MANAGEMENT SCIENCE
Vol. 51, No. 11, November 2005, pp. 1706–1719
ISSN 0025-1909 | EISSN 1526-5501 | 05 | 5111 | 1706

DOI 10.1287/mnsc.1050.0378


# Integrated Lot Sizing in Serial Supply Chains with Production Capacities

Stan van Hoesel
Faculty of Economics and Business Administration, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands, s.vanhoesel@ke.unimaas.nl

H. Edwin Romeijn
Department of Industrial and Systems Engineering, University of Florida, 303 Weil Hall, P.O. Box 116595, Gainesville, Florida 32611-6595, romeijn@ise.ufl.edu

Dolores Romero Morales
Saïd Business School, University of Oxford, Park End Street, Oxford OX1 1HP, United Kingdom, dolores.romero-morales@sbs.ox.ac.uk

Albert P. M. Wagelmans
Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands, wagelmans@few.eur.nl

We consider a model for a serial supply chain in which production, inventory, and transportation decisions are integrated in the presence of production capacities and concave cost functions. The model we study generalizes the uncapacitated serial single-item multilevel economic lot-sizing model by adding stationary production capacities at the manufacturer level. We present algorithms with a running time that is polynomial in the planning horizon when all cost functions are concave. In addition, we consider different transportation and inventory holding cost structures that yield improved running times: inventory holding cost functions that are linear and transportation cost functions that are either linear, or are concave with a fixed-charge structure. In the latter case, we make the additional common and reasonable assumption that the variable transportation and inventory costs are such that holding inventories at higher levels in the supply chain is more attractive from a variable cost perspective. While the running times of the algorithms are exponential in the number of levels in the supply chain in the general concave cost case, the running times are remarkably insensitive to the number of levels for the other two cost structures.

*Key words*: lot sizing; integration of production planning and transportation; dynamic programming; polynomial time algorithms
*History*: Accepted by Thomas M. Liebling, mathematical programming and networks; received June 17, 2002. This paper was with the authors 11 months for 2 revisions.

## 1. Introduction

In this paper, we consider a problem in which production, inventory, and transportation decisions in a basic supply chain are integrated. Traditional models usually consider only one or two of these aspects in isolation from the other(s). Substantial evidence exists (see, for instance, Arntzen et al. 1995, Chandra and Fisher 1994, Geoffrion and Powers 1995, and Thomas and Griffin 1996, as well as the references therein) that shows that integrating these decisions can lead to substantial increases in efficiency and effectiveness. Integrating different decisions in the supply chain is particularly important when resources are limited and when costs are nonlinear, e.g., exhibit economies of scale.

We will consider a serial supply chain for the production and distribution of a product. Such a supply chain will occur, for instance, when value is added to a product in a sequence of production facilities, and intermediate goods need to be transported between these facilities. Kaminsky and Simchi-Levi (2003) describe an example of such a chain as it arises in the pharmaceutical industry.

Another example is the third-party logistics industry. In this case, a downstream distribution center that satisfies demands in a certain geographical area may employ the services of a third-party warehouse before products are transported to the actual distribution center for distribution to its retailers. A serial supply chain model can then be used to represent part of a supply chain that is relevant to the distribution center (see Lee et al. 2003). A final example is a situation in which production takes place at a manufacturer. The items that are produced are then stored at the manufacturer level or transported to the first warehouse level. At each of the warehouse levels, products are again either stored or transported to the warehouse at the next level. From the final warehouse level, products are then (possibly after having been stored for some periods) transported to a retailer (possibly allowing for early deliveries, i.e., inventories at the retailer level). Such a structure may arise if a retailer actually represents an entire market, and the supply chain from manufacturer to this market is very long. This could make it advantageous to, in several stages, employ economies of scale by transporting larger quantities over long distances to intermediate storage facilities before being distributed in the actual market.

All of the situations described above can be represented by a generic model consisting of a manufacturer, several intermediate production or distribution levels, and a level where demand for the end product takes place, which we will refer to in this paper as the retailer level (although this does not necessarily represent the level at which actual demand consumption takes place). In fact, in such a model the intermediate production and transportation stages are indistinguishable from one another, so that in the remainder of this paper we will simply refer to all intermediate stages as transportation stages between warehouses.

The serial supply chain model sketched above can be viewed as a generalization of a fundamental problem, which in fact is one of the most widely studied problems in production and inventory planning, the *economic lot-sizing problem* (*ELSP*). The basic variant of this problem considers a production facility that produces and stores a single product to satisfy known demands over a finite planning horizon. The problem is then to determine production quantities for each period such that all demands are satisfied on time at minimal total production and inventory holding costs. The cost functions are nondecreasing in the amount produced or stored, and are usually assumed to be linear, fixed-charge, or general concave functions. The production facility may or may not face a capacity constraint on the amount produced in each period.

To model the serial supply chain, the classical ELSP can be extended to include transportation decisions, as well as the possibility of holding inventory at different levels in the chain. In addition to production and inventory holding costs, we then clearly also need to incorporate transportation costs, which adds the problem of the timing of transportation to the problem of timing of production. The objective will be to minimize the systemwide cost while satisfying all demand. Even if the manufacturer and retailer are in fact distinct participants in the supply chain, each of which faces a part of the supply chain costs, this problem will be relevant. In this case, the participants clearly still need to decide how to distribute the minimal total costs, which is a coordination problem that is outside the scope of this paper. Alternatively, however, we may interpret the holding costs at the retailer level as a penalty or a discount on the purchasing price of an item, which is given by the manufacturer to the retailer if items are delivered early. In this case, the costs minimized by our optimization model are all incurred by the manufacturer. As in standard lot-sizing problems, all cost functions are assumed to be nondecreasing in the amount produced, stored, or shipped. In addition, we will assume that all cost functions are concave.

In general, all levels in a serial supply chain, regardless of whether they correspond to production or transportation decisions, may face capacities. In this paper, we will concentrate on serial supply chains with capacities at the production (i.e., first) level only, as a first step towards the study of more general capacitated supply chains. Adding capacities at other (i.e., transportation) levels appears to significantly change the structure of the problem, and thereby the problem analysis. Therefore, such problems are outside the scope of this paper, but remain a topic of ongoing research. Note that under certain cost structures it may be possible to eliminate capacitated levels from the supply chain. One such example is provided by Kaminsky and Simchi-Levi (2003), who transform a three-level serial supply chain model in which the first and third levels are capacitated to a two-level serial supply chain model with capacities at the first level only.

We will call the problem of determining optimal production, transportation, and inventory lot sizes in a serial supply chain as described above and under production capacities at the production level the *multilevel lot-sizing problem with production capacities* (*MLSP-PC*). In general, this problem is NP-hard, as it is a direct generalization of the NP-hard ELSP with general production capacities (see Florian et al. 1980). The ELSP with stationary production capacities, however, is solvable in polynomial time (see Florian and Klein 1971). Because our goal is to identify polynomially solvable cases of the MLSP-PC, we will assume in most of this paper that the production capacities are stationary.

We study problems with general concave production, inventory holding, and transportation costs, as well as problems with linear inventory holding costs and two different transportation cost structures: (i) linear transportation costs; and (ii) fixed-charge transportation costs without speculative motives, which means that with respect to variable costs, holding inventory is less costly at higher levels than at lower levels in the supply chain. Our solution methods are based on a dynamic programming framework that uses a decomposition principle that generalizes the classical zero-inventory ordering (ZIO) property of solutions to uncapacitated lot-sizing problems as described in Zangwill (1969) for the multilevel case, and, for instance, in Wagner and Whitin (1958) for the single-level case. In particular, in our two-level model we work with the new concept of a subplan, and show that extreme solutions decompose into a number of consecutive subplans. Our algorithms for this model all run in polynomial time in the planning horizon of the problem. The direct generalization of this approach to the multilevel case leads to a very large running time. We achieve substantial savings by introducing the concept of a relaxed subplan. In contrast to existing approaches in the literature, our dynamic program does not necessarily represent all (or even only) extreme point solutions to the MLSP-PC. In addition, while the paths in the dynamic program do all correspond to feasible solutions of the problem, the costs of a path may overestimate the costs of the corresponding solution to the problem. We are nevertheless able to prove (based on the concavity of the cost functions) that our dynamic program solves the MLSP-PC to optimality. The resulting algorithm for the case of general concave cost functions is exponential in the number of levels in the supply chain. However, it is remarkably insensitive to the number of levels for the two specific cost structures mentioned above.

This paper is organized as follows. In §2, we introduce the MLSP with production costs and general nondecreasing concave production, transportation, and inventory holding cost functions. We characterize the extreme points of the feasible region of the problem, and prove a decomposition result that will form the basis of our algorithms. In §3, we study the two-level problem and provide a general dynamic programming framework based on the decomposition result derived earlier, which yields a polynomial time algorithm in the planning horizon for general concave costs. In §4, this algorithm is then generalized to the multilevel lot-sizing problem and is shown to still be polynomial in the planning horizon, and better running times are given for two variants of the model. The paper ends in §5 with some concluding remarks and issues for further research.

## 2. Model Formulation and Analysis

### 2.1. The Model

As described in the introduction, we will study a multilevel lot-sizing problem with a serial structure. In each period, production may take place at the manufacturer. The items that are produced may be stored at the manufacturer level or transported to the first warehouse level. At each of the warehouse levels, products are again either stored or transported to the warehouse at the next level. From the final warehouse level products are then (possibly after having been stored for some period) transported to the retailer.

