OPERATIONS RESEARCH
Vol. 57, No. 5, September–October 2009, pp. 1250–1261
ISSN 0030-364X | EISSN 1526-5463 | 09 | 5705 | 1250

DOI 10.1287/opre.1080.0649

Strategic Planning with Start-Time Dependent Variable Costs

James Luedtke
Department of Industrial and Systems Engineering, University of Wisconsin–Madison,
Madison, Wisconsin 53706, jrluedt1@wisc.edu

George L. Nemhauser
H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology,
Atlanta, Georgia 30332, gnemhaus@isye.gatech.edu

We present a strategic planning model in which the activities to be planned, such as production and distribution in a supply network, require technology to be installed before they can be performed. The technology improves over time, so that a decision maker has incentive to delay starting an activity to take advantage of better technology and lower operational costs. The model captures the fundamental trade-off between delaying the start time of an activity and the need for some activities to be performed now. Models of this type are used in the oil industry to plan the development of oil fields. This problem is naturally formulated as a mixed-integer program with a bilinear objective. We develop a series of progressively more compact mixed-integer linear formulations, along with classes of valid inequalities that make the formulations strong. We also present a specialized branch-and-cut algorithm to solve an extremely compact concave formulation. Computational results indicate that these formulations can be used to solve large-scale instances, whereas a straightforward linearization of the mixed-integer bilinear formulation fails to solve even small instances.

Subject classifications: integer programming: theory and applications; facilities/equipment planning: capacity expansion; technology.
Area of review: Optimization.
History: Received February 2007; revision received January 2008; accepted June 2008. Published online in *Articles in Advance* March 30, 2009.

1. Introduction

We study a multiple-period strategic planning model in which the activities to be planned, such as production at supply points, have technology-dependent costs. Specifically, we assume that technology must be installed before an activity can be performed, and that once installed it cannot be changed for the remainder of the planning horizon. In general, technology improves over time so that a decision maker has incentive to delay installation to benefit from reduced operating costs. On the other hand, immediate requirements such as customer demands may require some activity to be performed now. In this paper, we present a planning model that captures this basic trade-off.

This research was motivated by a strategic planning model in the upstream oil and gas industry in which a firm is planning the long-term development of oil fields and transportation modes between the fields and downstream processing facilities. This development involves the installation of facilities that cost billions of dollars. A generic version of this application consists of a production and distribution planning problem over a multiple-period planning horizon. In the motivating application, the time periods are years and the time horizon is about 30 years, reflecting the long-term planning of very expensive investments. In this application, the production and distribution costs are *start-time dependent*, that is, before we can produce at a node (or send flow on an arc), we must first install technology at that node (or on that arc), and the technology installed will determine the per-unit cost of production or distribution *over the entire planning horizon*. The period in which technology is installed also determines the fixed cost for the installation, if any. Therefore, the planning problem becomes a question of if and when to install technology at each supply node and distribution arc, and, given these decisions, how much to produce at each node and distribute along each arc to minimize the total cost of meeting demand over the entire horizon.

A natural formulation of the problem introduces binary variables to model the decision of which period technology will be installed (which we refer to as the start period), and leads to a mixed-integer program with bilinear objective. A simple linearization of this model yields a mixed-integer linear programming (MIP) formulation, but this formulation has poor lower bounds, and hence is not computationally useful. Another approach, due to Adams and Sherali (1993), that has been used in solving general mixed-integer bilinear programs is to create a linearization of the formulation by defining new variables that represent all product terms in the objective. This approach can yield a linearization that gives good lower bounds, but the resulting formulation is very large relative to the original. We follow a similar approach in attempting to generate strong linear formulations for our problem, but we exploit problem-specific structure to generate strong formulations that are relatively small. In particular, we focus on the substructure arising from a single-activity problem, and develop strong formulations for it.

The strategic planning problem we study is similar to the dynamic facility location problem, where installing technology corresponds to opening a facility. One difference is that, in the problem we study, technology must be installed on arcs before they can be used for distribution (in addition to at nodes before they can be used for production). Also, the dynamic facility location literature (e.g., Shulman 1991) generally does not address the dependence of variable operating costs on the period in which the facility is opened. An exception is the model of Van Roy and Erlenkotter (1982), which is basically equivalent to the extended formulation we present in §3.1. However, this model is not analyzed, and no method is proposed for its solution. Another related area of work is the capacity expansion literature (e.g., Hsu 2002, Li and Tirupati 1994, Rajagopalan 1998, Rajagopalan et al. 1998), in which operating costs may depend on the period in which technology is installed, but capacity is installed in continuous increments, as opposed to the discrete install versus do not install decision in our model. Rajagopalan and Soteriou (1994) study a model in which the capacity expansion decision is discrete, but the variable operating costs are ignored altogether.

The main contribution of this paper is a new strategic planning model in which variable costs depend on the period in which technology is installed and the study of formulations for this problem that are both strong and compact. We present two compact MIP formulations and a concave minimization formulation, along with classes of valid inequalities that make each of these formulations strong. This is a nontrivial task because the obvious compact formulation is a mixed-integer *bilinear* (MIBL) formulation, and the obvious MIP formulations are either very weak or very large. The first MIP formulation is obtained by strengthening the initially weak linearization of the MIBL formulation. The second and more compact MIP formulation is obtained by directly studying the MIBL formulation. The concave minimization formulation is the most compact, and we present a specialized branch-and-cut algorithm for solving it in which lower bounds are obtained by using a linear lower bound on the nonconvex objective. These lower bounds are strengthened using valid inequalities, and branching is used to enforce the true objective cost.

We present the problem definition in §2 and our study of MIP formulations in §3. We begin with an extended formulation in §3.1, which is theoretically interesting because it is integral in the single-activity case. In §3.2, we present the first MIP formulation which is based on linearization of the MIBL formulation. In §3.3, we derive an even more compact MIP formulation in which no linearizing variables are introduced. In §4, we present the concave minimization formulation and the branch-and-cut algorithm for solving it. We present computational results in §5 comparing the different formulations and testing the effectiveness of the valid inequalities developed for them. Section 6 contains some concluding remarks.

2. Problem Definition

We consider a multiple-period planning model with time horizon T periods, and let $\mathcal{T} = \{1, \ldots, T\}$. We let I denote a set of production nodes and J a set of demand nodes. The physical decision variables in our model are x_{it}, representing the quantity that is produced at production node $i \in I$ in period $t \in \mathcal{T}$, and x_{ijt}, representing the distribution quantity from production node i to demand node $j \in J$ in period t. For each demand node j and time period t, we are given a demand D_{jt}, which must be met. A model in which unmet demand is allowed but penalized can be achieved by introducing an additional supply node with sufficient capacity and distribution costs to each demand node equal to the penalty for unmet demand.

The primary complication in our model, and the motivation for the present work, is the requirement to install technology at nodes and arcs in this network before production and distribution can be done. This leads to a nonconvex objective that links the periods together. We refer to the period in which the technology is installed as the *start period*. If the start period at a node i is period $t \in \mathcal{T}$, then a fixed charge of $f_{it} \geqslant 0$ is incurred, and the variable cost of production *over the entire horizon* is $c_{it} \geqslant 0$. If production is never performed, then we take the start period, by definition, to be period $T+1$, and appropriately define $f_{i,T+1} = c_{i,T+1} = 0$. Similarly, if technology is installed on the arc from i to j in period t, then a fixed cost of $f_{ijt} \geqslant 0$ is incurred and $c_{ijt} \geqslant 0$ is the variable cost for shipping product from i to j over the entire horizon. Consider now either a fixed production node i, or a fixed arc from i to j, which we will call an *activity*. Dropping for now the subscripts except t, this activity has physical decision variables x_t representing the quantity of the activity in period t. To model the cost of performing this activity, we introduce a vector $y \in \{0, 1\}^T$ of binary decision variables where we have $y_t = 1$ if and only if period t is the start period of the activity. Then, the cost of performing this activity is given by

$$\sum_{s=1}^{T}\left(f_s + c_s \sum_{t=s}^{T} x_t\right) y_s, \tag{1}$$

where the first term captures the fixed costs and the second (bilinear) term captures the variable cost. Indeed, if period s is the start period, then $y_s = 1$ and the objective correctly records the cost $f_s + c_s \sum_{t=s}^{T} x_t$. We must add the constraints

