 0 \Rightarrow \mu_i = 0$ (by Lemma 1(a)). Thus, by Lemma 1(b), $u_i = u_j$ and $\mu_j = 0$. Analogously, it can be easily proven that $u_j = 0$, $\mu_j > 0 \Rightarrow u_i = 0$, and $\mu_i = \mu_j$. Finally, since both u_i and μ_i cannot be positive, by Lemma 1(b) we have $u_i = \mu_i = 0$ and $u_j = \mu_j = 0$.

Once the structures for u and μ have been determined, the structure for z follows by applying the KKTOC $u_i z_i = 0$.

(b) First, we note that if $q_i = q_j$ and an optimal solution has $z_i < b_i$ then there exists an optimal solution with either $z_j = 0$ or $z_i = b_i$. In order to prove this, we consider the optimal $z_i + z_j = \rho$. For any feasible z_i, z_j such that $z_i + z_j = \rho$ the primal objective function does not change given that $q_i = q_j$ and $W(.)$ is a function of $z_i + z_j$. Therefore, we can take $z_i = \rho$, $z_j = 0$ if $\rho \le b_i$ or $z_i = b_i$, $z_j = \rho - b_i$ if $\rho > b_i$. Thus, we may (and will) consider, without loss of generality, that if $i < j$ and $q_i = q_j$, then, at the optimal solution, $z_j > 0$ may occur only if $z_i = b_i$.

The result is easily extended for the case $q_i < q_j$. Also, working with the dual problem, $q_i < q_j$ implies $\mu_i = 0$ and then $\mu_j = 0$. Again, if $z_j \neq 0 \Rightarrow u_j = 0$.

Thus, by Lemma 1(b) we have $u_i - q_i = -q_j \Rightarrow u_i = q_i - q_j < 0$, which is a contradiction. □

Lemma 2. *If the marginal index r does not exits (i.e., $-q_1 < G(B_1)$) then*:

(a) $v = 0$.
(b) If $-q_1 \geq G(l)$ *then* $w = 0$.

Proof.

(a) We shall show that $v > 0$ contradicts the hypothesis.
By KKTOC, if $v > 0 \Rightarrow \tau = m$. Since $l < m = \tau$, it follows $w = 0$. Applying the stationary condition in KKTOC, $0 = G(m) = u_1 - q_1 - \mu_1 - v$. We will see that this equation leads to an inconsistency in the value of u_1: if we let $u_1 > 0$, then, by Theorem 1(a), $u_1 > 0 \Rightarrow u_j > 0$, $\forall j = 2, \ldots, I \Rightarrow z_i = 0$, $\forall i = 1, \ldots, I$. But if this is true, then $\tau = 0 < m$, in contradiction with $\tau = m$. On the other hand, if we take $u_1 = 0$, then $-q_1 - \mu_1 - v = 0$. This is also a contradiction since $q_1 \geq 0$, $\mu_1 \geq 0$ and $v > 0$. The conclusion is that $v = 0$.

(b) We consider two cases:

Case 1. $B_1 = b_1 \leq l$. Let us prove that this case is impossible: since G is an increasing function we have $G(B_1) \leq G(l)$. However, by the hypothesis $G(l) \leq -q_1$, the inequality $G(B_1) \leq -q_1$ must also hold. This contradicts the initial hypothesis $G(B_1) > -q_1$.

Case 2. $B_1 = b_1 > l$. Suppose $w > 0$. Then, $\tau = l$. We will prove that this leads to a contradiction in the value of u_1. By KKTOC, we have $w = \mu_1 + q_1 + G(l) - u_1$. Since u_1 and μ_1 cannot be strictly positive at the same time, and $q_1 + G(l) < 0$, it is necessary that $u_1 = 0$ and $\mu_1 > 0$ to maintain the assumption $w > 0$. However, $\mu_1 > 0 \Rightarrow z_1 = b_1$, which is impossible since $z_1 \leq \tau = l < b_1$. □

Theorem 2. *If* $-q_1 < G(B_1)$ *then the optimal solution to SP is provided by one of the following cases*:

(a) *If* $-q_1 \geq G(l)$ *then the optimal solution is*:

$$z_1 = G^{-1}(-q_1); \quad z_2 = \cdots = z_I = 0;$$

$$u_1 = 0; \quad u_j = q_j - q_1, \quad \forall j = 2, \ldots, I;$$

$$\mu_1 = \cdots = \mu_I = 0;$$

$$v = w = 0.$$

(b) *If* $-q_1 < G(l)$ *then let* k *be such that* $B_{k-1} < l \leq B_k$ *(taking* $B_0 = 0$ *if necessary). The optimal solution is*:

$$z_1 = b_1, \dots, z_{k-1} = b_{k-1}, \quad z_k = l - B_{k-1} \leq b_k; \quad z_j = 0, \quad \forall j = k+1, \dots, I;$$

$$\mu_{k+1} = \cdots = \mu_I = 0; \quad \mu_j = q_k - \mu_k - q_j, \quad \forall j < k;$$

$$u_1 = \cdots = u_k = 0; \quad u_j = q_j - q_k - \mu_k, \quad \forall j > k;$$

$$v = 0; \quad w = G(l) + q_k + \mu_k.$$

Moreover, if $B_k > l$ *then* $\mu_k = 0$*. Otherwise, if* $B_k = l$*, the dual problem has an optimal solution for each* $\mu_k \in [0, q_k - q_{k-1}]$*.*

Proof.

(a) By Lemma 2(a,b), $v = w = 0$. Next, we shall prove that there exists a feasible solution to SP (and its dual) verifying KKTOC. Consider the identity:

$$\mu_i = u_i - q_i - G(\tau), \quad \forall i = 1, \dots, k.$$

Here, it must be $\mu_1 = 0$: If not, $\mu_1 > 0 \Rightarrow u_1 = 0$ (Lemma 1(a)) and then $z_1 = b_1$ by KKTOC. In this case, $\mu_1 = -q_1 - G(\tau)$. Since $\tau = \sum_{i=1}^{I} z_i \geq z_1 = b_1$ and $G(\tau)$ is increasing, $\mu_1 \leq -q_1 - G(b_1) < 0$ (by hypothesis). Thus, $\mu_1 = 0$. Now, by Theorem 1(a), $\mu_2 = \cdots = \mu_I = 0$.

Let us calculate the values for the components of u: $u_j = u_1 + q_j - q_1, \forall j = 2, \dots, I$. Here, $u_1 = 0$. Otherwise, $u_i > 0$, $\forall i = 1, \dots, I$, and then $z_i = 0$, $\forall i = 1, \dots, I \Rightarrow \tau = 0$. Then, $G(\tau) = u_1 - q_1 \Rightarrow G(0) = u_1 - q_1$. Notice that $\tau = 0$ is feasible only when $l = 0$, but by the original hypothesis in this theorem we have $G(l) = G(0) \leq -q_1$. Thus, it follows that $G(0) = u_1 - q_1$ (where $u_1 > 0$), which is a contradiction. By KKTOC and the hypothesis of this theorem

$$u_1 = 0 \Rightarrow G(\tau) = -q_1 \Rightarrow \tau = G^{-1}(-q_1) < b_1.$$

Applying hypothesis (a) we have $\tau \geq l$. Also, $\tau \leq m$ because

$$-q_1 \leq 0 \Rightarrow G^{-1}(-q_1) \leq G^{-1}(0) = m,$$

given that G is strictly increasing.

(b) By Lemma 2(a), $v = 0$. By a similar argument as that in case (a), and taking into account that $\tau = G^{-1}(-q_1) < l$ is not feasible, we have to consider the first index k such that $\sum_{i=1}^{k} z_i = l$ for feasible values of z_i. This is clearly achieved by the following forward scheme:

$$0 < z_1 = b_1; \;\; 0 < z_2 = b_2; \;\; \cdots \;\; 0 < z_k = l - \sum_{i=1}^{k-1} z_i \leq b_k;$$

$$\text{and} \quad z_{k+1} = \cdots = z_I = 0.$$

By Theorem 1(a):

$$\mu_{k+1} = \cdots = \mu_I = 0.$$

Since $z_1 > 0, \dots, z_k > 0$, we also have $u_1 = \cdots = u_k = 0$. We calculate w using the identities $w = G(l) + q_i + \mu_i - u_i,\ \forall i = 1, \dots, I$. In particular, for $i = k$ we have:

$$w = G(l) + q_k + \mu_k.$$

The remaining components of μ and u are obtained by employing Lemma 1, yielding:

$$u_j = q_j - q_k - \mu_k, \quad \forall j > k,$$

$$\mu_j = q_k + \mu_k - q_j, \quad \forall j < k.$$

Note that the value of μ_k depends on z_k. If $z_k < b_k$ then $\mu_k = 0$. If $z_k = b_k$ (this only occurs if $\sum_{i=1}^{k} b_k = l$) then there exist an infinite number of suboptimal solutions by selecting any $\mu_k \in [0, q_k - q_{k-1}]$ (notice that feasibility of μ_{k-1} in the above expression is preserved). □