We consider a planning horizon of $T$ periods. In each period $t$, the retailer faces a nonnegative demand given by $d_t$, while the production capacity of the manufacturer in period $t$ is equal to $b_t$. We will consider a total of $L$ levels, which includes the manufacturer, the retailer, and $L-2$ intermediate warehouses. We say that the manufacturer is at the first level of the chain, and the retailer is at the $L$th level. Each of the intermediate levels corresponds to a warehouse. Let $\mathbb{R}^+$ denote the set of nonnegative real numbers. For each period $t = 1, \ldots, T$, the production costs are given by the function $p_t\colon \mathbb{R}^+ \to \mathbb{R}^+$, the transportation costs from level $\ell$ to level $\ell+1$ are given by the function $c_t^\ell\colon \mathbb{R}^+ \to \mathbb{R}^+$ ($\ell = 1, \ldots, L-1$), and the inventory holding costs at level $\ell$ are given by the function $h_t^\ell\colon \mathbb{R}^+ \to \mathbb{R}^+$ ($\ell = 1, \ldots, L$). Throughout the paper, we will assume that all cost functions are concave, nondecreasing, and equal to zero when their argument is zero.

The MLSP-PC can be formulated as follows:

$$\text{minimize} \sum_{t=1}^{T}\left(p_t(y_t)+\sum_{\ell=1}^{L-1}c_t^\ell(x_t^\ell)+\sum_{\ell=1}^{L}h_t^\ell(I_t^\ell)\right) \qquad \text{(P)}$$

subject to

$$x_t^1+I_t^1=y_t+I_{t-1}^1,\quad t=1,\ldots,T, \qquad (1)$$

$$x_t^\ell+I_t^\ell=x_t^{\ell-1}+I_{t-1}^\ell,\quad t=1,\ldots,T;\ \ell=2,\ldots,L-1, \qquad (2)$$

$$d_t+I_t^L=x_t^{L-1}+I_{t-1}^L,\quad t=1,\ldots,T, \qquad (3)$$

$$y_t\le b_t,\quad t=1,\ldots,T, \qquad (4)$$

$$I_0^\ell=0,\quad \ell=1,\ldots,L, \qquad (5)$$

$$y_t\ge 0,\quad t=1,\ldots,T,$$

$$x_t^\ell\ge 0,\quad t=1,\ldots,T;\ \ell=1,\ldots,L-1,$$

$$I_t^\ell\ge 0,\quad t=1,\ldots,T;\ \ell=1,\ldots,L,$$

where $y_t$ denotes the quantity produced in period $t$, $x_t^\ell$ is the quantity shipped from level $\ell$ to level $\ell+1$ in period $t$, and $I_t^\ell$ denotes the inventory quantity at level $\ell$ at the end of period $t$. Constraints (1)–(3) model the balance between inflow, storage, and outflow at the manufacturer, warehouse, and retailer levels, respectively, in each period. The production quantity in each period is restricted by constraints (4). Finally, constraints (5) state that all initial inventory levels are equal to zero. Unlike in the traditional single-level lot-sizing model, this is not an assumption that we can make without loss of generality, due to the nonlinearity of the transportation and inventory holding cost functions. Therefore, we will later discuss how to deal with problem instances where this constraint is absent, and instead (nonnegative) initial inventory quantities at all levels are considered as part of the problem data. Then, the developed algorithms can be applied in a rolling horizon scheme, in which new lot-sizing instances are solved—and their optimal solutions partially implemented—as time proceeds and new demand forecasts become available.

For convenience, we will define $d_{ts}$ to be the cumulative demand in periods $t,\dots,s$, i.e.,

$$d_{ts} \equiv \begin{cases} \sum_{\tau=t}^{s} d_\tau & \text{for } t=1,\dots,s;\ s=1,\dots,T, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

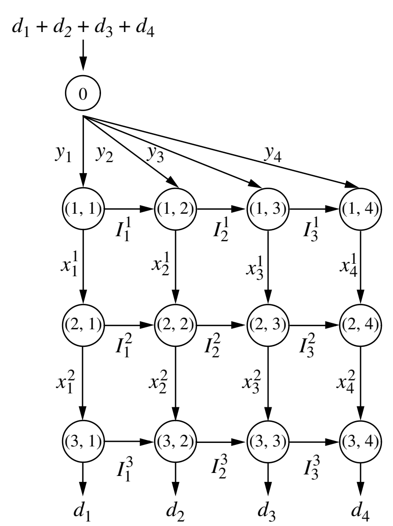
To ensure feasibility of (P), we will assume that the cumulative demand in the first $t$ periods cannot exceed the total production capacity in these periods, i.e.,

$$d_{1t} \le \sum_{\tau=1}^{t} b_\tau \quad \text{for each } t=1,\dots,T. \tag{7}$$

It is easy to see that this condition is both necessary and sufficient for (P) to have a nonempty feasible region.

We can also model the MLSP-PC as a capacitated minimum-cost network flow problem in a network with one source (see also Zangwill 1969 for a general discussion on such minimum-cost network flow problems, as well as a discussion of uncapacitated multilevel ELSPs). To this end, we define a network with a single source 0, $T$ transshipment nodes $(1,t)$ at the production level (level 1, $t=1,\dots,T$), $T$ transshipment nodes $(\ell,t)$ at each of the warehouse levels ($t=1,\dots,T$; $\ell=2,\dots,L-1$), and $T$ demand nodes $(L,t)$ with demand $d_t$ at the retailer level (level $L$, $t=1,\dots,T$). Finally, feasibility dictates that the source node 0 has a supply of $d_{1T}$ units. Figure 1 illustrates the network representation of the MLSP-PC for $L=3$ and $T=4$. This representation will facilitate the analysis of the structure of extreme points of the feasible region of (P) in §2.4. Before proceeding with this analysis, in §2.2 we will discuss related models and algorithms from the literature, as well as some special cases that reduce to single-level models in §2.3.

**Figure 1** **Network Representation of the MLSP-PC for $L=3$ and $T=4$**




### 2.2. Literature Review

The single-level variant of the MLSP-PC has received a lot of attention in the literature. The uncapacitated problem, the ELSP, is solvable in polynomial time in the length of the time horizon; see Wagner (1960) for this basic result. More efficient algorithms for special cases have been developed by Aggarwal and Park (1993), Federgruen and Tzur (1991), and Wagelmans et al. (1992). When production capacities are present, we obtain the so-called *capacitated lot-sizing problem* (*CLSP*). In contrast to the uncapacitated ELSP, this problem is known to be NP-hard, even in many special cases; see Florian et al. (1980) and Bitran and Yanasse (1982). An interesting and important special case that does allow for a polynomial time algorithm arises when the production capacities are stationary; see, e.g., Florian and Klein (1971), Florian et al. (1980), and van Hoesel and Wagelmans (1996). See also references in Baker et al. (1978) for other work on the CLSP with stationary production capacities, and Chung and Lin (1988) and van den Heuvel and Wagelmans (2003) for another special case of the CLSP that is solvable in polynomial time.

Zangwill (1969) studied the uncapacitated version of the MLSP-PC, and developed a dynamic programming algorithm that is polynomial in both the planning horizon and the number of levels $L$. We analyze this algorithm in the online appendix (available at http://mansci.pubs.informs.org/ecompanion.html) and conclude that it runs in $O(LT^4)$ time, where $L$ is the number of levels, or even in $O(T^3)$ for the special case of $L=2$. Lee et al. (2003) consider a two-level model where the transportation costs are nonconcave functions.

A study that is related to ours in the sense that it also considers capacities in a multilevel setting is the one by Kaminsky and Simchi-Levi (2003). They propose a three-level model in which the first and third levels are production stages, and the second level is a transportation stage. Both production stages are capacitated, while the transportation stage is uncapacitated. They consider linear inventory holding costs that increase with the level of the supply chain, and linear production costs at both levels 1 and 3 that satisfy a traditional nonspeculative motives condition (see also §2.3). The transportation costs at the second level are of the fixed-charge or general concave form and are assumed to satisfy a restrictive and nontraditional nonspeculative motives condition. By eliminating the third-level production decisions, they reduce the problem to a two-level model that inherits its cost function structures from the three-level model. For their class of fixed-charge transportation costs, they provide an $O(T^4)$ algorithm to solve the model, even in the case of nonstationary production capacities. For their class of concave transportation costs they provide an $O(T^8)$ algorithm to solve the model in the presence of stationary production capacities. They pose the complexity of their model for more general cost structures as an open question. In this paper, we address this question by deriving an $O(T^7)$ algorithm for solving the two-level problem in the presence of stationary capacities.