$$x_t - \sum_{s=1}^{t} M y_s \leqslant 0 \quad \forall\, t \in \mathcal{T}, \tag{2}$$

$$\sum_{s=1}^{T} y_s \leqslant 1, \tag{3}$$

to ensure that no activity is done before the start period and that there is only one start period. Here, $M \geqslant 0$ is a given upper bound on the amount of activity that can be performed in a period. Define the set $F(M) = \{(x, y) \in [0, M]^T \times \{0, 1\}^T\colon (2), (3)\}$. Then, reintroducing the subscripts for the production and distribution activities, and letting M_i represent the per-period production capacity at node i and M_{ij} represent the per-period distribution capacity from i to j, we state our strategic planning model as

$$
\begin{aligned}
\text{(SP)}\quad \min\ & \sum_{i\in I}\sum_{s\in\mathcal{T}}\left(f_{is} + c_{is}\sum_{t=s}^{T} x_{it}\right) y_{is} \\
& + \sum_{i\in I}\sum_{j\in J}\sum_{s\in\mathcal{T}}\left(f_{ijs} + c_{ijs}\sum_{t=s}^{T} x_{ijt}\right) y_{ijs} \\
\text{s.t.}\ & \sum_{j\in J} x_{ijt} - x_{it} = 0 \quad \forall\, i \in I,\ t\in\mathcal{T}, && (4)\\
& \sum_{i\in I} x_{ijt} = D_{jt} \quad \forall\, j\in J,\ t\in\mathcal{T}, && (5)\\
& (x_i, y_i) \in F(M_i) \quad \forall\, i\in I,\\
& (x_{ij}, y_{ij}) \in F(M_{ij}) \quad \forall\, i\in I,\ j\in J.
\end{aligned}
$$

Our approach to developing formulations for this problem is to study the substructure corresponding to a single fixed activity. Therefore, for the development of our theoretical results, we will suppress the activity-specific subscripts. Thus, the substructure of interest that we study is simply

$$
\text{(MIBL)}\quad \min_{x,\,y}\left\{\sum_{s=1}^{T} f_s y_s + \sum_{s=1}^{T} c_s y_s \sum_{t=s}^{T} x_t\colon (2),\ (3),\ x\in[0, M]^T,\ y\in\{0,1\}^T\right\}.
$$

This problem itself is trivial because it is optimal to perform no activity at zero cost. However, by developing strong formulations for MIBL, and for each activity including this in the formulation for SP, we achieve a strong formulation for SP in which the constraints (4) and (5) require activity to be performed to meet demand.

Problem SP is $\mathcal{NP}$-hard, even in the special case in which $T = 2$ and there are no fixed costs (Luedtke 2007). Thus, the start-time dependent variable cost nature of this problem adds additional complexity beyond just the presence of fixed costs for installing network components.

3. Mixed-Integer Linear Formulations

In this section, we present strong MIP formulations for MIBL. In doing so, we pay close attention to the size of the formulations developed. In particular, we are interested in solving (at least approximately) problems with a large number of activities and a long planning horizon, so that formulations that involve a large number of auxiliary variables may be undesirable.

3.1. Extended Formulation

We begin with an extended formulation, which introduces $O(T^2)$ auxiliary variables, directly contradicting our stated goal of keeping the formulations small. However, this formulation has the nice property that it is integral for the single-activity problem, and this property is useful for proving tightness of the more compact formulations we present in the sequel.

We introduce auxiliary variables w_{st} for $1 \leqslant s \leqslant t \leqslant T$ to represent the amount of the activity that is done in period t, given that the start period was $s \leqslant t$. MIBL is then reformulated as

$$
\begin{aligned}
\min\ & \sum_{s=1}^{T} c_s \sum_{t=s}^{T} w_{st} + \sum_{s=1}^{T} f_s y_s \\
\text{s.t.}\ & \sum_{s=1}^{t} w_{st} - x_t \geqslant 0 \quad \forall\, t\in\mathcal{T}, && (6)\\
& w_{st} - M y_s \leqslant 0 \quad \forall\, 1\leqslant s\leqslant t\leqslant T, && (7)\\
& \sum_{s=1}^{T} y_s \leqslant 1, && (8)\\
& w\geqslant 0,\quad x\geqslant 0,\quad y\in\{0,1\}^T.
\end{aligned}
$$

We refer to this formulation as EF. The variables x_t and inequalities (6) could actually be eliminated from the formulation, but including them will be useful for the study of the more compact formulations we will present in the sequel.

We now study the tightness of this formulation. Let

$$
F^E = \left\{(x, y, w) \in [0, M]^T \times \{0, 1\}^T \times \mathbb{R}_+^{T(T+1)/2}\colon (6)\text{–}(8)\right\}
$$

represent the feasible region of the extended formulation, and let P^E represent the polytope obtained by dropping the integrality restriction on the binary variables in F^E. For a set F, let $\operatorname{conv}(F)$ represent the convex hull of F. Analogous to the similar result for the extended formulation of the standard lot-sizing problem (Bárány et al. 1986, Nemhauser and Wolsey 1988), we have:

Theorem 1. $P^E = \operatorname{conv}(F^E)$.

Proof. We prove the equivalent result that the polytope P^E has y integer in all extreme points. Let P_1^E be the polytope given by P^E with $M = 1$. We will show that P_1^E is an integral polytope. The result then follows because $(Mx, y, Mw) \in P^E$ if and only if $(x, y, w) \in P_1^E$, so that if y is integral in all extreme points of P_1^E, then it is integral in all extreme points of P^E. We claim that the system defining P_1^E is *totally dual integral* (e.g., Nemhauser and Wolsey 1988, p. 537). Indeed, consider the linear programming relaxation of EF having $M = 1$ and arbitrary objective given by

$$
\max \sum_{t=1}^{T} \tilde{c}_t x_t + \sum_{s=1}^{T}\sum_{t=s}^{T} \tilde{d}_{st} w_{st} + \sum_{t=1}^{T} \tilde{f}_t y_t,
$$

where the coefficients $(\tilde{c}, \tilde{f}, \tilde{d}) \in \mathbb{Z}^{2T+T(T+1)/2}$ are integer. Associating dual variables σ_{st} with inequalities (7), γ_t with inequalities (6), and π_0 with (8), the dual of this linear program is

$$\begin{aligned}
\min\ & \pi_0 \\
\text{s.t.}\ & \pi_0 - \sum_{t=s}^{T} \sigma_{st} \geqslant \tilde{f}_s \quad \forall s \in \mathcal{T}, \\
& \sigma_{st} - \gamma_t \geqslant \tilde{d}_{st} \quad \forall 1 \leqslant s \leqslant t \leqslant T, \\
& \gamma_t \geqslant \tilde{c}_t \quad \forall t \in \mathcal{T}, \\
& \sigma \geqslant 0, \quad \gamma \geqslant 0, \quad \pi_0 \geqslant 0.
\end{aligned}$$

It is simple to see that the optimal dual solution is

$$\begin{aligned}
&\pi_0 = \left(\max_{s\in\mathcal{T}} \left\{\tilde{f}_s + \sum_{t=s}^{T} \sigma_{st}\right\}\right)^+, \\
&\sigma_{st} = (\tilde{d}_{st} + \gamma_t)^+ \quad \forall 1 \leqslant s \leqslant t \leqslant T, \\
&\gamma_t = (\tilde{c}_t)^+ \quad \forall t \in \mathcal{T},
\end{aligned}$$

where we use the notation $(\cdot)^+ = \max\{\cdot, 0\}$. Thus, the dual solution is integral, and so the system defining the primal is totally dual integral, and hence P_1^E is an integral polytope. □

We remark that if M is integral, the arguments in the above proof can be used to establish that the extreme points of P^E have x and w integer as well as y.