Theorem 3. *If the marginal index is $r = I$ (and so, $-q_I \geq G\ (B_I)$), the optimal solution to the primal and dual of SP is:*

$$z_i = b_i, \quad u_i = 0, \quad \forall i = 1, \dots, I;$$

$$\mu_j = -G(B_I) - q_j + w, \quad \forall j = 1, \dots, I;$$

$$v = 0.$$

Moreover, if the primal problem does not have a unique feasible solution, then $w = 0$. Otherwise, there exist an infinite number of optimal solutions to the dual SP for any $w \geq 0$.

Proof. Since $G(m) = 0$, $q_I > 0$ and G is a strictly increasing function, in order to satisfy $G(B_I) + q_I \leq 0$ we require $B_I < m$. Thus, $\tau \leq B_I < m \Rightarrow v = 0$.

Next, we prove that $z_i = b_i, \forall i = 1, \dots, I$. Suppose that, for some j, $z_j < b_j$. Then, $G(\tau) + q_j < 0$ because:

$$\begin{cases} G \text{ is strictly increasing,} \\ q_j < q_I, \\ G(B_I) + q_I \leq 0 \text{ by hypothesis.} \end{cases}$$

So, if $z_j < b_j$ the second KKTOC implies that $\mu_j = 0$. This makes the first KKTOC for subscript j become $G(\tau) + q_j + \mu_j + v - w - u_j = 0$, which is impossible because $G(\tau) + q_j < 0$, $\mu_j = v = 0$, $w \geq 0$ and $u_j \geq 0$.

Therefore, it has to be $z_i = b_i$, $\forall i$. Applying Theorem 1 it follows that $u_i = 0$, $\forall i = 1, \dots, I$.

To obtain w, we note that if $\tau = \sum_{i=1}^{I} z_i = \sum_{i=1}^{I} b_i > l$then $w = 0$. It follows:

$$\mu_I = -G(B_I) - q_I, \quad \text{(by the first KKTOC)},$$

$$\mu_j = \mu_I + q_I - q_j = -G(B_I) - q_j, \quad \forall j = 1, \dots, I-1 \quad \text{(by Lemma 1(b))}.$$

As in Theorem 2, there exists a degenerate case for which the dual problem has infinity many suboptimal solutions: if $\sum_{i=1}^{I} b_i = l$, then, the primal SP has only one feasible solution (and so, optimal): $\tau = l$. However, any $w \geq 0$ provides an optimal dual solution:

$$\mu_I = -G(B_I) - q_I + w,$$

$$\mu_j = \mu_I + q_I - q_j = -G(B_I) - q_j + w, \quad \forall j = 1, \ldots, I-1.$$

□

Theorem 4. *If the marginal index verifies* $1 \leq r \leq I-1$ *(i.e.,* $-q_r \geq G(B_r)$ *and* $-q_{r+1} < G(B_{r+1})$*), then, the optimal solution to SP depends on which of the following four exclusive cases occurs*:

$$\begin{cases} \textit{Case 1}: & q_{r+1} + G(B_r) \geq 0 \textit{ and } l \leq B_r. \\ \textit{Case 2}: & q_{r+1} + G(B_r) \geq 0 \textit{ and } B_r < l < B_{r+1}. \\ \textit{Case 3}: & q_{r+1} + G(B_r) < 0 \textit{ and } l < B_{r+1}. \\ \textit{Case 4}: & B_{r+1} \leq l. \end{cases}$$

The explicit values of the primal and dual optimal solutions are derived and presented in the proof.

Proof. Analogously to Theorem 3, the hypothesis $-q_r \geq G(B_r)$ and the fact that $G(m) = 0$, $q_r > 0$, and G strictly increasing ensure that $B_r < m$. Now, let us study each case individually:

Case 1: By the first KKTOC, it is true that:

$$G(\tau) + q_i + \mu_i + v - w - u_i = 0, \quad \forall i = 1, \ldots, I \tag{A.1}$$

We first prove that $\tau = B_r$.