### 2.3. Special Cases

It is common in lot-sizing problems to model the inventory holding costs as linear functions, i.e., $h_t^\ell(I_t^\ell)=h_t^\ell I_t^\ell$ for $t=1,\ldots,T$; $\ell=1,\ldots,L$, with $h_t^\ell\ge 0$ for all $t$ and $\ell$. We will therefore consider this class of problems in §§4.3 and 4.4. In §4.3, we will in addition assume that the transportation costs have a fixed-charge structure without speculative motives. More formally, $c_t^\ell(x)=f_t^\ell 1_{\{x>0\}}+g_t^\ell x$, where $1_{\{x>0\}}$ is an indicator function taking the value 1 if $x>0$, and 0 otherwise. The assumption that there are no speculative motives, which is commonly assumed for the production and inventory holding costs in traditional economic lot-sizing models, means in this context that, with respect to variable inventory and transportation costs only, it is attractive to transport as late as possible. More formally, $g_t^\ell+h_t^{\ell+1}\ge h_t^\ell+g_{t+1}^\ell$ for $t=1,\ldots,T-1$; $\ell=1,\ldots,L-1$.

Note that if the transportation cost functions are both linear and exhibit no speculative motives, it is always optimal to store production at the manufacturer and transport only when demand needs to be satisfied. Hence, without loss of optimality, we can assume that $I_t^\ell=0$ for all $t=1,\ldots,T$ and $\ell=2,\ldots,L$. Similarly, if the transportation costs are linear and $g_t^\ell+h_t^{\ell+1}\le h_t^\ell+g_{t+1}^\ell$ for $t=1,\ldots,T-1$; $\ell=1,\ldots,L-1$, it is cheaper to transport as soon as we produce and store the production at the retailer level. Then, without loss of optimality, we can assume that $I_t^\ell=0$ for all $t=1,\ldots,T$ and $\ell=1,\ldots,L-1$. These two special cases of the MLSP-PC therefore yield a standard CLSP.

Finally, a variant of the uncapacitated two-level MLSP-PC can easily be reduced to an uncapacitated ELSP. When production costs as well as the inventory holding costs at both levels are linear, given that we decide to transport in a certain period, we can easily determine the best production period, i.e., the period that yields the minimum total unit production and manufacturer-level inventory costs for transport in period $t$. Redefining the transportation cost function accordingly, which can be done in $O(LT)$ time, allows us to eliminate the production variables as well as the inventory variables at the manufacturer, yielding a standard uncapacitated ELSP. The resulting problem can be solved in $O(T^2)$ time for general concave transportation costs (see Wagner 1960), and in $O(T\log T)$ time for fixed-charge transportation costs (see Aggarwal and Park 1993, Federgruen and Tzur 1991, and Wagelmans et al. 1992).

### 2.4. Characterization of Extreme Points

Problem (P) has a concave objective function, and its feasible region is defined by linear constraints. This implies that there exists an extreme point optimal solution to (P). Consider the flow in the network corresponding to any extreme point feasible solution. As is common in network flow problems, we will call the arcs that carry an amount of flow that is both strictly positive and strictly less than its capacity *free arcs*. It is well known (see, e.g., Ahuja et al. 1993) that the subnetwork containing only the free arcs contains no cycle.

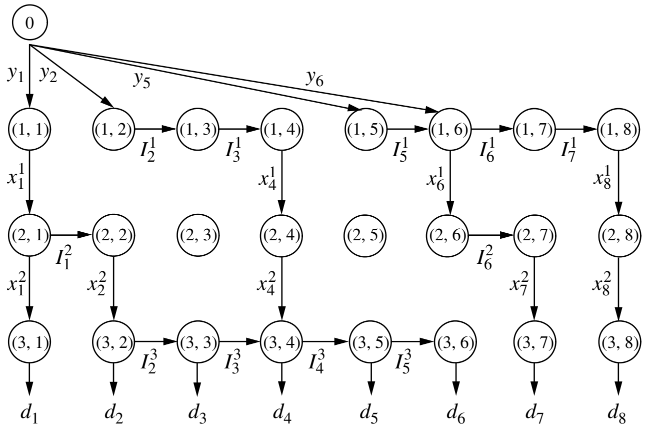
**2.4.1. Subplans.** Note that only arcs that have a finite upper bound (which in our case are only the production arcs) may carry flow while they are not free. Removing all production arcs, the network containing all remaining free arcs decomposes into a number of connected components. Limiting ourselves for now to connected components that do indeed carry flow, we identify the first and last nodes in the component at each level. For a given component, these nodes can be denoted by $(\ell,s_{\ell 1}+1)$ and $(\ell,s_{\ell 2})$ for $\ell=1,\ldots,L$, where

$$s_{\ell 1}\le s_{\ell+1,1}<s_{\ell 2}\le s_{\ell+1,2}\quad \text{for } \ell=1,\ldots,L-1. \tag{8}$$

(Note that the strict inequality holds due to the definition of the subplan: The first period included at level $\ell$ is $s_{\ell 1}+1$.) With this approach, some nodes may be isolated and not included in any component carrying flow. We assign each of those to the component that is adjacent to the left of them. The assignment of the isolated nodes is illustrated in Figure 2. After eliminating the production arcs, we obtain two components. The first one is defined by the nodes $(1,1)$ and $(1,4)$ in Level 1, $(2,1)$ and $(2,4)$ in Level 2, and $(3,1)$ and $(3,6)$ in Level 3, and the second one by nodes $(1,5)$ and $(1,8)$ in Level 1, $(2,6)$ and $(2,8)$ in Level 2, and $(3,7)$ and $(3,8)$ in Level 3. We may observe that node $(2,3)$ is part of the first component, even though no flow passes through this node. As mentioned above, the isolated node $(2,5)$ is assigned to the left component.

Summarizing, we can decompose an extreme point solution to (P) into components, each of which contains a set of nodes $\{(\ell,s_{\ell 1}+1),\ldots,(\ell,s_{\ell 2})\}$ $(\ell=1,\ldots,L)$ satisfying (8). We will call the components thus

**Figure 2** **The Structure of an Extreme Point Solution to the MLSP-PC, $L=3$ and $T=8$**




obtained *subplans*. We will represent a subplan by the $2L$ periods that identify it: $((\varsigma_{\ell 1},\varsigma_{\ell 2})_{\ell=1}^L)$. It will often be convenient to refer to the production and demand periods in a subplan separately, and we will then often use the notation $(t_1,t_2,\tau_1,\tau_2)\equiv(\varsigma_{11},\varsigma_{12},\varsigma_{L1},\varsigma_{L2})$. By construction, no inventory is carried between subplans, so the only flow entering a subplan comes from production arcs associated with the manufacturer nodes in the subplan. The total quantity produced in all production periods in the subplan, i.e., the total production in periods $t_1+1,\ldots,t_2$, is used to supply the demand of all retailer nodes in the subplan, i.e., the total demand in periods $\tau_1+1,\ldots,\tau_2$. We will call two subplans $((\varsigma_{\ell 1},\varsigma_{\ell 2})_{\ell=1}^L)$ and $((\varsigma'_{\ell 1},\varsigma'_{\ell 2})_{\ell=1}^L)$ *consecutive* if $\varsigma'_{\ell 1}=\varsigma_{\ell 2}$ for $\ell=1,\ldots,L$. We can summarize the structure of extreme point solutions as follows.

Proposition 2.1. *Any extreme point feasible solution can be decomposed into a sequence of consecutive subplans.*

The extreme solution given in Figure 2 decomposes into two subplans, namely, $((0,4),(0,5),(0,6))$ and $((4,8),(5,8),(6,8))$.

Note that the first subplan obtained by decomposing an extreme point solution as described above has $\varsigma_{\ell 1}=0$ for $\ell=1,\ldots,L$. However, in the remainder of this paper it will be convenient to also include subplans $((\varsigma_{\ell 1},\varsigma_{\ell 2})_{\ell=1}^L)$ satisfying (8), for which some but not all values of $\varsigma_{\ell 1}$ are zero.

**2.4.2. Production Quantities in a Subplan.** The fact that the extreme flows are acyclic implies that, although there may be multiple production arcs associated with a subplan that carry flow, there is at most one such arc with production below capacity. In other words, there is at most one free production arc entering the subplan. This yields the following generalization of the characterization of extreme points of single-level CLSPs by Florian and Klein (1971).

Proposition 2.2. *A subplan can contain at most one free production arc.*

If the problem is uncapacitated, this proposition implies that only one production arc carrying flow enters each of the subplans, which in turn means that the extreme flows are arborescent. The dynamic programming algorithm proposed for this problem by Zangwill (1969) is based on this property; see the online appendix.

As an example, in Figure 2 we know that in the subplan $((0,4),(0,5),(0,6))$ the production arcs $y_1$ and $y_2$ cannot both be free; the same holds for production arcs $y_5$ and $y_6$ in subplan $((4,8),(5,8),(6,8))$.

**2.4.3. Transportation Quantities in a Subplan.** The absence of cycles consisting of free arcs only in an extreme point solution can also be used to identify structural properties of the transportation quantities. Consider a period, say $t$, in which transportation takes place between levels $\ell$ and $\ell+1$, i.e., the flow on the arc between nodes $(\ell,t)$ and $(\ell+1,t)$ is $x_t^\ell>0$. Two situations can then occur with respect to the total flow into nodes $(\ell+1,\varsigma_{\ell+1,1}+1),\ldots,(\ell+1,t)$, i.e., the *cumulative* shipments between levels $\ell$ and $\ell+1$ up to and including period $t$ within the subplan:

- It is equal to the cumulative production in periods $t_1+1,\ldots,s$ for some $s\in\{t_1+1,\ldots,t\}$;
- It satisfies the demand of periods $\tau_1+1,\ldots,s$ for some $s\in\{\tau_1+1,\ldots,\tau_2\}$.

