Mathematical Programming Models and Formulations for Deterministic Production Planning Problems*

Yves Pochet

CORE and IAG, Université Catholique de Louvain, Belgium
ypochet@core.ucl.ac.be

Abstract. We study in this lecture the literature on mixed integer programming models and formulations for a specific problem class, namely deterministic production planning problems. The objective is to present the classical optimization approaches used, and the known models, for dealing with such management problems.

We describe first production planning models in the general context of manufacturing planning and control systems, and explain in which sense most optimization solution approaches are based on the decomposition of the problem into single-item subproblems.

Then we study in detail the reformulations for the core or simplest subproblem in production planning, the single-item uncapacitated lot-sizing problem, and some of its variants. Such reformulations are either obtained by adding variables – to obtain so called extended reformulations – or by adding constraints to the initial formulation. This typically allows one to obtain a linear description of the convex hull of the feasible solutions of the subproblem. Such tight reformulations for the subproblems play an important role in solving the original planning problem to optimality.

We then review two important classes of extensions for the production planning models, capacitated models and multi-stage or multi-level models. For each, we describe the classical modeling approaches used.

Finally, we conclude by giving our personal view on some new directions to be investigated in modeling production planning problems. These include better models for capacity utilization and setup times, new models to represent the product structure – or recipes – in process industries, and the study of continuous time planning and scheduling models as opposed to the discrete time models studied in this review.

1 Introduction

Two chapters in this book present the general theory of mixed integer programming [⟶ Martin], and the lift-and-project approach to combinatorial op-

* Research carried out with financial support of the project TMR-DONET nr. ERB FMRX-CT98-0202 of the European Union, and within the framework of the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming.

M. Jünger and D. Naddef (Eds.): Computat. Comb. Optimization, LNCS 2241, pp. 57–111, 2001.

timization [⟶ Balas]. They propose and study methods for improving the formulations and solving general mixed integer programs, without considering or exploiting the specific nature of the problems. In this chapter we take a different viewpoint. We study the literature on mixed integer programming models and formulations for a specific problem class, namely deterministic production planning problems. We illustrate how the specific structure of these problems can be exploited to obtain good or better formulations and algorithms for solving them. Implementation issues are not considered in this chapter, we refer the reader to two chapters [⟶ Elf/Gutwenger/Jünger/Rinaldi], and [⟶ Ladányi/Ralphs/Trotter].

In manufacturing environments, production planning problems deal with decisions about the size of the production lots of the different products to manufacture or to process, about the time at which those lots have to be produced, and sometimes about the machine or production facility where the production must take place. The production decisions are typically taken by looking at the best trade-off between financial and customer service or satisfaction objectives. In production planning and operations management, the financial objectives are usually represented by production costs – for machines, materials, manpower, startup costs, overhead costs, ...– and inventory costs – opportunity costs of the capital tied up in the stocks, insurances, ...–. Customer service objectives are represented by the ability to deliver the right product, in ordered quantity, at the promised date and place. The final aim of such modelling approaches is to provide tools allowing one to better plan and control the flow of materials and information within the firms.

These production planning problems abound in practice and the literature contains many heuristic and exact optimization algorithmic approaches to solve them. Our objective here is to present a classification of the mixed integer programming (MIP) optimization models, and mathematical formulations, used for dealing with such management problems.

In Section 2, we describe production planning models in the general context of manufacturing planning and control systems, including Material requirements planning (MRP-I), Manufacturing resource planning (MRP-II) and Hierarchical production planning (HPP). We describe briefly the elements of production planning models, and define a typology of the models encountered in practice. This part is a self contained introduction to production planning models, and to make the presentation more practical at this early stage, we describe example planning models such as the uncapacitated lot-sizing model, the multi-item master production scheduling (MPS) model and the materials requirements planning basic model.

Although practical models are multi-item, we explain in which sense most optimization solution approaches are based on the decomposition of the problem into single-item subproblems. This motivates the analysis of the formulations of a variety of single-item models in the next section. The goal is to obtain a complete linear description of the convex hull of the feasible solutions of the subproblem. Such tight formulations play an important role in solving or finding good solutions for the original multi-item planning problem.