3.2. Linearizing the Bilinear Formulation

A natural way to deal with the bilinear objective term appearing in (1) in a compact way is to introduce linearization variables z_s to capture the bilinear terms $y_s \sum_{t=s}^{T} x_s$ for each $s \in \mathcal{T}$. That is, z_s represents the amount of activity that is *charged* at the variable cost of period s, c_s. Then, we obtain the mixed-integer linear formulation

$$\begin{aligned}
\min\ & \sum_{s\in\mathcal{T}} f_s y_s + \sum_{s\in\mathcal{T}} c_s z_s \\
\text{s.t.}\ & z_s - \sum_{t=s}^{T} x_t + (1-y_s)(T-s+1)M \geqslant 0 \quad \forall s \in \mathcal{T}, \qquad (9) \\
& z \geqslant 0, \quad x \geqslant 0, \quad y \in \{0,1\}^T, \quad (2) \text{ and } (3),
\end{aligned}$$

which we refer to as the *weak linearization* (WL) formulation. If period s is the start period, then $y_s = 1$ and constraint (9) ensures that $z_s \geqslant \sum_{t=s}^{T} x_t$ so the variable cost c_s is charged on the activity over the entire horizon. On the other hand, if the activity does not begin in period s, $y_s = 0$ so that (9) is not binding, and because we are minimizing we will have $z_s = 0$, so that none of the activity is charged at the variable cost of period s, as desired. Although this yields a correct formulation, we will see in the computational results in §5 that the bounds from the linear programming relaxation of this formulation are extremely weak.

3.2.1. Strengthening Using Ideas from Lot Sizing. The main problem with the WL formulation is the presence of the weak constraints (9). Indeed, when the binary variables are fractional, it is possible to have positive activity levels, and yet have $z_s = 0$ for all s, so that we pay nothing for the activity we perform. Fortunately, it turns out that constraints (9) can be eliminated by using an idea from lot sizing. Note that we can interpret z_s as an *economic* amount that is paid for in period s, which can be used by the *physical* activities x_t in any period $t \geqslant s$. With this interpretation, our formulation appears similar to the lot-sizing problem, in which we have to produce to meet demands over time (Bárány et al. 1984, Nemhauser and Wolsey 1988). Using this analogy, we can add the constraints

$$\sum_{s=1}^{t} z_s \geqslant \sum_{s=1}^{t} x_s \quad \forall t \in \mathcal{T}, \qquad (10)$$

which state that the cumulative amount we pay for up to each time period t must be at least as much as the physical activity levels up to period t. The difference between our problem and the standard lot-sizing problem is that all activities must be charged the variable cost corresponding to the start period, or equivalently, our economic production variables, z_s, can only be positive in the start period. Thus, we obtain the lot-sizing inspired formulation, LS, given by

$$\begin{aligned}
\min\ & \sum_{s\in\mathcal{T}} f_s y_s + \sum_{s\in\mathcal{T}} c_s z_s \\
\text{s.t.}\ & z_s - (T-s+1)My_s \leqslant 0 \quad \forall s \in \mathcal{T}, \qquad (11) \\
& x \geqslant 0, \quad z \geqslant 0, \quad y \in \{0,1\}^T, \quad (2), (3), \text{ and } (10),
\end{aligned}$$

where (11) guarantees that z_s can only be positive in the start period. We include (2), which could be eliminated and replaced with bounds $x_t \leqslant M$, because their presence tightens the formulation, and computational results indicate that this benefit far outweighs the increased formulation size.

3.2.2. The Convex Hull. In this section, we characterize the convex hull of feasible solutions to the LS formulation. Thus, we study the set

$$F^{\mathrm{LS}} = \left\{(x, y, z) \in [0, M]^T \times \{0,1\}^T \times \mathbb{R}_+^T\colon (8), (10), (11)\right\},$$

where here we have included the constraint $x \in [0, M]^T$ but not the inequalities (2), which are not needed to define the feasible region.

Theorem 2. conv(F^{LS}) *is given by the set of* $(x, y, z) \in \mathbb{R}_+^{3T}$, *which satisfy* (8), (11), *and*

$$\sum_{t\in S} x_t \leqslant \sum_{t\in L} z_t + M \sum_{t\in S} \sum_{s\in\{1,\ldots,t\}\setminus L} y_s \quad \forall S, L \subseteq \mathcal{T}. \qquad (12)$$

Proof. Note that (10) and (2) are special cases of (12). Let P^{LS} be the set of (x, y, z) that satisfy the inequalities stated in the theorem. We prove conv$(F^{\mathrm{LS}}) \subseteq P^{\mathrm{LS}}$ by showing that (12) are valid for F^{LS}. Let $(x, y, z) \in F^{\mathrm{LS}}$, and let $i \in \mathcal{T}$

be such that $y_i = 1$. If $y_t = 0$ for all t, set $i = T+1$. If $i \in L$, then $\sum_{t\in L} z_t = z_i \geqslant \sum_{t\in\mathcal{T}} x_t \geqslant \sum_{t\in S} x_t$ and (12) holds. If $i \notin L$, then for $t \in S$ we have $x_t = 0 \leqslant M \sum_{s\in\{1,\dots,t\}\setminus L} y_s$ if $t < i$ and $x_t \leqslant M = M\sum_{s\in\{1,\dots,t\}\setminus L} y_s$ if $t \geqslant i$. Hence,

$$\sum_{t\in S} x_t \leqslant M \sum_{t\in S} \sum_{s\in\{1,\dots,t\}\setminus L} y_s$$

and (12) holds.

Now, suppose that $(x, y, z) \in P^{\mathrm{LS}}$. We show that there exists w such that

$$\sum_{t=s}^{T} w_{st} \leqslant z_s \quad \forall\, s \in \mathcal{T} \tag{13}$$

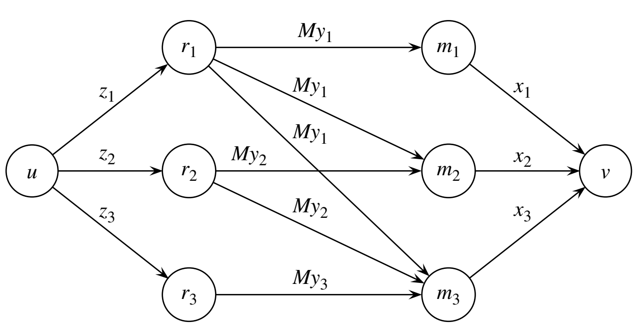
and $(x, y, w) \in \operatorname{conv}(F^E)$, as defined in §3.1. By Theorem 1, $P^E = \operatorname{conv}(F^E)$, and so the latter condition can be verified by checking that (x, y, w) satisfy the inequalities defining the polyhedron P^E. Consider a network G with node set given by $V = \{u, v, r_s\ \forall\, s\in\mathcal{T}, m_t\ \forall\, t\in\mathcal{T}\}$. The arcs in this network consist of arcs from u to r_s with capacity z_s for $s \in \mathcal{T}$, arcs from r_s to m_t with capacity My_s for all $1 \leqslant s \leqslant t \leqslant T$, and arcs from m_t to v with capacity x_t for all $t \in \mathcal{T}$. An example of this graph for $T = 3$ is given in Figure 1, where the label on each arc represents its capacity. Because $(x, y, z) \in P^{\mathrm{LS}}$, we have $x \in \mathbb{R}^T_+$, $y \in \mathbb{R}^T_+$ and y satisfies (8). Thus, if we associate w_{st} with the flow from r_s to m_t in this network, it is easy to check that if this network has a $u-v$ flow of $\sum_{t\in T} x_t$, then the desired w exists. It follows from the max-flow min-cut theorem that if the capacity of every $u-v$ cut in G is at least $\sum_{t\in\mathcal{T}} x_t$, then there exists a $u-v$ flow of this value. Let $U \subset V$ with $u \in U$ and $v \notin U$ and consider the cut defined by U and $V\setminus U$. Let $S = \{t \in \mathcal{T}\colon m_t \notin U\}$ and $L = \{s \in \mathcal{T}\colon r_s \notin U\}$. The capacity of this cut is

$$\sum_{s\in L} z_s + \sum_{t\in S} \sum_{s\in\{1,\dots,t\}\setminus L} My_s - \sum_{t\in S} x_t + \sum_{t\in\mathcal{T}} x_t.$$

Because (x, y, z) satisfies (12) for all $S, L \subseteq \mathcal{T}$, it follows that the capacity of this cut is at least $\sum_{t\in\mathcal{T}} x_t$. Thus, there exists a $u-v$ flow of $\sum_{t\in\mathcal{T}} x_t$.