If not, consider the last production period in which some of the transported quantity $x_t^\ell$ was produced, say $s'$. There will then be a period whose demand is satisfied partially from the quantity $x_t^\ell$ and partially from production in period $s'$ that remains in inventory at level $\ell$ at the end of period $t$, creating a cycle containing only free arcs. This result can be summarized as follows.

Proposition 2.3. *In a subplan, the transported quantity between levels $\ell$ and $\ell+1$ in some period either makes the cumulative transported quantities thus far in the subplan equal to the cumulative production quantities of an initial sequence of consecutive production periods in the subplan, or to the cumulative demand of an initial sequence of demand periods in the subplan.*

The two possibilities for cumulative transport can be illustrated using Figure 2. In subplan $((0,4),(0,5),(0,6))$,

- $x_1^1$ is equal to the (cumulative) production in Period 1, while $x_1^1+x_2^1+x_3^1+x_4^1$ is both equal to the cumulative production in Periods $1,\ldots,4$ and satisfies the demand of Periods $1,\ldots,6$;
- $x_1^2$ satisfies the demand of Period 1; $x_1^2+x_2^2$ is equal to the (cumulative) production in Period 1; and $x_1^2+x_2^2+x_3^2+x_4^2$ is both equal to the cumulative production in Periods $1,\ldots,4$ and satisfies the demand of Periods $1,\ldots,6$.

# 3. The Two-Level Capacitated Lot-Sizing Problem with Concave Costs

For clarity of exposition, we will first consider the two-level version of the MLSP-PC, which we will call the 2LSP-PC. In the next section, we will show how the methodology can be extended to chains with more than two levels.

## 3.1. A Dynamic Programming Approach

In this section, we will outline a general dynamic programming approach to the 2LSP-PC. This approach is based on the decomposition of extreme point solutions to (P) into consecutive subplans (see Proposition 2.1). In particular, define $F(t,\tau)$ to be the minimum cost associated with satisfying the retailer demands in periods $\tau+1,\dots,T$ using production in periods $t+1,\dots,T$. We are then clearly interested in computing $F(0,0)$. This can be achieved using a two-phase approach:

*Phase* 1. For each subplan $(t_1,t_2,\tau_1,\tau_2)$, compute the minimum costs that are incurred for satisfying the demand of that subplan under the condition that at most one free production arc enters the subplan. Denote these costs by $\phi(t_1,t_2,\tau_1,\tau_2)$.

*Phase* 2. Compute the values $F(t_1,\tau_1)$ for all $0\le t_1\le \tau_1\le T$ by realizing that an extreme point solution to the corresponding subproblem is given by a subplan $(t_1,t_2,\tau_1,\tau_2)$ and the remaining subproblem $F(t_2,\tau_2)$ for some $t_2$ and $\tau_2$. This gives rise to the following backward recursion:

$$F(t_1,\tau_1)=\min_{(t_2,\tau_2):\tau_2\ge t_2>\tau_1}\{\phi(t_1,t_2,\tau_1,\tau_2)+F(t_2,\tau_2)\}\quad \text{for } 0\le t_1\le\tau_1<T,$$

$$F(t_1,T)=0\quad \text{for } 0\le t_1\le T.$$

Note that in Phase 1 we need to compute $O(T^4)$ values. Phase 2 is, in fact, a shortest path problem in a network with nodes representing all period-pairs $(t,\tau)$ such that $0\le t\le\tau\le T$, and arcs representing the subplans with corresponding costs. The minimum-cost path from node $(0,0)$ to any of the nodes $(t_1,T)$ in this acyclic network can be found in linear time in the number of arcs, i.e., in $O(T^4)$ time (see Ahuja et al. 1993). Florian and Klein (1971) used this general dynamic programming framework to develop an $O(T^4)$ dynamic programming algorithm to solve the CLSP with stationary capacities and general concave production and inventory holding cost functions.

When the value of $\phi(\cdot)$ is given for each subplan, the 2LSP-PC is polynomially solvable. To achieve a polynomial time algorithm for the 2LSP-PC, the challenge is therefore to provide a polynomial time algorithm for computing the costs corresponding to all subplans. Because we know that the 2LSP-PC is NP-hard for general production capacities, we will restrict our attention to the case of stationary production capacities, i.e., $b_t=b$ for $t=1,\dots,T$. In the remainder of this section, we will derive a polynomial time algorithm for computing the optimal costs of all subplans, and thereby for the 2LSP-PC. Before studying the subproblems of computing the optimal subplan costs, we will first study the implications of the assumption that the production capacities are stationary in the next section.

## 3.2. Implications of Stationary Production Capacities

In Phase 1 of the dynamic programming approach, we need to compute the optimal costs of all subplans, under the additional constraint that all but one of the production arcs entering the subplan carry a flow equal to 0 or $b$. Consider a particular subplan, say $(t_1,t_2,\tau_1,\tau_2)$, in which the total demand of periods $\tau_1+1,\dots,\tau_2$ needs to be satisfied using production in periods $t_1+1,\dots,t_2$. Following Florian and Klein (1971), note that the constraint on the values of the production arcs entering the subplan implies that the number of production arcs that carry flow equal to the production capacity is exactly equal to $K=\lfloor d_{\tau_1+1,\tau_2}/b\rfloor$, and the remaining production quantity is equal to $\varepsilon=d_{\tau_1+1,\tau_2}-Kb$. Clearly, we have that $0\le\varepsilon<b$. If $\varepsilon>0$, there will be exactly one production arc entering the subplan carrying this flow.

## 3.3. The Subplan Costs

We will formulate the problem of determining the optimal costs of a subplan as a dynamic programming problem. Put differently, for each subplan $(t_1,t_2,\tau_1,\tau_2)$, we will define a network with the property that $\phi(t_1,t_2,\tau_1,\tau_2)$ is equal to the length of the shortest path between a pair of source and sink nodes in this network.

We choose the nodes in this network to be of the form $(t,Y,X)$, where $t$ indicates a period, $Y$ is equal to the cumulative production quantity up to and including period $t$, and $X$ is equal to the cumulative transportation quantity up to and including period $t$. Node $(t_1,0,0)$ is the source, while node $(t_2,Kb+\varepsilon,Kb+\varepsilon)\equiv(t_2,d_{\tau_1+1,\tau_2},d_{\tau_1+1,\tau_2})$ is the sink. By Proposition 2.2 and the discussion in §3.2, we know that the production quantity in any period can only assume one of the values $\{0,\varepsilon,b\}$, with the value $\varepsilon$ only in one period. This immediately implies that $Y$ can only assume the values

$$Y\in\bigcup_{k=0}^{K}\{kb,kb+\varepsilon\},$$

where, in addition, $Y=0$ if $t=t_1$, $d_{\tau_1+1,t}\le Y\le(t-t_1)K$ for $t=t_1+1,\dots,t_2-1$, and $Y=Kb+\varepsilon$ if $t\ge t_2$ to ensure that all demand is produced within the set of production periods allowed in the subplan. Because clearly $K \le T$, the number of allowable values for $Y$ is $O(T)$. Furthermore, by Proposition 2.3 we know that the cumulative transported quantity up to and including some period is either equal to the total production quantity of an initial sequence of production periods or satisfies the demand of an initial sequence of demand periods in the subplan. More formally, this means that

$$X \in \left(\bigcup_{k=0}^{K}\{kb, kb+\varepsilon\}\right) \cup \left(\bigcup_{s=\tau_1+1}^{\tau_2}\{d_{\tau_1+1,s}\}\right),$$

where, in addition, $d_{\tau_1+1,t} \le X \le Y$ to ensure that demands are satisfied on time and products are not transported before they are produced, and $X=0$ if $t \le \tau_1$ and $X = Kb+\varepsilon$ if $t \ge t_2$ to ensure that transportation takes place within the subplan. The number of allowable values for $X$ is thus $O(T)$ as well, so that the total number of nodes in the network is $O(T^3)$.

Arcs in the network represent production, transportation, and inventory decisions. Arcs are present between pairs of nodes in the network of the form $(t, Y, X)$ and $(t+1, \overline{Y}, \overline{X})$, with $\overline{Y} - Y \in \{0, \varepsilon, b\}$ (where the value $\varepsilon$ is only allowed if $Y = kb$ for some $k = 0, 1, \ldots, K$), and $\overline{X} \ge X$ (where $\overline{X} \in \{X, \overline{Y}\} \cup (\bigcup_{s=\tau_1+1}^{\tau_2}\{d_{\tau_1+1,s}\})$). It is easy to see that there are $O(T)$ arcs emanating from each node in the network, so that the entire network has $O(T^4)$ arcs.

From the information contained in the nodes defining an arc, we can easily compute the production quantity in period $t+1$ $(\overline{Y} - Y)$, the transportation quantity in period $t+1$ $(\overline{X} - X)$, the inventory held at the manufacturer level at the end of period $t+1$ $(\overline{Y} - \overline{X})$, and the inventory held at the retailer level at the end of period $t+1$ $(\overline{X} - d_{\tau_1+1,t+1})$. The costs of an arc are thus given by

$$p_{t+1}(\overline{Y} - Y) + c^1_{t+1}(\overline{X} - X) + h^1_{t+1}(\overline{Y} - \overline{X}) + h^2_{t+1}(\overline{X} - d_{\tau_1+1,t+1}).$$

If all cost functions can be evaluated in constant time, the costs of a given arc can be computed in constant time provided that we determine all cumulative demands $d_{tt'}$ (in $O(T^2)$ time) in a preprocessing step.