If we allow $\tau > B_r$, then $\tau > l \Rightarrow w = 0$ and:

$$\tau > B_r \Rightarrow G(\tau) > G(B_r) \Rightarrow G(\tau) + q_{r+1} > 0.$$

In addition, by Theorem 1, $\tau > B_r \Rightarrow z_{r+1} > 0$.

This implies (following Theorem 1(a)) that $z_{r+1} = 0 \Rightarrow u_{r+1} = 0$. However, it is clear that Eq. (A.1) for $i = r+1$ could not be held, which rules out the hypothesis $\tau > B_r$. If we allow $\tau < B_r$ then $G(\tau) < G(B_r) \Rightarrow G(\tau) + q_r < 0$, and $\tau < B_r < m \Rightarrow v = 0$. Also, $\tau < B_r \Rightarrow z_r < b_r \Rightarrow \mu_r = 0$ (because if $z_r = b_r \Rightarrow z_1 = b_1, \ldots, z_r = b_r \Rightarrow \tau \geq B_r$ by Theorem 1(b)). Again, Eq. (A.1) cannot be verified, in this case for index $i = r$. Thus, we rule out $\tau < B_r$.

The only feasible solution is, therefore, $\tau = B_r$. Now, by Theorem 1(b):

$$z_j = b_j, \ \forall j = 1, \ldots, r; \quad z_j = 0, \ \forall j = r+1, \ldots, I.$$

Moreover, since $\tau = B_r < m$ it will be $v = 0$.

The optimal values of the dual variables are easily obtained by KKTOC and Lemma 1:

$$u_j = 0, \quad \forall j = 1, \ldots, r;$$

$$\mu_r = -G(B_r) - q_r + w;$$

$$\mu_j = \mu_r + (q_r - q_j) = -G(B_r) - q_j + w, \ j = 1, \ldots, r-1; \quad \mu_j = 0, \ \forall j = r+1, \ldots, I;$$

$$u_j = q_j + G(B_r) - w, \quad \forall j = r+1, \ldots, I.$$

The value of w is obtained as follows: if $l < B_r = \tau \Rightarrow w = 0$. Likewise, for Theorems 2 and 3, the dual optimal solution may not be unique. This happens when $l = B_r$. Then, w may be any value such that $w \in [0, q_{r+1} + G(B_r)]$.

Case 2: The proof is analogous to Case 1, but we take into account that $\tau > B_r$ (because $l \leq B_r$ is not feasible).

The optimal solution will be achieved for the nearest point in the boundary $\tau = l$, that is

$$z_1 = b_1, \ldots, z_r = b_r; \quad z_{r+1} = l - B_r < b_{r+1}; \quad z_{r+2} = \cdots = z_I = 0.$$

We also obtain the value of vector u:

$$u_j = 0, \ \forall j = 1, \ldots, r+1; \quad u_{r+j} = q_{r+j} + G(\tau) - w, \ \forall j = 1, \ldots, I - r.$$

We use the first KKTOC to obtain:

$$u_{r+1} = q_{r+1} + G(\tau) - w$$

and w is calculated as $w = q_{r+1} + G(\tau)$.

Remark: Note that, since $\tau \geq B_r$ and, by hypothesis, $w = q_{r+1} + G(\tau) > q_{r+1} + G(B_r) \geq 0$, the resulting w is feasible. Also, $v = 0$ (because $\tau = l < m$).

Finally, we obtain the value of the components of μ:

$$z_{r+1} < b_{r+1} \Rightarrow \mu_{r+1} = \cdots = \mu_I = 0,$$

$$\mu_j = u_j + w - v - q_j - G(\tau) = q_{r+1} - q_j, \quad \forall j = 1, \ldots, r.$$

Case 3: For this case we make a similar reasoning as in Case 1. The only difference is that the value obtained for the dual variable $u_{r+1} = q_{r+1} + G(\tau)$ where $\tau = B_r$ (in Case 1) is no longer feasible (notice that the new hypothesis implies that this value is strictly less than 0).

To achieve feasibility, it is necessary to increase the value of τ to a certain value $\tau = \tilde{B}_r = B_r + \triangle$ so as to restore dual feasibility for u_{r+1} (i.e., $u_{r+1} = 0$). To accomplish this task, we must solve the equation

$$q_{r+1} + G(\tilde{B}_r) = 0$$

for $\tilde{B}_r$. This entails computing $\tilde{B}_r = G^{-1}(-q_{r+1})$. Notice that $\tilde{B}_r$ verifies $\tilde{B}_r \in (B_r, B_{r+1})$ because $q_{r+1} + G(B_r) < 0$, $q_{r+1} + G(B_{r+1}) > 0$, and G is a continuous function.