We complete the proof by demonstrating that the existence of w such that $(x, y, w) \in \operatorname{conv}(F^E)$ and (13) is satisfied implies $(x, y, z) \in \operatorname{conv}(F^{\mathrm{LS}})$. Therefore, suppose such a w exists. We first observe that there exists a w' such that $(x, y, w') \in \operatorname{conv}(F^E)$ and $\sum_{t=s}^{T} w'_{st} = z_s\ \forall\, s \in \mathcal{T}$. This follows because $z_s \leqslant (T-s+1)My_s$, so that where necessary w_{st} can be increased to obtain equality in (13) without violating the inequalities (7), i.e., $w_{st} \leqslant My_s$. Next, let (x^i, y^i, w^i), $i \in I$ be a set of points of F^E and $\lambda \in \mathbb{R}^{|I|}_+$ such that $\sum_{i\in I} \lambda_i = 1$ and $(x, y, w') = \sum_{i\in I} \lambda_i (x^i, y^i, w^i)$. For each $i \in I$, define z^i by $z^i_s = \sum_{t=s}^{T} w^i_{st}\ \ \forall\, s \in \mathcal{T}$. Then, it is easy to check that $(x, y, z) = \sum_{i\in I} \lambda_i (x^i, y^i, z^i)$ and that $(x^i, y^i, z^i) \in F^{\mathrm{LS}}$ for $i \in I$, thus establishing that $(x, y, z) \in \operatorname{conv}(F^{\mathrm{LS}})$. □

Figure 1. Example of network G with $T = 3$.



Although the inequalities (12) have a structure similar to the (l, S) inequalities (Bárány et al. 1984) for the classical lot-sizing problem, they cannot be derived from the (l, S) inequalities. Indeed, inequalities (12) crucially use the constraint $\sum_{t\in\mathcal{T}} y_t \leqslant 1$, which is not present on the binary variables in the classical lot-sizing problem.

The proof of Theorem 2 demonstrates how separation of (12) can be accomplished by finding the minimum cut in a network with $O(T)$ nodes. This immediately implies that separation can be accomplished with $O(T^3)$ complexity by finding the maximum flow in this network. Using dynamic programming, it is possible to perform separation in $O(T^2)$ by exploiting the special structure of the corresponding network; see Luedtke (2007) for details.

3.3. Formulation Based on the Bilinear Model

We now present a linear formulation that does not introduce the auxiliary variables z_t, $t \in \mathcal{T}$. To obtain a linear objective in this case, we introduce an upper-bound variable, μ, on the objective, and move the bilinear objective into the constraints. That is, we simply reformulate MIBL as

$$\begin{aligned} \min\ \ & \mu + \sum_{s=1}^{T} f_s y_s \\ \text{s.t.}\ \ & \mu - \sum_{s=1}^{T} y_s c_s \sum_{t=s}^{T} x_t \geqslant 0, \qquad (14) \\ & x \geqslant 0, \quad y \in \{0,1\}^T, \quad (2) \text{ and } (3). \end{aligned}$$

Now note that for any fixed y, the constraints reduce to a set of linear constraints. Furthermore, a feasible y can be equal to one of only $T+1$ binary vectors, so that the feasible region is the union of exactly $T+1$ polyhedra. Therefore, optimizing a general linear function over this feasible region is easy, and consequently separating over the convex hull of this feasible region is also theoretically easy. In fact, disjunctive programming theory can be used to write an explicit, polynomial-sized linear program to separate over this convex hull; see Balas (1998). However, we prefer to have an explicit characterization of inequalities defining the convex hull, and an efficient combinatorial algorithm for separation over these inequalities. We therefore study the convex hull of the set

$$F^{\mathrm{BL}} = \{(\mu, x, y) \in \mathbb{R} \times [0, M]^T \times \{0,1\}^T\colon (2), (3), (14)\}.$$

THEOREM 3. *Let $i_t \in \{1, \ldots, t\}$ for each $t \in \mathcal{T}$. Then, the inequality*

$$\mu \geqslant \sum_{t=1}^{T} c_{i_t} x_t - \sum_{s=1}^{T}\sum_{t=s}^{T} M(c_{i_t} - c_s)^+ y_s \tag{15}$$

is valid for F^{BL}.

PROOF. Let $(\mu, x, y) \in F^{\mathrm{BL}}$ and let $i_t \in \{1, \ldots, t\}$ $\forall\, t \in \mathcal{T}$. If $x = y = \mathbf{0}$, the inequality is trivially valid. Otherwise, let k be the period such that $y_k = 1$ and $y_t = 0$ for all $t \neq k$. Then, $x_t = 0$ for $t = 1, \ldots, k-1$ and

$$\mu \geqslant c_k \sum_{t=k}^{T} x_t \geqslant c_k \sum_{t=k}^{T} x_t - \sum_{t=k}^{T} (c_{i_t} - c_k)^+ (M - x_t) \tag{16}$$

$$\geqslant c_k \sum_{t=k}^{T} x_t + \sum_{t=k}^{T} (c_{i_t} - c_k) x_t - \sum_{t=k}^{T} M(c_{i_t} - c_k)^+ \tag{17}$$

$$= \sum_{t=1}^{T} c_{i_t} x_t - \sum_{s=1}^{T}\sum_{t=s}^{T} M(c_{i_t} - c_s)^+ y_s,$$

where the first inequality in (16) follows from (14), the second inequality in (16) follows because $x_t \leqslant M$ for all $t \in \mathcal{T}$, and (17) follows because $x \geqslant 0$. This proves that (15) is valid for F^{BL}. □

In fact, together with the inequalities defining F^{BL}, the inequalities (15) are sufficient to define the convex hull of F^{BL}. The proof can be found in Luedtke (2007).

THEOREM 4. $\mathrm{conv}(F^{\mathrm{BL}})$ *is given by the set of* $(\mu, x, y) \in \mathbb{R}_+^{2T+1}$ *that satisfy* (2), (3), *and* (15) *for all* $i_t \in \{1, \ldots, t\}\, \forall\, t \in \mathcal{T}$.

Despite this characterization of $\mathrm{conv}(F^{\mathrm{BL}})$, it is still not obvious how to obtain a *compact* valid MIP formulation when the nonlinear constraints (14) are dropped. A simple option is to use

$$\mu \geqslant c_s \sum_{t=s}^{T} x_t - (T - s + 1) M c_s (1 - y_s) \quad \forall\, s \in \mathcal{T}, \tag{18}$$

and then add inequalities (15) as needed to strengthen the formulation. This yields a valid formulation because if $y_s = 1$, then the right-hand side of (18) yields the correct cost lower bound for μ, whereas if $y_s = 0$, the right-hand side of (18) will not be positive, and hence will not constrain μ. However, we can avoid adding (18), which are likely to be weak, by observing that a small subset of the inequalities (15) are sufficient to guarantee a valid mixed-integer linear formulation.

THEOREM 5. F^{BL} *is given by the set of* $(\mu, x, y) \in \mathbb{R} \times [0, M]^T \times \{0, 1\}^T$ *which satisfy* (2), (3) *and*

$$\mu \geqslant \sum_{s=1}^{k-1} c_s x_s + \sum_{s=k}^{T} c_k x_s - \sum_{t=1}^{k-1}\sum_{s=1}^{t} M(c_t - c_s)^+ y_s - \sum_{t=k}^{T}\sum_{s=1}^{t} M(c_k - c_s)^+ y_s \quad \forall\, k \in \mathcal{T}. \tag{19}$$

The proof follows by first observing that each inequality in (19) corresponds to (15) with a particular choice of i_t for each t, so that (19) are valid by Theorem 3, and then noting that when $y_k = 1$, the kth inequality in (19) enforces $\mu \geqslant c_k \sum_{s=k}^{T} x_s$ so that μ records the correct cost when y is integer feasible. A formal proof can be found in Luedtke (2007).