Any path in the network from the source $(t_1, 0, 0)$ to the sink $(t_2, d_{\tau_1+1,\tau_2}, d_{\tau_1+1,\tau_2})$ represents a feasible flow in the subplan $(t_1, t_2, \tau_1, \tau_2)$ with just one free production arc. Moreover, it is easy to see that the reverse is also true. Therefore, the subplan costs are given by the minimal cost path in this network from the source node to the sink node. The time required for finding this minimal cost path is proportional to the number of arcs in the network, so that the cost of a single subplan can be determined in $O(T^4)$ time.

Because there are $O(T^4)$ subplans, a straightforward application of the dynamic programming algorithm defined above to each individual subplan would yield an algorithm with running time $O(T^8)$ for computing the costs of all subplans. However, the running time can be reduced by observing that the costs of many subplans are related. In particular, observe that the dynamic programming network corresponding to any subplan of the form $(t_1, t_2, \tau_1, \tau_2)$ is actually a subnetwork of the dynamic programming network for the subplan $(0, t_2, \tau_1, \tau_2)$. Therefore, using backward recursion to solve for the shortest path between nodes $(0,0,0)$ and $(t_2, d_{\tau_1+1,\tau_2}, d_{\tau_1+1,\tau_2})$ in the latter network yields, as a byproduct, the shortest paths between nodes $(t, 0, 0)$ and $(t_2, d_{\tau_1+1,t_2}, d_{\tau_1+1,t_2})$ for each $t = 1, \ldots, \tau_1$. It thus follows that we only need to consider the $O(T^3)$ subplans of the form $(0, t_2, \tau_1, \tau_2)$, the costs of which can be determined in $O(T^7)$ time.

### 3.4. Dealing with Initial Inventories

If the initial inventories at the manufacturer and/or retailer levels, $I^1_0$ and $I^2_0$, are strictly positive, there is a slight change in the construction of subplans. Recall that we construct subplans corresponding to a given extreme point solution by considering all arcs (except production arcs) that carry positive flow. The subplans are then formed by the resulting connected components together with some isolated nodes. When there are initial inventories, however, there may be one or more components that carry flow but do not contain a production period. In these components, demand is satisfied using initial inventories at warehouse and retailer levels only, and they can be assigned to the component containing production Period 1 (i.e., the component containing node $(1,1)$). The results in §§2.4.2 and 2.4.3 are clearly still valid for subplans in which $t_1 > 0$. However, for subplans with $t_1 = \tau_1 = 0$, the results continue to hold provided we view the total initial inventories $I^1_0 + I^2_0$ as a cumulative production quantity up to and including Period 0, and the initial inventory $I^2_0$ at Level 2 as the cumulative transportation quantity up to and including Period 0. Unless $\tau_2 = T$, these subplans can only have a feasible solution if the total initial inventories do not exceed the total demand that needs to be satisfied in the subplan. For subplans with $d_{1\tau_2} \ge I^1_0 + I^2_0$, we obtain $K = \lfloor (d_{1\tau_2} - I^1_0 - I^2_0)/b \rfloor$ and $\varepsilon = d_{1\tau_2} - I^1_0 - I^2_0 - Kb$. As already mentioned, $d_{1\tau_2} < I^1_0 + I^2_0$ can only occur if $\tau_2 = T$. If indeed $d_{1T} < I^1_0 + I^2_0$, an extreme point solution will contain only a single subplan: $(0, T, 0, T)$, and no production will take place in any period in that subplan, i.e., $K = \varepsilon = 0$. The only remaining difficulty in this case is that we do not want to specify in advance in which level the excess inventory will end up as ending inventory. This can easily be dealt with by extending the planning horizon by one period, say $T+1$. Then, define the production cost function for that period as $p_{T+1}(0) = 0$ and $p_{T+1}(y_{T+1}) = \infty$ for all $0 < y_{T+1} \leq b$ and the transportation cost function as $c^1_{T+1}(x_{T+1}) = 0$ for all $x_{T+1} \geq 0$. Finally, set $d_{T+1} = I^1_0 + I^2_0 - d_{1T}$. The costs of the single subplan $(0, T, 0, T)$ in the original problem can then be found by finding the costs of the subplan $(0, T+1, 0, T+1)$ in the modified problem.

Now consider the dynamic programming network used to compute the costs of a subplan. For subplans that contain initial inventories, we let the source node be $(0, I^1_0 + I^2_0, I^2_0)$ and the sink node be $(t_2, I^1_0 + I^2_0 + Kb + \varepsilon, I^1_0 + I^2_0 + Kb + \varepsilon)$. For a state $(t, Y, X)$, this also means that

$$Y \in \bigcup_{k=0}^{K} \{I^1_0 + I^2_0 + kb, I^1_0 + I^2_0 + kb + \varepsilon\}$$

and

$$X \in \left(\bigcup_{k=0}^{K} \{I^1_0 + I^2_0 + kb, I^1_0 + I^2_0 + kb + \varepsilon\}\right) \cup \left(\bigcup_{s=\tau_1+1}^{\tau_2} \{d_{1s}\}\right).$$

Finally, note that for subplans with $t_1 > 0$, we should have no positive inventory inflow. Therefore, in case there are nonzero initial inventory levels, we actually need to compute the costs of all subplans $(0, t_2, 0, \tau_2)$ while taking into account the initial inventory levels, as well as the costs of all subplans $(0, t_2, \tau_1, \tau_2)$ for $\tau_1 > 0$ without taking into account the initial inventory level at the manufacturer. This clearly does not influence the overall running time of the algorithm.

# 4. The Multilevel Case

## 4.1. Introduction

We may extend the dynamic programming approach developed in §3.1 for the two-level case to the multilevel case, where again a Phase 2 dynamic programming network represents all extreme point solutions to the MLSP-PC. To this end, we should define $\tilde{F}((\varsigma_{\ell 1})^L_{\ell=1})$ to be the minimum cost associated with satisfying the retailer demands in periods $\{\varsigma_{L1}+1, \ldots, T\}$ using production in periods $\{\varsigma_{11}+1, \ldots, T\}$, and warehouse $\ell$ in periods $\{\varsigma_{\ell+1,1}+1, \ldots, T\}$ for each $\ell = 1, \ldots, L-2$. We would then be interested in computing $\tilde{F}((0)^L_{\ell=1})$. It is easy to see that the running time of the corresponding generalization of the Phase 2 dynamic program would be $O(T^{2L})$. In this section, we will derive a modification of the Phase 2 dynamic program that runs in $O(T^4)$ time. This modification does not make Phase 1 computationally more expensive, and may even make it less expensive.

In particular, we will develop a more efficient approach in which the Phase 2 dynamic program does not necessarily represent all (or even only) extreme point solutions to the MLSP-PC, and in addition overestimates the costs of many of the nonextreme point solutions that it represents. However, as we will show, it does contain an optimal extreme point solution and is guaranteed to find this solution. This approach is based on the idea that the most important information present in the definition of a subplan is the set of production periods $t_1+1, \ldots, t_2$ and the set of demand periods $\tau_1+1, \ldots, \tau_2$. The basis of our improved algorithm is then to allow transportation in the periods $t_1+1, \ldots, \tau_2$ (while of course retaining the given production and demand periods). We can then use the same dynamic programming approach as in the two-level case, where we replace the two-level subplan costs $\phi(t_1, t_2, \tau_1, \tau_2)$ by the minimum costs of satisfying demand in periods $\tau_1+1, \ldots, \tau_2$ using production in periods $t_1+1, \ldots, t_2$, where at most one of the production quantities may be different from both 0 and $b$, and where transportation at all levels is allowed in periods $t_1+1, \ldots, \tau_2$. We will denote the latter costs by $\psi(t_1, t_2, \tau_1, \tau_2)$, and refer to vectors $(t_1, t_2, \tau_1, \tau_2)$ as *relaxed subplans*.

To illustrate the concept of relaxed subplans, consider the following problem instance of the 2LSP-PC. All demands are equal to 1; the production and transportation costs are given by

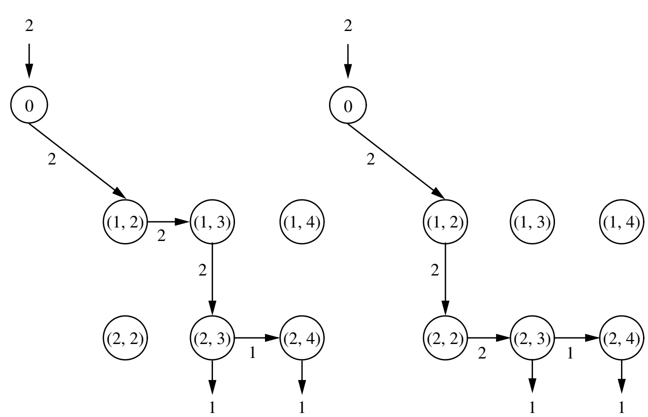
$$p_t(y) = \begin{cases} 100 \cdot 1_{\{y>0\}} + y & \text{if } t \neq 2, \\ 1 \cdot 1_{\{y>0\}} + y & \text{otherwise}, \end{cases}$$

and

$$c^1_t(x) = \begin{cases} 50 \cdot 1_{\{x>0\}} + x & \text{if } t \neq 2, \\ 1 \cdot 1_{\{x>0\}} + x & \text{otherwise}. \end{cases}$$

Finally, let all inventory holding cost functions be equal to zero. The optimal flows in $\phi(1,4,2,4)$ and $\psi(1,4,2,4)$ are given in Figure 3. When calculating the costs $\phi(1,4,2,4)$, transportation is only allowed in periods in which both production may take place and demand is satisfied (i.e., in Periods 3 and 4 in the example), while in the relaxed version of the same subplan transportation is permitted in any period where either production may take place or demand is satisfied (i.e., in Periods 2, 3, and 4 in the example). Therefore, the costs $\psi(1,4,2,4)$ are lower than

**Figure 3 The Optimal Flows in $\phi(1,4,2,4)$ and $\psi(1,4,2,4)$**



$\phi(1,4,2,4)$ because in the relaxed subplan we can transport in the second period.