If $l < \tilde{B}_r \Rightarrow w = 0$ and the optimal solution for all primal and dual variables will be:

$$\mu_{r+1} = 0; \quad z_{r+1} = \tilde{B}_r - B_r = \delta; \quad u_{r+1} = 0;$$

$$\mu_r = -G(\tilde{B}_r) - q_r = q_{r+1} - q_r; \quad z_r = b_r; \quad u_r = 0;$$

$$\mu_j = \mu_r + q_r - q_j = q_{r+1} - q_j; \quad z_j = b_j; \quad u_j = 0, \qquad \forall j < r;$$

$$\mu_j = 0; \quad z_j = 0; \quad u_j = q_j - q_{r+1}, \qquad \forall j > r + 1.$$

If $l \geq \tilde{B}_r$, it will be necessary to proceed as in Case 2, by setting:

$$z_{r+1} = l - B_r < b_{r+1},$$

which is feasible given that, by hypothesis, $l < B_{r+1}$.

The remaining primal variables are:

$$z_j = 0, \quad \forall j > r + 1.$$

Case 4: $B_{r+1} \leq l$. Analogously to Theorem 2 (Case 2), the optimal solution to the primal problem is achieved for $\tau = l$. This is obtained by finding the index s such that:

$$z_1 = b_1, \ldots, z_{s-1} = b_{s-1}, \quad z_s = l - B_{s-1} \leq b_s$$

and proceeding as in that case to obtain the dual variables. □

Theorem 5. *Let us consider* SP_∞. *Let* $\tau = \sum_{i=1}^{I} z_i$ *and* $G(\tau) = \partial W(\tau)/\partial z_1 = \cdots = \partial W(\tau)/\partial z_I$. *Suppose (without loss of generality) that* $q_1 \leq \cdots \leq q_I$. *An optimal solution to this problem can be obtained as follows*:

Case 1: *If* $G(l) < -q_1$, *the solution is*: $z_1 = G^{-1}(-q_1)$; $z_2 = \cdots = z_I = 0$.
Case 2: *If* $G(l) \geq -q_1$, *the solution is*: $z_1 = l$; $z_2 = \cdots = z_I = 0$.

Proof. If we write the KKTOCs for the proposed problem, we need to find $I + 2$ non-negative dual variables $u, v, (u_i)_{i=1,\ldots,I}$ such that:

$$\begin{cases} \frac{\partial W(\tau)}{\partial z_i} + q_i + v - w - u_i = 0, & \forall i = 1, \ldots, I, \\ v\left(\sum_{i=1}^{I} z_i - m\right) = 0, & \\ w\left(\sum_{i=1}^{I} z_i - l\right) = 0, & \\ u_i z_i = 0, & \forall i = 1, \ldots, I. \end{cases}$$

Case 1: We have $G(\tau) = u_i + w - v - q_i$, $\forall i = 1, \ldots, I$. Note that τ, w and v do not depend on i. Thus, $u_j = u_i + q_j - q_i$ hold for any $i, j = 1, \ldots, I$. If $j > i$ then $q_j > q_i \Rightarrow u_j > u_i$.

We shall next prove that $u_1 = 0$. If $u_1 > 0$ then $u_j > 0$, $\forall j = 1, \ldots, I$. Since:

$$u_i z_i = 0, \quad \forall i = 1, \ldots, I \Rightarrow z_i = 0, \quad \forall i = 1, \ldots, I$$

it is clear that the constraint $\tau \geq l$ cannot be held. Thus, it must be $u_1 = 0$, and then

$$0 = u_1 = u_2 = \cdots = u_k \quad \text{for } q_1 = q_2 = \cdots = q_k,$$

where k is the first subscript such that $q_k < q_{k+1}$ (we may assume $k = 1$ without loss of generality).

The remaining variables take the values:

$$u_l = q_l - q_1 > 0 \Rightarrow z_l = 0, \quad \forall l = 1, \ldots, I - k.$$

Next, we will prove that $v = w = 0$. Since $l < m$, by the optimality conditions it is clear that either $v = 0$ or $w = 0$ (or both).