As a consequence of this theorem, we obtain a new valid mixed-integer linear formulation for this problem that we refer to as the LBL formulation, and a class of valid inequalities given in Theorem 4 that can be added to make the formulation as tight as possible for a single activity.

We now discuss separation of the inequalities (15). Given a point (μ^*, x^*, y^*), testing whether there is an inequality of the form (15) that this point violates amounts to testing whether $\mathrm{RHS}^* > \mu^*$, where

$$\mathrm{RHS}^* = \sum_{t=1}^{T} \max_{i_t \in \{1, \ldots, t\}} \left\{ c_{i_t} x_t^* - \sum_{s=1}^{t} M(c_{i_t} - c_s)^+ y_s^* \right\}.$$

Then, if we define $v(i, t) = \sum_{s=1}^{t} (c_i - c_s)^+ y_s^*$ for $1 \leqslant i \leqslant t \leqslant T$, it is easy to see that these quantities can be calculated in $O(T^2)$ time. Rewriting RHS^* as

$$\mathrm{RHS}^* = \sum_{t=1}^{T} \max_{i=1, \ldots, t} \{ c_i x_t^* - v(i, t) \},$$

we see that we can subsequently calculate RHS^* in $O(T^2)$ time, leading to separation in $O(T^2)$.

4. Concave Formulation

In this section, we demonstrate how problem SP can be formulated as a very compact concave minimization problem, and present a specialized branch-and-cut algorithm to solve this formulation. This formulation is based on rewriting the single-activity problem MIBL as

(CM) $\quad \min\{\mathrm{h(x)}\colon \mathrm{x} \in [0, \mathrm{M}]^{\mathrm{T}}\}$,

where $h\colon [0, M]^T \to \mathbb{R}$ is given by

$$h(x) = \min_y \ \sum_{s=1}^{T} f_s y_s + \sum_{s=1}^{T} y_s c_s \sum_{t=s}^{T} x_t$$

$$\text{s.t.} \ \sum_{s=1}^{t} M y_s \geqslant x_t \quad \forall\, t \in \mathcal{T},$$

$$\sum_{t \in \mathcal{T}} y_t \leqslant 1, \quad y \in \{0, 1\}^T.$$

For any vector $x \in \mathbb{R}_+^T$, $x \neq \mathbf{0}$, we let $t_{\min}(x) = \min\{t \in \mathcal{T}\colon x_t > 0\}$. We also define $t_{\min}(\mathbf{0}) = T + 1$. Then, we can write $h(x)$ as

$$h(x) = \min \left\{ f_p + c_p \sum_{t=p}^{T} x_t \colon p = 1, \ldots, t_{\min}(x) \right\}. \tag{20}$$

That is, for each $x \in [0, M]^T$, $h(x)$ is the minimum cost to perform the activity, subject to the feasibility requirement that the start period, p, of the activity must occur on or before the first period in which the activity level is positive.

We next observe that the function $h(x)$ is concave, although not separable concave. This is almost immediate from (20) because this shows that $h(x)$ is the minimum of finitely many affine functions of x. However, the set of affine functions depends on x through the function $t_{\min}(x)$, and so the concavity of h requires justification.

THEOREM 6. *h is concave over $[0, M]^T$.*

PROOF. Let x^k, $k \in K$ be a set of points in $[0, M]^T$ and $\lambda \in \mathbb{R}_+^{|K|}$ be such that $\sum_{k\in K} \lambda_k = 1$. Also, let $x = \sum_{k\in K} \lambda_k x^k$. Then,

$$\begin{aligned} h(x) &= \min\left\{ f_p + c_p \sum_{t=p}^{T} x_t \colon p = 1, \ldots, t_{\min}(x) \right\} \\ &= \min\left\{ \sum_{k\in K} \lambda_k \left(f_p + c_p \sum_{t=p}^{T} x_t^k \right) \colon p = 1, \ldots, t_{\min}(x) \right\} \\ &\geqslant \sum_{k\in K} \lambda_k \min\left\{ f_p + c_p \sum_{t=p}^{T} x_t^k \colon p = 1, \ldots, t_{\min}(x^k) \right\} \\ &= \sum_{k\in K} \lambda_k h(x^k), \end{aligned} \tag{21}$$

where the inequality (21) follows because $t_{\min}(x) \leqslant t_{\min}(x^k)$ for all k such that $\lambda_k > 0$. □

We therefore have a concave minimization formulation of problem SP with simple linear constraints (4) and (5). General purpose methods have been developed to solve such types of $\mathcal{NP}$-hard problems; see, e.g., Horst and Tuy (1990). However, our problem has a special structure that we exploit to enable us to solve large-scale instances.

We propose a specialized branch-and-cut method to solve formulation CM. First, we reformulate CM by introducing an objective upper-bound variable μ to obtain the formulation

$$\min\{\mu\colon \mu \geqslant h(x),\ x \in [0, M]^T\}.$$

In our method, we relax the nonlinear inequality $\mu \geqslant h(x)$ and subsequently enforce it by branching. In addition, we generate valid inequalities to approximate the nonconvex set

$$E = \{(\mu, x) \in \mathbb{R} \times [0, M]^T \colon \mu \geqslant h(x)\} \tag{22}$$

to obtain tight lower bounds at nodes in our branch-and-bound tree.

In describing this method, we continue to focus primarily on the single-activity problem CM, but it should be understood that in the context of the overall problem SP, the branching may have to be done on each of the activities in the formulation, and consequently, at each node in the tree a choice may have to be made as to which activity on which to branch. Furthermore, we present node relaxations for a single activity, but it should be understood that when we refer to solving a node relaxation, we are solving the relaxation of the overall problem, using the relaxations from all the activities together.

4.1. Branching and Lower Bounds

In the previous formulations in which we introduced binary variables, branching on these variables meant we were branching on the decision of which period would be the start period. In this approach, we still branch on this decision, implicit in the definition of h, but we do so without introducing the binary variables. Specifically, we branch on the implicit variable, p, representing the start period of the activity. At every node n in the branch-and-bound tree, we will have that $p \in \{l(n), \ldots, u(n)\}$, where $1 \leqslant l(n) \leqslant u(n) \leqslant T+1$. Recall that if $T+1$ is the start period, this means the activity never starts. At the root node, node 0, we set $l(0) = 1$ and $u(0) = T+1$.

Lower Bounds. We are interested in deriving lower bounds on the cost function $h(x)$ subject to the restriction that the start period p satisfies $l \leqslant p \leqslant u$. Therefore, define the cost function given this restriction by

$$h(x; l, u) = \min\left\{ f_p + c_p \sum_{t=p}^{T} x_t \colon l \leqslant p \leqslant \min(u, t_{\min}(x)) \right\}$$

for x such that $x_t = 0$ for $t = 1, \ldots, l-1$. Also, define

$$\underline{f}(l, u) = \min\{f_p\colon l \leqslant p \leqslant u\},$$
$$\underline{c}_t(l, u) = \min\{c_p\colon l \leqslant p \leqslant \min(t, u)\} \quad \text{for } t = l, \ldots, T+1.$$

Then, the following lower bound is valid (see Luedtke 2007 for a proof).

THEOREM 7.

$$h(x; l, u) \geqslant \underline{f}(l, u) + \sum_{t=l}^{T} \underline{c}_t(l, u) x_t. \tag{23}$$

At the root node we have no restrictions on the start time so that $l = 1$ and $u = T+1$ and we obtain the lower bound $h(x) \geqslant \sum_{t=1}^{T} \underline{c}(1, t) x_t$, where the fixed cost term vanished because, by definition, $f_{T+1} = 0$.

At a node in the branch-and-bound tree, if we have $l \leqslant p \leqslant u$, we obtain a lower bound by replacing $\mu \geqslant h(x)$ with

$$\mu \geqslant \underline{f}(l, u) + \sum_{t=l}^{T} \underline{c}_t(l, u) x_t. \tag{24}$$

An important property of this lower bound is that it is exact when $l = u$. That is,

$$h(x; l, l) = f_l + c_l \sum_{t=l}^{T} x_t = \underline{f}(l, l) + \sum_{t=l}^{T} \underline{c}_t(l, l) x_t$$

so that if (24) is enforced when $l = u$, we necessarily have $\mu \geqslant h(x)$.