In Section 3, we present known results on the MIP formulation of the basic or core planning subproblem, namely the single-item uncapacitated lot-sizing subproblem (ULS), and some of its single-item variants including startup costs, backlogging, constant capacity restriction, Wagner-Within costs and profit maximization objective function. We illustrate on the ULS problem the classical approach used to study the reformulations of these polynomially solvable single item problems. The reformulations are either obtained by adding variables – to obtain so called extended reformulations – or by adding constraints to the initial formulation.

For the ULS problem, several extended formulations are described, including the shortest path formulation that can be derived directly from the polynomial time dynamic program solving the problem. Other classical extended reformulations such as multicommodity and facility location reformulations are presented.

The alternative approach is to work with the initial set of variables, and derive valid inequalities that describe the convex hull of the incidence vectors of feasible solutions. By solving the separation problem associated with such inequalities, it is possible to develop a branch and cut algorithm for solving these problems. For each model studied in this section, we describe the known valid inequalities, indicate whether or not they describe the convex hull of solutions.

In Sections 4 and 5, we review two important classes of extensions for the production planning models, capacitated models and multi-stage or multi-level models.

We separate capacitated planning models into big buckets (large time periods) and small buckets (small time periods) models. Big buckets capacitated models are used in order to control the capacity utilization in each period. We describe how to obtain classes of valid inequalities to strengthen the formulation of these models. Small buckets models are used when, in order to model accurately the capacity utilization, one has to control the production sequence of the different items – because there are variable sequence dependent change over costs and times –. We describe several formulations that are used to represent these production sequences using small buckets.

For multi-stage models, we explain the classical stage by stage decomposition used in MRP-I, and describe the well known echelon stock reformulation which is the basis of all MIP approaches for these problems.

In Section 6, we give our personal view on some new directions to be investigated in modeling production planning problems. These include better models for capacity utilization and setup times, new models to represent the product structure – or recipes – in process industries, and the study of continuous time planning and scheduling models as opposed to the discrete time models studied in this review.

We conclude with some challenges for the future of this field of research.

2 Production Planning

2.1 Manufacturing Planning and Control Systems

The production process can be defined as the transformation process of raw materials into end products, usually through a series of transformation steps producing and consuming intermediate products. These raw materials, intermediate and end products can often be inventoried, allowing one to produce and consume them at different moments and rates in time. Each transformation step may require several input products and may produce one or several outputs. The raw materials are purchased from suppliers, and the end products are sold to external customers. Sometimes, intermediate products are also sold to customers (spare parts, . . .). This general definition of the production as a transformation process is illustrated in Figure 1, where materials inventories are represented by triangles, transformation processes are represented by circles, and flows of materials through the process (i.e. inputs into or outputs from transformations steps) are represented by arrows.

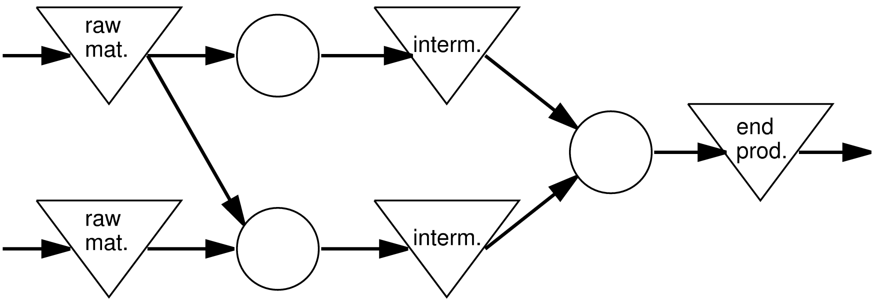

Fig. 1. The production process

Production planning is defined as the planning (that is the acquisition, time of usage, quantity used, . . .) of the resources required to perform these transformation steps, in order to satisfy the customers in the most efficient or economical way. In other words, the production decisions are typically taken by looking at the best trade-off between financial objectives and customer service or satisfaction objectives. In production planning and operations management, the financial objectives are usually represented by production costs – for machines, materials, manpower, startup costs, overhead costs, . . . – and inventory costs – opportunity costs of the capital tied up in the stocks, insurances, . . . –. Customer service objectives are represented by the ability to deliver the right product, in ordered quantity, at the promised date and place.

According to Anthony [5] and Salomon [62] –among others–, the production planning problems can be classified into strategic, tactical and operational planning problems.