Suppose $v = 0$. Then, $G(\tau) = w - q_1$. But, if $w > 0$ then

$$w > 0 \Rightarrow \tau = l \Rightarrow G(l) = G(\tau) = w - q_1 > -q_1,$$

which is a contradiction with the hypothesis $G(l) < -q_1$.

On the other hand, if $w = 0$, then $v > 0$ cannot be satisfied because it would imply $\tau = m$, which is impossible given that $G(m) = 0$, and $0 = v + q_1$ could not be held. Thus, it is necessary that $v = w = 0$. This leads to

$$G(\tau) = -q_1 \Rightarrow \tau = G^{-1}(-q_1) = z_1.$$

Case 2: The arguments to prove the result in this case are almost the same as in Case 1. The only difference with that is the way to obtain w. Now, $w = 0$ is not justified because the hypothesis implies $G(l) \geq -q_1$. Instead, the value of w is obtained by considering:

$$G(l) = G(\tau) = w - q_1 \ \Rightarrow \ w = G(\tau) + q_1 \geq 0,$$

which yields $z_1 = l;\ z_2 = \cdots = z_I = 0$ as optimal solution. □

Extensions of Hypotheses 4 and 5 for $W(.)$

Hypothesis 4: $W(.)$ *convex* (*but not strictly*): If $W(.)$ is not strictly convex, we cannot ensure $G(\tau)$ is strictly increasing. Hence, although all proofs throughout the appendix remain true, we can find alternative solutions when computing $G^{-1}(-Q)$ for some values of $-Q < 0$ (i.e., $G^{-1}(.)$ is not uniquely defined and some criterion must be adopted when the inverse is not unique, for example, the least τ such that $G(\tau) = -Q$).

Hypothesis 5: $\lim_{\tau \to m} (dW/d\tau) = \lim_{\tau \to m} G(\tau) = k \leq 0$:

Lemma 1 and Theorems 1, 4, and 5: No changes are needed.

Lemma 2: In the proof of part (a), we may change "$G(m) = 0$" by "$G(m) \leq 0$."

Theorem 2: In Theorem 2(a), $G^{-1}(-q_1)$ is still well-defined given that $-q_1 \leq G(B_1)$ by hypothesis.

Theorem 3: Proof needs some tuning to accommodate non-strict convexity. The reason is that we cannot define $G^{-1}(-Q)$ for $0 < -Q < k$. Note that in that proof, changing "$G(m) = 0$" by "$G(m) \leq 0$" does not guarantee that the thesis $B_I < m$ will hold. The reason is that it may occur as $B_I = m$ and $G(B_I) + q_I = -k + q_I \leq 0$.

We distinguish two cases: If $B_I < m$ no changes are needed. However, if $B_I = m$, in order to attain feasibility for constraint $\tau \leq m$ it is necessary that $\tau = m$ holds. Thus, we need to modify our solution as follows:

$$z_1 = b_1, \ldots, z_{I-1} = b_{I-1}, \quad z_I = m - (b_1 + \cdots + b_{I-1}),$$

$$u_i = 0, \quad \forall\, i = 1, \ldots, I.$$

See that $w = 0$ because $\tau = m > l$. Also, $v = -q_I - G(\tau)$, which is feasible $(v \geq 0)$ because $-q_I - G(m) \geq -q_I - G(B_I) \geq 0$.

Acknowledgments

The authors wish to thank two anonymous referees for their valuable reviews and suggestions, which improved the content and exposition of this paper.

References

Álvarez, M, C Cuevas, L Escudero, J De la Fuente, C García and F Prieto (1994). Network planning under uncertainty with an application to hydropower generation. *Top*, 2(1), 25–58.

Bazaraa, M and C Shetty (1993). *Nonlinear Programming* New York: J. Wiley & Sons, Inc.

Bean, J, J Higle and R Smith (1992). Capacity expansion under stochastic demands. *Operations Research*, 40(S2), S210–S216.

Bloom, J (1983). Solving an electricity generation expansion planning problem by generalized Benders' decomposition. *Operations Research*, 31(1), 84–100.

Brooke, A, D Kendrick and A Meeraus (1998). *Gams, a User's Guide*. California: The Scientific Press.

Cornuejols, G, G Nemhauser and L Wolsey (1990). The uncapacitated facility location Problem. In *Discrete Location Theory*, PB Mirchandani and RL Francis (eds.) New York: Wiley-Interscience.