Branching. At any node n in the tree, we first solve the relaxation for that node obtained by including (24) in the linear program. Specifically, the relaxation solved has linear constraints (4) and (5), simple bounds $0 \leqslant x_{it} \leqslant M_i$ and $0 \leqslant x_{ijt} \leqslant M_{ij}$, and for each activity (node i or arc from i to j) an objective upper bound variable μ is introduced and the corresponding inequality (24) is added. After solving the relaxation, we obtain a solution x for each activity. If the optimal objective of the relaxation exceeds the cost of the best incumbent solution, we can fathom this node without branching further. Otherwise, we check whether the inequality $\mu \geqslant h(x)$ is violated (for any activity). If not, we have a new incumbent solution, and we need not explore this node any further. If so, then we select an activity with $\mu < h(x)$, for example, by selecting the activity with maximum value of $h(x) - \mu$. Because we have enforced (24), we must have $l(n) < u(n)$ for this activity, so we can select $k \in \{l(n), \ldots, u(n) - 1\}$. We then create two nodes, d_1 and d_2, by enforcing in d_1 that $p \leqslant k$, and enforcing in d_2 that $p > k$. This is achieved by setting $l(d_1) = l(n)$ and $u(d_1) = k$, and $l(d_2) = k + 1$ and $u(d_2) = u(n)$. Thus, the lower bound (24) will be updated in the two nodes, and in addition, we will enforce that $x_t = 0$ for $t = 1, \ldots, k$ in node d_2, reflecting the restriction that the activity will not start until period $k + 1$.

This branching scheme ensures that each path in the resulting branch-and-bound tree will finitely terminate with a leaf node in which $l(n) = u(n)$ (if not earlier) and will therefore not have to be explored further. Thus, the branching scheme is finite for each single activity, and hence will be finite for finitely many activities.

4.2. Improving the Lower Bounds

In the LS and LBL formulations, we were able to explicitly characterize the convex hull of the set of feasible solutions. We have not been able to do that for the feasible set in this formulation, given by (22). Fortunately, we can still separate all inequalities valid for E by solving a linear program. The proof, given in the appendix, follows a direct derivation of this linear program, but we note that it could also be obtained by disjunctive programming theory (Balas 1998).

Theorem 8. *Let* $(\mu, x) \in \mathbb{R} \times [0, M]^T$. *Then,* $(\mu, x) \in \operatorname{conv}(E)$ *if and only if* $\mu \geqslant v^*$, *where*

$$v^* = \max \ \beta + \sum_{t=1}^{T} \alpha_t x_t$$

$$\text{s.t.} \ \ M \sum_{t=p}^{T} \sigma_{pt} + \beta \leqslant f_p \quad \forall\, p \in \mathcal{T}, \tag{25}$$

$$\alpha_t - \sigma_{pt} \leqslant c_p \quad \forall\, 1 \leqslant p \leqslant t \leqslant T,$$

$$\beta \leqslant 0, \quad \sigma_{pt} \geqslant 0 \quad \forall\, 1 \leqslant p \leqslant t \leqslant T.$$

Moreover, if $\mu < v^*$, *the optimal solution yields an inequality of the form*

$$\mu \geqslant \beta + \sum_{t=1}^{T} \alpha_t x_t, \tag{26}$$

which cuts off (μ, x) *and is valid for* $\operatorname{conv}(E)$.

4.3. Feasible Solutions

Any solution to a node relaxation in our branch-and-bound tree will yield a solution that satisfies the physical constraints of our formulation. Thus, all we need to do to obtain a feasible solution at any node is to calculate the true cost of each activity for the levels given by the relaxation solution.

5. Computational Results

We performed computational tests to compare the different formulations and to investigate the effect of using the valid inequalities. We tested six different formulations: WL, LS, LS.C, LBL, LBL.C, and CM. WL refers to the weak linearization formulation of §3.2. LS refers to the lot-sizing inspired formulation of §3.2.1, and LBL refers to the formulation presented in §3.3. LS.C and LBL.C refer to the LS and LBL formulations, using the valid inequalities (12) and (15), respectively. CM refers to the compact concave minimization formulation of §4, solved with the specialized branch-and-cut algorithm. We did not test the extended formulation which is not practical for large instances.

Table 1 summarizes the number of variables and rows in each of the formulations, where we let $n = |I|$ and $m = |J|$. The number of rows does not include cuts or the $T(n + m)$ constraints (4) and (5), which are present in all formulations. For the extended formulation, we list only the approximate size, to emphasize that it is quadratic in the number of periods. The number of rows in the LS formulations includes the inequalities (2), which are not necessary for the LS formulation, but yield significantly better computational results.

5.1. Test Instances

We randomly generated instances that have characteristics, including the problem size, similar to data in the application that motivated this work. In all cases, the variable costs decrease at a constant rate as the start period is delayed. Fixed costs for installing technology were not considered in the motivating application, but because fixed costs may be present in other applications, we generated instances with and without fixed costs. For instances with fixed costs, the fixed cost does not depend on the start period.

5.2. Implementation Comments

We used CPLEX 9.0 as our mixed-integer programming solver, and implemented the addition of valid inequalities using CPLEX cut callback routines. For the implementation

Table 1. Sizes of the different formulations.

Formulation	Number of variables	Number of rows
Extended	$O(T^2 nm)$	$O(T^2 nm)$
LS	$3T(nm + n)$	$(3T + 1)(nm + n)$
LBL	$2T(nm + n)$	$(2T + 1)(nm + n)$
CM	$(T + 1)(nm + n)$	$nm + n$

of the specialized branch-and-cut algorithm to solve the concave minimization formulation, we used CPLEX to solve the linear programming relaxations, select nodes to explore, and manage the branch-and-cut tree. We implemented our customized branching strategy using the branch callback routine provided by CPLEX.

We let CPLEX generate the cuts it generates by default, and in particular, because of the network structure of our test instances, CPLEX was able to generate many flow cover inequalities for the LS and LBL formulations.

We investigated different strategies for using the valid inequalities we have developed, including generating locally and globally valid inequalities at nodes throughout the search tree, and at varying frequencies. However, we found that the simple strategy of generating globally valid inequalities at the root node was most effective, and therefore this is the strategy we used.

In problem SP, we are deciding if and when to install technology at supply points as well as on distribution arcs in the network. Because the decisions of when to start a supply node affect the decisions of when to start distribution on arcs from that node, it makes sense to put priority on these decisions, and we have done so in our implementations. For the LS and LBL formulations, this is done by giving the corresponding binary variables appropriate priority levels in CPLEX. When solving the CM formulation, the activity selection criterion we use favors selection of a supply activity over a distribution activity. The criterion is simply to branch on the activity that has the maximum disparity between the actual cost of the activity over the horizon $h(x)$ and the lower bound on the cost given by the current value of μ. This tends to favor selection of supply activities because the supply activity levels are generally larger than the distribution activity levels (because one supply output can be split on many distribution arcs).

5.3. Results

We first conducted tests on a set of small instances to compare the solution times required to solve instances to optimality. For these instances, we used a time limit of one hour. We used a set of 30 instances, half with and half without fixed costs. Table 2 lists each of the different sizes of instances in this test set and the geometric average time, taken over five instances at each size, to solve these instances to optimality. For the WL formulation, none of the instances could be solved to optimality within the one-hour time limit, so we report the average remaining optimality gaps and the average gap between the best solution found by WL and the optimal solution (UB Gap). The average remaining optimality gaps with the WL formulation were huge, 57% on average. More significantly from a practical standpoint, the best solution found within the hour time limit was 6.1% more costly than the optimal solution on average. In contrast, Table 2 indicates that these instances could be solved in minutes using the LS and LBL formulations with and without the new valid inequalities. This table also indicates that for the LS formulation on small instances, using the new cuts slightly reduces the average solution time for instances without fixed costs, but does not help the solution times for instances with fixed costs. In addition, the LBL.C formulation tends to solve these sized instances to optimality slightly faster than the alternatives, whereas solving the concave minimization formulation takes significantly longer than the alternatives.