These changes have two major consequences. Consider a path from the source to a sink in the Phase 2 network. First, while it is easy to see that the corresponding solution of the MLSP-PC is indeed feasible, it is not necessarily an extreme point solution because production and demand nodes in two relaxed subplans contained in the solution may be connected by arcs containing positive flow. Second, it is possible that certain arcs are used in more than one relaxed subplan. This means that the length of the path in the network may not be the same as the costs of the corresponding solution to the MLSP-PC. Dealing with this latter issue first, the following theorem shows that the path length is never smaller than the actual costs of the solution, and is equal to the costs of the solution if all transportation and inventory cost functions are linear.

Theorem 4.1. *Each path from the source to a sink in the Phase* 2 *dynamic programming network corresponds to a feasible solution to the MLSP-PC. The length of this path cannot be smaller than the cost of the corresponding solution, and is equal to the solution cost if all transportation and inventory cost functions are linear.*

Proof. The fact that a path from the source to a sink in the Phase 2 dynamic programming network corresponds to a feasible solution to the lot-sizing problem follows immediately from the fact that all production capacity constraints, as well as all demands, are satisfied. However, certain transportation and inventory arcs may carry positive flow in the partial solutions corresponding to more than one arc in the path, and each of the partial flows is charged separately according to the corresponding cost function. Due to the concavity of all cost functions, it follows that the cost of the total flow will not exceed the sum of the costs of the individual flows on any particular arc, and therefore the length of a path will never be less than the costs of the corresponding solution. In addition, when all transportation and inventory cost functions are linear, the path length and solution costs are clearly equal. □

The next lemma gives a relationship between the costs associated with a subplan and the corresponding relaxed subplan.

Lemma 4.2. *For any subplan* $((s_{\ell 1}, s_{\ell 2})_{\ell=1}^{L})$, *we have that* $\phi((s_{\ell 1}, s_{\ell 2})_{\ell=1}^{L}) \geq \psi(s_{11}, s_{12}, s_{L1}, s_{L2})$.

Proof. This result follows immediately by noting that both $\phi((s_{\ell 1}, s_{\ell 2})_{\ell=1}^{L})$ and $\psi(s_{11}, s_{12}, s_{L1}, s_{L2})$ are the optimal value of an optimization problem with identical cost functions, but where the feasible region of the former is a subset of the feasible region of the latter. □

The next theorem shows that there exists an optimal solution to the lot-sizing problem that is represented by a path in the Phase 2 dynamic programming network whose length is equal to the optimal costs.

Theorem 4.3. *The Phase* 2 *dynamic programming network contains a path that corresponds to an optimal solution to our lot-sizing problem, and the length of the path is equal to the cost of this solution.*

Proof. Consider an extreme point optimal solution to the lot-sizing problem, say with cost $\Phi^*$. As discussed in §2.4, this optimal solution decomposes into a sequence of consecutive subplans. It is easy to see that the Phase 2 dynamic programming network contains a path for which the production and demand periods of each of the arcs correspond to this sequence of subplans. Lemma 4.2 now says that the length of the path in the dynamic programming network, say $\Psi$, will not exceed $\Phi^*$. However, by Theorem 4.1 we know that $\Psi$ is an overestimation of the costs of a corresponding feasible solution. Optimality of $\Phi^*$ now implies that in fact $\Psi = \Phi^*$, which proves the desired result. □

Theorems 4.1 and 4.3 clearly imply that our two-phase algorithm solves the MLSP-PC. We can now conclude that Phase 2 of the algorithm runs in $O(T^4)$ time, given all values $\psi(t_1, t_2, \tau_1, \tau_2)$. The remaining challenge is thus to provide efficient algorithms for computing these values.

### 4.2. Concave Costs

**4.2.1. The Costs of the Relaxed Subplans.** In this section, we will formulate the problem of determining the costs $\psi(t_1, t_2, \tau_1, \tau_2)$ as a dynamic programming problem. Put differently, we define, for each $(t_1, t_2, \tau_1, \tau_2)$, a network with the property that $\psi(t_1, t_2, \tau_1, \tau_2)$ is equal to the shortest path between a pair of source and sink nodes in this network. The nodes in this network are of the form $(t, Y, X^1, \ldots, X^{L-1})$, where $t$ indicates a period, $Y$ is equal to the cumulative production quantity up to and including period $t$, and $X^\ell$ is equal to the cumulative transportation quantity from level $\ell$ to level $\ell+1$ up to and including period $t$. Note that feasibility dictates that we should restrict ourselves to values $d_{\tau_1+1,t} \leq X^{L-1} \leq \cdots \leq X^1 \leq Y$. The source is the node $(t_1, 0, \ldots, 0)$, while the sink is the node $(t_2, Kb+\varepsilon, \ldots, Kb+\varepsilon)$. As in §3.3, we have that

$$Y \in \bigcup_{k=0}^{K} \{kb, kb+\varepsilon\},$$

and the number of allowable values for $Y$ is $O(T)$. Moreover, in a similar fashion to that of the two-level case, we have that

$$X^\ell \in \left(\bigcup_{k=0}^{K} \{kb, kb+\varepsilon\}\right) \cup \left(\bigcup_{s=\tau_1+1}^{\tau_2} \{d_{\tau_1+1,s}\}\right), \quad \ell = 1, \ldots, L-1,$$

so that the number of allowable values for $X^\ell$ is $O(T)$ as well. This means that the total number of nodes in the network is $O(T^{L+1})$.

Arcs in the network represent production, transportation, and inventory decisions, and are present between pairs of nodes in the network of the form $(t, Y, X^1, \ldots, X^{L-1})$ and $(t+1, \overline{Y}, \overline{X}^1, \ldots, \overline{X}^{L-1})$, where $\overline{Y} - Y \in \{0, \varepsilon, b\}$ and $\overline{X}^\ell \geq X^\ell$ $(\ell = 1, \ldots, L-1)$. It is easy to see that there are $O(T^{L-1})$ arcs emanating from each node, for a total of $O(T^{2L})$ arcs in the network.

Similar to §3.3, we can easily compute the production quantity in period $t+1$ $(\overline{Y} - Y)$, the transportation quantity between levels $\ell$ and $\ell+1$ in period $t+1$ $(\overline{X}^\ell - X^\ell)$, the inventory held at the manufacturer level at the end of period $t+1$ $(\overline{Y} - \overline{X}^1)$, and the inventory held at the retailer level at the end of period $t+1$ $(\overline{X}^{L-1} - d_{\tau_1+1, t+1})$. The costs of an arc are thus given by

$$p_{t+1}(\overline{Y} - Y) + \sum_{\ell=1}^{L-1} c_{t+1}^\ell(\overline{X}^\ell - X^\ell) + h_{t+1}^1(\overline{Y} - \overline{X}) + \sum_{\ell=2}^{L-1} h_{t+1}^\ell(\overline{X}^{\ell-1} - \overline{X}^\ell) + h_{t+1}^L(\overline{X}^{L-1} - d_{\tau_1+1, t+1}).$$

If all cost functions can be evaluated in constant time, the costs of a given arc can be computed in $O(L)$ time in the same way as in the 2LSP-PC after a preprocessing step taking $O(T^2)$ time. We conclude that the cost of a single relaxed subplan can be determined in $O(LT^{2L})$ time.

Finally, noting that there are $O(T^4)$ relaxed subplans and applying the same technique for reducing the running time as used at the end of §3.3, we obtain an algorithm for the MLSP-PC with arbitrary concave production, transportation, and inventory holding costs and stationary capacities that runs in $O(LT^{2L+3})$ time. Although this time is exponential in the number of levels, the order of the running time will be limited by the fact that the number of levels will typically be relatively small.

This approach can easily be extended to deal with initial inventories. Recall from §3.4 that only relaxed subplans with $t_1 = 0$ need to be considered. For such a relaxed subplan, we should view the total initial inventories $\sum_{\ell=1}^L I_0^\ell$ as a cumulative production quantity up to and including Period 0, and the initial inventory $\sum_{\ell=s+1}^L I_0^\ell$ as the cumulative transportation quantity up to and including Period 0 from level $s$ to level $s+1$, for all $s \in \{1, \ldots, L-1\}$. As in §3.4, without increasing the running time, these initial inventories can be incorporated into the dynamic programming approach to calculating $\psi(t_1, t_2, \tau_1, \tau_2)$ by appropriate redefinitions of the possible values of $Y$ and $X$.