Current, J, S Ratick and Ch. ReVelle (1998). Dynamic facility location when the total number of facilities is uncertain: a decision analysis approach. *European Journal of Operational Research*, 110, 597–609.

Dafermos, S and A Nagurney (1989). Supply and demand equilibration algorithms for a class of market equilibrium problems. *Transportation Science*, 23, 118–124.

Daskin, M, W Hopp and B Medina (1992). Forecast horizons and dynamic facility location planning. *Annals of Operations Research*, 40, 123–152.

De Groot, M (1970). *Optimal Statistical Decisions*. New York: McGraw-Hill.

Drezner, Z (1995). Dynamic facility location: the progressive p-median problem. *Location Science*, 3, 1–7.

Fourer, R (1988). A simplex algorithm for piecewise-linear programming II: finiteness, feasibility and degeneracy. *Mathematical Programming*, 41, 281–315.

Ferland, J, V Gascon and A Benchakroun (1991). Benders decomposition for network design problems. Research Report University of Montreal. Publication #762.
Fisher, M (1981). The lagrangian relaxation method for solving integer programming problems. *Management Science*, 27(1), 1–18.
Geoffrion, A (1972). Generalized Benders decomposition. *Journal of Optimization Theory and Applications*, 10(4), 237–260.
Guignard, M and S Kim, (1987). Lagrangian decomposition: a model yielding stronger lagrangian bounds. *Mathematical Programming*, 39, 215–228.
Hobbs, B (1995). Optimization methods for electric utility resource planning. *European Journal of Operational Research*, 83, 1–20.
Jack, C, S Kai and A Shulman (1992). Design and implementation of an interactive optimization system for telephone network planning. *Operations Research*, 40(1), 14–25.
Jacobsen, S (1990). Multiperiod capacitated location models. In *Discrete Location Theory*, PB Mirchandani and RL Francis (eds.). New York: Wiley-Interscience.
Kabadi, S and Y Aneja (1997). An efficient, strongly polynomial, ϵ-approximation parametric optimization scheme. *Information Processing Letters*, 64, 173–177.
Kontogiorgis, S (2000). Practical piecewise-linear approximation for monotropic optimization. *INFORMS Journal on Computing*, 12(4), 324–340.
Laguna, M (1998). Applying robust optimization to capacity expansion of one location in telecommunications with demand uncertainty. *Management Science*, 44(11, Part 2 of 2), S101–S110.
Magnanti, T and R Wong (1990). Decomposition methdos for facility location problems. In *Discrete Location Theory*, PB Mirchandani and RL Francis (eds.). New York: Wiley-Interscience.
Malcom, S and S Zenios (1994). Robust optimization for power systems expansion under uncertainty. *Journal of Operational Research*, 45(9), 1040–1049.
Marín, A and J Salmerón (1998). Electric capacity expansion under uncertain demand: decomposition approaches. *IEEE Transactions on Power Systems*, 13(2), 333–339.
Marín, A and J Salmerón (2001). A risk function for the stochastic modeling of electric capacity expansion. *Naval Research Logistics*, 48(8), 662–683.
Murphy, F and Z Wang (1993). A network reformulation of an electric utility expansion planning model. *Naval Research Logistics*, 40, 451–457.
Sherali, D and K Staschus (1990). A two-phase decomposition approach for electric utility capacity expansion planning including nondispatchable technologies. *Operations Research*, 38(5), 773–791.
Verter, V and M Cemal (1995). Facility location and capacity acquisition: an integrated approach. *Naval Research Logistics*, 42, 1141–1160.

Javier Salmerón is a Research Assistant Professor in the Operations Research Department at the Naval Postgraduate School in Monterey, California. He achieved his PhD in Mathematics from Universidad Politécnica de Madrid in 1998. His research interests lie in the area of applied integer and stochastic programming.

Ángel Marín is a full Professor at the Department of Statics and Applied Mathematics of the Polytechnic University of Madrid. He achieved his PhD in Aeronautical Engineering from this university in 1982. His research interests include

decomposition methods, bilevel programming, traffic management (urban, rail, air, etc.), energy generation, telecommunications, local access, and wireless networks. He has published in several journals such as *Transportation Research*, *Networks*, *IEEE Transactions on Power Systems*, *European Journal of Operational Research*, and *International Transactions on Operational Research.*