We also tested the formulations, excluding the WL formulation, on a set of 40 large instances, half without and half with fixed costs. In these instances, the number of demand points is fixed at $|J| = 20$. We used a time limit of eight hours for these instances, and none of the formulations were able to provably solve any of these instances to optimality within this time limit. Table 3 lists the instance sizes and the average optimality gap obtained within two and eight hours using the different formulations. Each entry in this table is an average over five instances of the size given in the row. From this table, we observe that the LBL.C formulation yields the smallest optimality gap after two and eight hours in almost all cases. In particular, using the valid inequalities in the LBL formulation significantly reduces the optimality gaps. In contrast to this, using the valid inequalities in the LS formulation does not always reduce the optimality gap. The explanation for this is that in the larger LS formulation, the additional time spent solving the linear programming relaxations when using the valid inequalities outweighs the improved lower bounds obtained by using them. Table 3 indicates that although these formulations could not solve these large instances to optimality within eight hours, the optimality gaps were usually reasonable even after just two hours, with most being not much larger than 1%. The instances with fixed costs and $T = 20$ periods are an exception, with average optimality gaps in the 3%–4% range in the best cases after two hours. However, when run for eight hours, the optimality gaps are reduced to less than 2% when using one of the more compact formulations (LBL, LBL.C, or CM). It is also evident from Table 3 that the instances with fixed costs are more difficult to solve than those without fixed costs.

We next study how the formulations do in terms of generating good feasible solutions. Table 4 presents the average percent by which the best feasible solution found using

Table 2. Results for small instances.

		WL gaps (%)		Average time (s)				
FC?	n, m, T	Opt.	UB	LS	LS.C	LBL	LBL.C	CM
No	10, 5, 10	78.9	1.4	30	28	18	16	54[a]
	15, 5, 10	76.8	1.5	46	45	41	29	498[a]
	10, 10, 10	55.6	0.9	70	51	53	35	506[a]
Yes	10, 5, 10	46.0	10.3	17	21	12	11	228[a]
	15, 5, 10	49.6	11.3	162	196	128	116	1,577[a]
	10, 10, 10	38.8	11.2	47	51	36	33	373[a]

[a]Times were truncated at one hour: reported average is a lower bound.

Table 3. Average optimality gaps for large instances.

		Average percentage optimality gap after (2 hrs., 8 hrs.)				
FC?	n, T	LS	LS.C	LBL	LBL.C	CM
No	50, 10	(0.39, 0.36)	(0.36, 0.29)	(0.47, 0.43)	(0.32, 0.29)	(0.67, 0.62)
	50, 20	(2.42, 1.25)	(2.38, 1.43)	(1.88, 1.52)	(1.25, 0.89)	(1.28, 1.25)
	100, 10	(0.26, 0.22)	(0.19, 0.17)	(0.33, 0.28)	(0.18, 0.16)	(0.54, 0.53)
	100, 20	(1.72, 1.24)	(1.87, 1.27)	(1.92, 1.10)	(1.26, 0.63)	(1.36, 0.95)
Yes	50, 10	(0.62, 0.49)	(0.48, 0.43)	(0.60, 0.54)	(0.56, 0.52)	(0.91, 0.84)
	50, 20	(5.32, 1.63)	(5.26, 3.59)	(1.86, 1.83)	(1.85, 1.77)	(2.96, 2.00)
	100, 10	(0.44, 0.44)	(0.59, 0.38)	(0.52, 0.49)	(0.47, 0.45)	(0.81, 0.73)
	100, 20	(3.39,[a] 3.02)	(3.42,[a] 2.98)	(3.34, 1.46)	(3.31, 1.33)	(3.98, 1.83)

[a]Average is based on four of the five instances.

each formulation after two and eight hours exceeds the overall best lower bound, LB*. Specifically, LB* is the best lower bound over all formulations run for eight hours, and the quantity reported is the average over the five instances at each size of $(\text{UB}-\text{LB}^*)/\text{LB}^*$, where UB is the value of the best feasible solution found by the formulation. These results once again favor the LBL formulation (with and without cuts). However, the CM formulation does nearly as well, and does better for the largest instances (those with $T=20$) without fixed costs. This indicates that the majority of the weakness in the optimality gap provided by the CM formulation can be attributed to a relatively weaker lower bound. The LS formulation generally yields the worst results in terms of solution quality, particularly for the largest instances. This can be explained by looking at the average number of nodes processed in eight hours, given in Table 5. From this, we see that for the largest instances the average number of nodes processed when using the LS formulation is close to just one. For these instances, all or nearly all of the eight hours are spent at the root node: solving the relaxation, generating cuts, running CPLEX's heuristics, and determining which variables to branch on. Thus, even in eight hours, the size of the LS formulation limits the amount of search for improved feasible solutions that can be performed.

Finally, Table 6 presents how long it takes to find the first feasible solution for each of the formulations (which in all cases is slightly longer than the time to solve the root relaxation), and the quality of the first feasible solution, measured as the average gap between the solution value and LB*. For the LS and LBL formulations, these results are the same with and without cuts, so only the without cuts case is reported. Table 6 indicates that the CM formulation is able to find a feasible solution much more quickly than the others, which is a natural consequence of the small relaxation solved. In addition, the quality of this solution is similar to the quality of the first solution found in much longer time by the other formulations, and for instances with fixed costs is often better.

6. Concluding Remarks

We have studied a strategic planning model that addresses the question of when to install technology in an environment in which technology is improving over time. We have developed a series of progressively more compact formulations that can be used to solve large-scale instances of this problem. Some of the results we have developed to strengthen these formulations have been implemented in practice and have led to significant reductions in computation time.

Each of the formulations we have presented may be useful, depending on the context. If a practitioner wants to implement a formulation with minimum effort, then the linearized formulation strengthened using ideas from lot sizing is a good choice. If a practitioner is willing to implement

Table 4. Solution quality results for large instances.

		Average percentage UB away from LB* after (2 hrs., 8 hrs.)				
FC?	n, T	LS	LS.C	LBL	LBL.C	CM
No	50, 10	(0.32, 0.31)	(0.35, 0.29)	(0.34, 0.31)	(0.31, 0.29)	(0.33, 0.36)
	50, 20	(2.11, 0.94)	(2.19, 1.40)	(1.38, 1.05)	(1.22, 0.89)	(0.81, 0.79)
	100, 10	(0.20, 0.17)	(0.17, 0.17)	(0.23, 0.19)	(0.18, 0.16)	(0.22, 0.23)
	100, 20	(1.40, 1.02)	(1.57, 1.17)	(1.58, 0.75)	(1.12, 0.63)	(0.91, 0.50)
Yes	50, 10	(0.54, 0.43)	(0.46, 0.43)	(0.47, 0.43)	(0.46, 0.43)	(0.54, 0.56)
	50, 20	(5.09, 1.41)	(5.08, 3.59)	(1.47, 1.45)	(1.49, 1.42)	(2.46, 1.52)
	100, 10	(0.39, 0.39)	(0.58, 0.38)	(0.43, 0.40)	(0.40, 0.38)	(0.52, 0.46)
	100, 20	(3.20,[a] 2.97)	(3.20,[a] 2.98)	(3.19, 1.31)	(3.19, 1.22)	(3.78, 1.63)

[a]Average is based on four of the five instances.

Table 5. Average nodes processed for large instances.

		Average nodes processed after 8 hours				
FC?	n, T	LS	LS.C	LBL	LBL.C	CM
No	50, 10	8,503.9	8,023.0	29,939.2	17,999.1	43,122.7
	50, 20	67.1	1.6	1,048.5	7.1	8,293.1
	100, 10	2,821.9	2,663.1	10,233.0	6,476.6	18,937.4
	100, 20	1.6	1.0	33.3	1.0	5,985.8
Yes	50, 10	2,719.0	1,540.1	9,030.1	8,815.7	44,127.5
	50, 20	8.0	1.0	297.6	199.1	4,704.8
	100, 10	508.0	255.2	2,449.3	2,511.2	15,545.1
	100, 20	1.0	1.0	1.2	3.2	549.2

a relatively simple routine for generating valid inequalities, the more compact formulation obtained from directly studying the mixed-integer bilinear model can yield faster computation times. Finally, the specialized branch-and-cut algorithm for the concave minimization formulation can be used to quickly generate good solutions and reasonable optimality bounds for large-scale instances. For very large-scale instances, just solving the relaxation of the MIP formulations may take prohibitively long, so that using the concave minimization formulation may be the only viable option.