In the next sections, we will show how the running time can be dramatically reduced for problem instances that have stationary production capacities, general concave production costs, and linear inventory holding costs at all levels, as well as one of the following two transportation cost structures: (i) fixed charge without speculative motives; or (ii) linear.

### 4.3. Fixed-Charge Transportation Costs Without Speculative Motives

**4.3.1. Introduction.** In this section, we consider the case of fixed-charge transportation costs without speculative motives and linear inventory holding costs. As before, we will determine the costs of each relaxed subplan using dynamic programming. After a preprocessing step that runs in $O(LT^4)$, this dynamic program runs in $O(T^4)$ time for each individual relaxed subplan. By using the reduction technique at the end of §3.3, the cost of all $O(T^4)$ relaxed subplans can be computed simultaneously in $O(T^7)$ time. Therefore, the running time of the dynamic programming approach for this special case of the MLSP-PC is $O(T^7 + LT^4)$. When $L = 2$, we can reduce this running time to $O(T^6)$.

**4.3.2. Zero-Inventory-Ordering Property at the Retailer.** We will show that, under fixed-charge transportation costs without speculative motives, solutions satisfying the zero-inventory-ordering (ZIO) property at all levels in $\{2, \ldots, L\}$, i.e., $I_t^\ell x_{t+1}^{\ell-1} = 0$ for $t = 1, \ldots, T-1$; $\ell = 2, \ldots, L$, are dominant. That is, given any feasible solution to the relaxed subplan $(t_1, t_2, \tau_1, \tau_2)$, there always exists another solution that is at least as good and satisfies the ZIO property at all levels in $\{2, \ldots, L\}$.

Theorem 4.4. *Given a relaxed subplan* $(t_1, t_2, \tau_1, \tau_2)$, *the set of solutions with the ZIO property at all levels in* $\{2, \ldots, L\}$ *is dominant.*

Proof. Let $(\bar{y}, \bar{x}, \bar{I})$ be a feasible solution to the relaxed subplan $(t_1, t_2, \tau_1, \tau_2)$ that does not satisfy the ZIO property at some level. Let $\bar{\ell}$ be the last level, such that the ZIO property holds for all $\ell \in \{\bar{\ell}+1, \ldots, L\}$, but is not true for level $\bar{\ell}$. We can construct a new solution at least as good as $(\bar{y}, \bar{x}, \bar{I})$, such that the ZIO property holds for all $\ell \in \{\bar{\ell}, \ldots, L\}$. If $\bar{\ell} = 2$, then we have obtained the desired result. Otherwise, we repeat the procedure with the new solution. Observe that this procedure converges because the new $\bar{\ell}$ has decreased by at least one unit.

Let $\bar{t} \in \{t_1+1, \ldots, \tau_2-1\}$ be a period so that $\bar{I}_{\bar{t}}^{\bar{\ell}} \bar{x}_{\bar{t}+1}^{\bar{\ell}-1} > 0$. The positive inventory $\bar{I}_{\bar{t}}^{\bar{\ell}}$ has been transported to level $\bar{\ell}$ in some earlier period. However, due to the absence of speculative motives, we can reschedule the transportation of the $\bar{I}_{\bar{t}}^{\bar{\ell}}$ units to period $\bar{t}+1$ without increasing the costs. Repeating this argument for each period $\bar{t}$ violating the ZIO property at level $\bar{\ell}$, we obtain a solution where the ZIO property is true for each level $\ell \in \{\bar{\ell}, \ldots, L\}$. □

We may recall that $\psi(t_1, t_2, \tau_1, \tau_2)$ is equal to the minimal costs among the solutions of the relaxed subplan $(t_1, t_2, \tau_1, \tau_2)$ with at most one free production arc. The following corollary to Theorem 4.4 states that for finding this constrained minimum we can again restrict our search to solutions satisfying the ZIO property at the retailer.

Corollary 4.5. *The cost associated with the relaxed subplan $(t_1, t_2, \tau_1, \tau_2)$ can be found among all feasible solutions satisfying the ZIO property at all levels in $\{2, \ldots, L\}$.*

Proof. This follows immediately from the proof of Theorem 4.4 by observing that the modification of the solution to obtain a solution satisfying the ZIO property does not alter the production flows. □

This corollary implies that when searching for $\psi(t_1, t_2, \tau_1, \tau_2)$, we can assume that any amount shipped is equal to the demand of a set of consecutive periods. This will help to reduce the information maintained in the dynamic programming approach described in §4.2.1.

**4.3.3. The Costs of a Relaxed Subplan.** In this section, we will formulate the problem of determining the costs $\psi(t_1, t_2, \tau_1, \tau_2)$ as a simplification of the dynamic programming problem defined in §4.2.1. All nodes in the dynamic programming network are of the form $(t, Y, s)$, where $t$ indicates a period, $Y$ is equal to the cumulative production quantity up to and including period $t$, and $s$ represents the last period whose demand is satisfied using transportation *from Level 1 to Level 2* up to and including period $t$, where $d_{\tau_1+1,s} \leq Y$ and $t \leq s$. We may observe that from Theorem 4.4, we have that $X^1 = d_{\tau_1+1,s}$ in the dynamic program of §4.2.1. The source is the node $(t_1, 0, \tau_1)$, while the sink is the node $(t_2, Kb+\varepsilon, \tau_2)$. As before, we know that $Y$ can only assume the values

$$Y \in \bigcup_{k=0}^{K} \{kb, kb+\varepsilon\}.$$

Arcs are present between pairs of nodes in the network of the form $(t, Y, s)$ and $(t, \overline{Y}, \bar{s})$, where $\overline{Y} - Y \in \{0, \varepsilon, b\}$ and $\bar{s} \geq s$. It is easy to see that there are $O(T)$ arcs emanating from each node in the network, so that the entire network has $O(T^4)$ arcs.

The costs of an arc between nodes $(t, Y, s)$ and $(t+1, \overline{Y}, \bar{s})$ are now given by

$$p_{t+1}(\overline{Y} - Y) + h^1_{t+1}(\overline{Y} - d_{\tau_1+1,\bar{s}}) + c^1_{t+1}(d_{s+1,\bar{s}}) + C_{t+1,2}(s+1, \bar{s}),$$

where $C_{t\ell}(s_1, s_2)$ are defined as the optimal costs of shipping $d_{s_1 s_2}$ units from node $(t, \ell)$ to their destinations, i.e., demand nodes $(L, s_1), \ldots, (L, s_2)$. We can use Zangwill's algorithm, in a preprocessing stage, to determine the values $C_{t2}(s_1, s_2)$ for all $t = 2, \ldots, T$; $s_1 = t, \ldots, T$; and $s_2 = s_1, \ldots, T$ in $O(LT^4)$ time; see the online appendix. It is important to note that although Zangwill's model allows for general concave transportation and inventory holding cost functions, we cannot use the same approach as described above in the presence of production capacities. The reason is that in the uncapacitated case, the ZIO property holds for arbitrary concave arc cost functions, while this is not the case in the capacitated case. However, as we have shown, in the case of fixed-charge transportation costs that exhibit no speculative motives, we also obtain the ZIO property, enabling the use of Zangwill's algorithm to determine inputs to our algorithm.

The problem of determining $\psi(t_1, t_2, \tau_1, \tau_2)$ reduces to finding the length of the shortest path in the network from the source to the sink, which can be done in linear time in the number of arcs. It is easy to see that the number of nodes in the network is $O(T^3)$ and the number of arcs $O(T^4)$. Using the same approach to computing multiple values of the function $\psi$ at once as we have discussed for the function $\psi$ at the end of §3.3, this yields an $O(T^7 + LT^4)$ algorithm for solving the multilevel variant of this problem.

When $L = 2$, this running time can be reduced to $O(T^6)$. Recall that the number of nodes in the dynamic programming approach above is $O(T^3)$. We will show that the number of arcs is also $O(T^3)$. For each $t$ there are $O(T)$ nodes of the form $(t, \cdot, t)$, and $O(T^2)$ nodes of the form $(t, \cdot, s)$ with $s > t$. Each node of the form $(t, \cdot, t)$ has $O(T)$ successors, and each node of the form $(t, \cdot, s)$ with $s > t$ has $O(1)$ successors, which makes for a total of $O(T) \cdot (O(T) \cdot O(T) + O(T^2) \cdot O(1)) = O(T^3)$ arcs in the network. This yields an $O(T^6)$ algorithm for solving the two-level variant of this problem.

Unfortunately, in the presence of nonzero initial inventory levels the ZIO property is not necessarily dominant anymore. However, in these cases the more general procedure developed for the case of arbitrary concave cost functions of course still applies.

### 4.4. Linear Transportation Costs

**4.4.1. Introduction.** In this section, we will consider the case where the transportation costs and inventory holding costs are linear. We will develop a dynamic programming approach that finds the optimal costs of each relaxed subplan. After a preprocessing step that runs in $O(LT^2)$ time, this algorithm runs in $O(T^2)$ time for a single relaxed subplan, but the costs of all $O(T^4)$ relaxed subplans can be computed simultaneously in $O(T^5)$ time. This results in an $O(T^5 + LT^2)$ algorithm for solving this class of instances of MLSP-PC.

**4.4.2. Preprocessing.** In terms of the underlying network (as described in §2.1), one unit produced in period $t$ for satisfying demand in period $\tau \geq t$ will, in the optimal solution, flow along the minimum-cost path from $(1, t)$ to $(L, \tau)$. In a preprocessing stage, we can determine the minimal unit transportation costs associated with producing one unit in period $t$ for consumption in period $\tau$, which we will call $G_{t\tau}$. All these values can be computed in $O(LT^2)$ time by solving $T$ shortest path problems in acyclic graphs with $O(TL)$ arcs using backward recursion. Using these values, we can then again determine the total transportation costs associated with producing, in period $t$, the entire demand of the consecutive periods $\tau_1 + 1, \ldots, \tau_2$, assuming that transportation is allowed in all periods $t, \ldots, \tau_2$, i.e., $GD_{t\tau_1\tau_2} \equiv \sum_{r=\tau_1+1}^{\tau_2} d_r G_{tr}$. In $O(T^3)$ time, these costs can be calculated for all $t = 1, \ldots, T$ and $t \leq \tau_1 + 1 \leq \tau_2 \leq T$. This information will enable us to compute the total transportation costs associated with production in period $t$ in constant time.

**4.4.3. The Costs of a Relaxed Subplan.** In this section, we will formulate the problem of determining the costs $\psi(t_1, t_2, \tau_1, \tau_2)$ as a further simplification of the dynamic programming problem defined in §4.2.1. All nodes in the dynamic programming network are of the form $(t, Y)$, where $t$ indicates a period, and $Y$ is equal to the cumulative production quantity up to and including period $t$, where $d_{\tau_1+1,t} \leq Y$ and

$$Y \in \bigcup_{k=0}^{K} \{kb, kb + \varepsilon\}.$$

The source is the node $(t_1, 0)$, while the sink is the node $(t_2, Kb + \varepsilon)$.

Arcs are present between pairs of nodes in the network of the form $(t, Y)$ and $(t+1, \overline{Y})$ when $\overline{Y} - Y \in \{0, \varepsilon, b\}$. Each arc of the network described above represents a possible production decision. We let the costs of the arcs be equal to the total costs associated with the production amount. It remains to show that the transportation and inventory holding costs can be computed in constant time. In addition to the information gathered in the preprocessing phase described in §4.4.2, we also will find, for each node $(t, Y)$ in the network, the first period whose demand is not fully satisfied by the cumulative production $Y$ (say $s$) as well as the part of the demand of that period that remains to be satisfied (say $\delta$). Using the cumulative demands $d_{\tau_1+1,t'}$ $(t' = \tau_1 + 1, \ldots, \tau_2)$ as well as the fact that the value of $Y$ can only be equal to $kb$ or $kb + \varepsilon$ for $k = 0, \ldots, K$, this additional information can be obtained in $O(T)$ time. As we will see later, this does not increase the running time of finding the costs of a single relaxed subplan.

Now consider an arc connecting the two nodes $(t, Y)$ (with first remaining demand period $s$ with remaining demand $\delta$) and $(t+1, \overline{Y})$ (with first remaining demand period $\bar{s}$ with remaining demand $\bar{\delta}$). When $\overline{Y} - Y \leq \delta$, the unit transportation costs of the quantity produced in period $t+1$ are equal to $G_{t+1,s}$. When $\overline{Y} - Y > \delta$, the transportation and inventory holding costs for this arc consist of up to three components: $G_{t+1,s}\delta + GD_{t+1,s,\bar{s}-1} + G_{t+1,\bar{s}}(d_{\bar{s}} - \bar{\delta})$, and can thus indeed be computed in constant time.

The problem of determining $\psi(t_1, t_2, \tau_1, \tau_2)$ reduces to finding the length of the shortest path in the network from the source to the sink, which can be done in linear time in the number of arcs. It is easy to see that the number of nodes in the network is $O(T^2)$, and the number of arcs $O(T^2)$. Using the same approach to computing multiple values of the function $\psi$ at once as in §4.3, this yields an $O(T^5 + LT^2)$ algorithm for solving this variant of the MLSP-PC.

As in the CLSP, initial inventories can be incorporated when all transportation and inventory holding cost functions are linear. In particular, the initial inventories are used to satisfy the earliest demands via the appropriate shortest paths in the network, after which the demands are updated and the remaining problem without initial inventories is solved.

## 5. Concluding Remarks and Future Research

In this paper, we have considered a generalization of the classical ELSP with stationary production capacities that allows for multiple levels of storage, as well as corresponding transportation decisions for transporting between the different levels. We have identified two important special cases of this problem that are solvable in polynomial time. The running times of the corresponding algorithms are remarkably insensitive to the number of levels in the supply chain.

Open issues for future research in this area can be divided into three general directions. First, the complexities, although polynomial in the planning horizon, are of relatively high order: $O(T^5)$ to $O(T^7)$ for the two-level cases. It would be interesting if the order of the running time could be reduced, for instance, by investigating whether more time can be saved by determining the costs of many or all subplans simultaneously. In addition, although the number of levels will generally be relatively small, it would nevertheless be interesting to determine if the multilevel case with general concave cost functions can be solved in polynomial time in both the time horizon and the number of levels. A second direction is the study of serial supply chains in the presence of capacities at other or additional levels in the chain. Finally, it would be interesting to consider more complex supply chain structures, including, for example, product assembly structures at the producer level, or multiple retailers.

An online appendix to this paper is available at http://mansci.pubs.informs.org/ecompanion.html.

## Acknowledgments

The work of the second and third authors was supported by the National Science Foundation under Grants No. DMI-0085682 and DMI-0355533, Maastricht Research School of Economics of Technology and Organizations (METEOR), and the Netherlands Organization for Scientific Research (NWO).

## References

Aggarwal, A., J. K. Park. 1993. Improved algorithms for economic lot-size problems. *Oper. Res.* **41**(3) 549–571.

Ahuja, R. K., T. L. Magnanti, J. B. Orlin. 1993. *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, Englewood Cliffs, NJ.

Arntzen, B. C., G. G. Brown, T. P. Harrison, L. L. Trafton. 1995. Global supply chain management at Digital Equipment Corporation. *Interfaces* **25**(1) 69–93.

Baker, K. R., P. Dixon, M. J. Magazine, E. A. Silver. 1978. An algorithm for the dynamic lot-size problem with time-varying production capacity constraints. *Management Sci.* **24**(16) 1710–1720.

Bitran, G. R., H. H. Yanasse. 1982. Computational complexity of the capacitated lot size problem. *Management Sci.* **28**(10) 1174–1186.

Chandra, P., M. L. Fisher. 1994. Coordination of production and distribution planning. *Eur. J. Oper. Res.* **72** 503–517.

Chung, C.-S., C.-H. M. Lin. 1988. An $O(T^2)$ algorithm for the $NI/G/NI/ND$ capacitated lot size problem. *Management Sci.* **34**(3) 420–426.

Federgruen, A., M. Tzur. 1991. A simple forward algorithm to solve general dynamic lot sizing models with $n$ periods in $O(n \log n)$ or $O(n)$. *Management Sci.* **37** 909–925.

Florian, M., M. Klein. 1971. Deterministic production planning with concave costs and capacity constraints. *Management Sci.* **18** 12–20.

Florian, M., J. K. Lenstra, A. H. G. Rinnooy Kan. 1980. Deterministic production planning: Algorithms and complexity. *Management Sci.* **26**(7) 669–679.

Geoffrion, A. M., R. F. Powers. 1995. Twenty years of strategic distribution system design: An evolutionary perspective. *Interfaces* **25**(5) 105–127.

Kaminsky, P., D. Simchi-Levi. 2003. Production and distribution lot sizing in a two stage supply chain. *IIE Trans.* **35**(11) 1065–1075.

Lee, C. Y., S. Çetinkaya, W. Jaruphongsa. 2003. A dynamic model for inventory lot sizing and outbound shipment scheduling at a third-party warehouse. *Oper. Res.* **51**(5) 735–747.

Thomas, D. J., P. M. Griffin. 1996. Coordinated supply chain management. *Eur. J. Oper. Res.* **94**(1) 1–15.

van den Heuvel, W., A. P. M. Wagelmans. 2003. A geometric algorithm to solve the $NI/G/NI/ND$ capacitated lot-sizing problem in $O(T^2)$ time. Technical report 2003-24, Econometric Institute Report, Rotterdam, The Netherlands.

van Hoesel, C. P. M., A. P. M. Wagelmans. 1996. An $O(T^3)$ algorithm for the economic lot-sizing problem with constant capacities. *Management Sci.* **42**(1) 142–150.

Wagelmans, A., S. van Hoesel, A. Kolen. 1992. Economic lot sizing: An $O(n \log n)$ algorithm that runs in linear time in the Wagner-Whitin case. *Oper. Res.* **40**(1) S145–S156.

Wagner, H. M. 1960. A postscript to dynamic problems of the theory of the firm. *Naval Res. Logist. Quart.* **7** 7–12.

Wagner, H. M., T. M. Whitin. 1958. Dynamic version of the economic lot size model. *Management Sci.* **5** 89–96.

Zangwill, W. I. 1969. A backlogging model and a multi-echelon model of a dynamic economic lot size production system—A network approach. *Management Sci.* **15**(9) 506–527.