We have assumed that variable costs depend only on the start period. A natural extension is to allow variable costs to depend both on the start period and on the period in which an activity is performed. This may be an important extension because, for example, it allows discounting to be incorporated into the model. To simplify the exposition, we have not considered this extension here, but in Luedtke (2007) we have shown that the majority of our results can be simply extended to this case. The only exception is the LS formulation. Although the LS formulation cannot be directly extended to this case, it is possible to use a hybrid between the LS and extended formulations that has $O(T^2)$ variables, but only $O(T)$ rows.

We have focused our discussion on the strategic production and distribution planning problem presented in §2. However, because our approach was based on developing

Table 6. Time and quality of first feasible solutions found.

		Time (s)			Average percentage away from LB* (%)		
FC?	n, T	LS	LBL	CM	LS	LBL	CM
No	50, 10	78.8	28.7	5.5	1.72	1.64	1.74
	50, 20	1,237.4	409.2	22.7	3.50	3.97	5.96
	100, 10	292.4	94.1	10.8	0.92	0.96	1.74
	100, 20	3,863.4	1,428.6	32.4	1.74	2.42	4.86
Yes	50, 10	189.9	50.6	12.7	2.56	2.46	2.17
	50, 20	2,188.4	665.7	132.1	6.54	7.00	4.27
	100, 10	650.2	150.9	33.3	1.92	2.31	1.60
	100, 20	10,765.2	2,530.9	373.9	3.30	4.20	2.98

strong formulations for a single activity, the formulations we have developed could actually be used much more generally. Specifically, it can be used in any context in which (1) technology (capacity) must be installed before an activity can be performed, (2) the installation is "all or nothing," (3) once installed the technology is fixed for the planning horizon, and (4) the technology installed determines the variable costs over the entire planning horizon.

In this study, we have assumed that the activity levels are constrained by an upper bound, common over all time periods. However, our approach can be used when more general constraints on the activity levels are present. One example of such a constraint, which was present in a variant of the application that motivated this work, is that the activities are restricted to be nondecreasing over time. Another plausible example is a ramping restriction on the activity levels, which would state that once an activity is begun, it must be performed within certain levels over time. Our approach can still be used to yield formulations and valid inequalities in these cases. However, it may be possible to make use of these additional restrictions on the activities over time to yield stronger formulations. This topic will be addressed in a companion paper.

Finally, the ideas we have proposed in this paper may be useful in other applications with bilinear objective functions.

Appendix

Proof of Theorem 8. Consider a generic valid inequality for E given by $\gamma\mu \geqslant \beta + \sum_{t=1}^{T} \alpha_t x_t$. We are interested only in nontrivial inequalities, i.e., those that are not implied by the bounds on x. Note that any such inequality will have a nonzero coefficient on μ. Furthermore, γ must be positive because otherwise the inequality would be violated by a solution $(\lambda, \mathbf{0})$ for large λ ($\lambda > \beta/\gamma$). Therefore, by scaling we can assume that any nontrivial inequality for E is of the form (26).

Now, observe that $E = \bigcup_{p=1}^{T+1} E_p$, where

$$E_p = \left\{ (\mu, x) \in \mathbb{R} \times [0, M]^T \colon \mu \geqslant f_p + c_p \sum_{t=p}^{T} x_t, \right.$$
$$\left. x_t = 0,\ t = 1, \ldots, p-1 \right\}.$$

Note that for each $p \in \mathcal{T}$, the extreme points of E_p are given by $x_t = M\ \forall\, t \in S$ and $x_t = 0$ otherwise, and $\mu = f_p + c_p M|S|$ for all $S \subseteq \{p, \ldots, T\}$. Also note that $E_{T+1} = \{(0, \mathbf{0})\}$.

Because any extreme point of $\operatorname{conv}(E)$ must be an extreme point of E_p for some $p \in \{1, \ldots, T+1\}$, we conclude that an inequality (26) is valid for E if and only if $\beta \leqslant 0$ and

$$M \max_{S \subseteq \{p, \ldots, T\}} \sum_{t \in S} (\alpha_t - c_p) + \beta \leqslant f_p \quad \forall\, p = 1, \ldots, T. \tag{27}$$

The pth condition in (27) is equivalent to $f_p \geqslant \beta + M \cdot \text{LP}'_p$, where

$$\begin{aligned} \text{LP}'_p = \max\ & \sum_{t=p}^{T} \omega_{pt}(\alpha_t - c_p) \\ \text{s.t.}\ & 0 \leqslant \omega_{pt} \leqslant 1 \quad \forall\, t = p, \ldots, T \\ = \min\ & \sum_{t=p}^{T} \sigma_{pt} \\ \text{s.t.}\ & \sigma_{pt} \geqslant \alpha_t - c_p \quad \forall\, t = p, \ldots, T, \\ & \sigma_{pt} \geqslant 0 \quad \forall\, t = p, \ldots, T, \end{aligned}$$

by linear programming duality. It follows that inequality (26) is valid for E if and only if there exists σ such that (β, α, σ) is feasible to (25).

It remains only to prove that v^* exists and is finite. This follows because the linear program (25) is feasible because $\mathbf{0}$ is a feasible solution, and bounded because

$$\begin{aligned} \beta + \sum_{t=1}^{T} \alpha_t x_t &\leqslant f_1 - M \sum_{t=1}^{T} \sigma_{1t} + \sum_{t=1}^{T} x_t(c_1 + \sigma_{1t}) \\ &\leqslant f_1 + c_1 \sum_{t=1}^{T} x_t. \quad \square \end{aligned}$$

Acknowledgments

This research was supported in part by the National Science Foundation under grants DMI-0100020, DMI-0121495, and DMI-0522485, and by a grant from ExxonMobil. The authors thank Cassandra McZeal for introducing them to the model presented in this paper. They also thank the anonymous referees for their comments, which helped improve the presentation of the paper.

References

Adams, W. P., H. D. Sherali. 1993. Mixed-integer bilinear programming problems. *Math. Programming* **59** 279–305.

Balas, E. 1998. Disjunctive programming: Properties of the convex hull of feasible points. *Discrete Appl. Math.* **89** 3–44.

Bárány, I., J. Edmonds, L. A. Wolsey. 1986. Packing and covering a tree by subtrees. *Combinatorica* **6** 221–233.

Bárány, I., T. J. Van Roy, L. A. Wolsey. 1984. Strong formulations for multi-item capacitated lot sizing. *Management Sci.* **30** 1255–1261.

Horst, R., H. Tuy. 1990. *Global Optimization*. Springer-Verlag, Berlin.

Hsu, V. N. 2002. Dynamic capacity expansion problem with deferred expansion and age-dependent shortage cost. *Manufacturing Service Oper. Management* **4** 44–54.

Li, S., D. Tirupati. 1994. Dynamic capacity expansion problem with multiple products: Technology selection and timing of capacity additions. *Oper. Res.* **42** 958–976.

Luedtke, J. 2007. Integer programming approaches for some non-convex and stochastic optimization problems. Ph. D. thesis, Georgia Institute of Technology, Atlanta. http://etd.gatech.edu.

Nemhauser, G. L., L. A. Wolsey. 1988. *Integer and Combinatorial Optimization*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, New York.

Rajagopalan, S. 1998. Capacity expansion and equipment replacement: A unified approach. *Oper. Res.* **46** 846–857.

Rajagopalan, S., A. C. Soteriou. 1994. Capacity acquisition and disposal with discrete facility sizes. *Management Sci.* **40** 903–917.

Rajagopalan, S., M. R. Singh, T. E. Morton. 1998. Capacity expansion and replacement in growing markets with uncertain technological breakthroughs. *Management Sci.* **44** 12–30.

Shulman, A. 1991. An algorithm for solving dynamic capacitated plant location problems with discrete expansion sizes. *Oper. Res.* **39** 423–436.

Van Roy, T. J., D. Erlenkotter. 1982. A dual-based procedure for dynamic facility location. *Management Sci.* **28** 1091–1105.