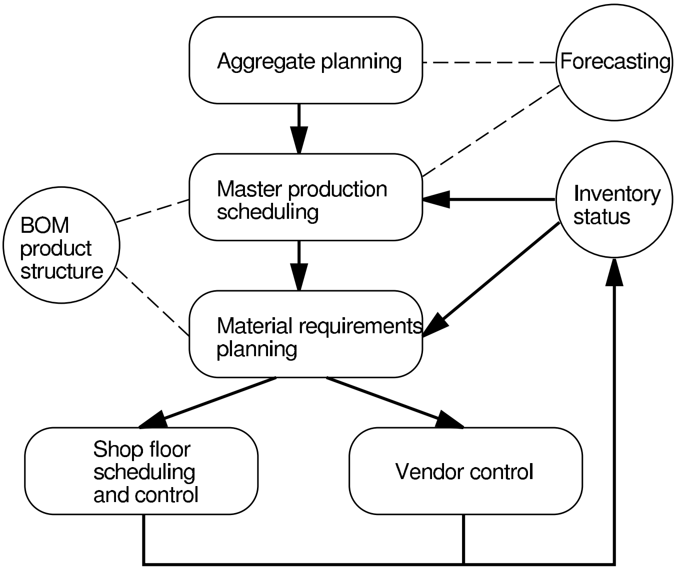

Fig. 2. An MRP-II system (adapted from [76])

Strategic problems deal with the management of change in the production process and the acquisition of the resources needed to produce. This includes for example product-mix, plan layout as well as location, supply chain design and investment decisions. The objective pursued in solving these strategic problems is to maintain the competitive advantage and capabilities, and to sustain the growth rate. This requires to model the decisions over a long term horizon using aggregate information.

Tactical planning problems analyse the resource utilization problems over a medium term planning horizon using aggregate information. This consists in making decisions about, for instance, materials flow, inventory, capacity utilization, maintenance planning. The usual objective at this stage is to improve the cost efficiency and the customer's satisfaction.

Operational Planning problems aim at planning and controling the execution of the production tasks. For instance, production sequencing and input/output analysis models fit into this category. Here the goal is to obtain an efficient and accurate execution of the plans, over a very short term horizon, and using very detailed information.

Manufacturing planning and control systems (MPC) (see Vollman, Berry, Whybark [76]) are developed to cope with these complex planning environments, and integrate these multi-level multi-horizon planning problems into a single integrated management system.

For instance, Figure 2 describes how the tactical and operational planning problems are integrated in classical manufacturing resources planning (MRP-II) systems, an example of MPC system. In these systems, medium term production planning (PP) consists in deciding about capacity utilization and aggregate or global inventory levels to meet forecasted demand over a medium term horizon of about one year. A medium term horizon is needed to be able to take into account some seasonal pattern in the demand. Master production scheduling (MPS) consists in planning the detailed short term production of end-products in order to meet forecasted demand and firm customer orders, taking into account the capacity utilization and global inventory levels decided at the PP stage. Here the time horizon is usually expressed in weeks and corresponds to the duration of the production cycle. Materials requirements planning (MRP-I) establishes the short term production plans for all components (intermediate products and raw materials) from the production plan of end-products decided at the MPS stage, and from the product structure database (Bills of Materials, BOM). Then, shop-floor control systems (for manufactured components) and vendor follow-up systems (for purchased components) control the very short term execution of the plans decided at the MRP-I stage. The time horizon at this last stage is usually of a few days.

Other well known integrated production planning concepts and systems fit into this general manufacturing planning and control framework. For instance the MRP-II system represented in Figure 2 subsumes the original MRP-I system defined by Orlicky [49], and follows the hierarchical production planning (HPP) principles defined by Hax and Meal [29].

In these systems, the production decisions are taken, one after the other, using a sequential (decomposition) approach. For instance, in the classical MRP-II systems, the production plans of the end-items are decided first at the MPS level, without considering the capacity or inventory available for intermediate products and raw materials. Then a tentative production plan for intermediate items is decided assuming that the production capacity is infinite (or at least does not impose any constraint on the production plans). Finally, the available capacity is taken into account, and the production plans are modified (smoothed) by finite loading heuristics.

This sequential approach is suboptimal because it is not capable to integrate some important constraints in a global model. The complicating factors and constraints that would need to be integrated are: multilevel product structures (raw materials, intermediate products and end products), capacity constraints (and in particular the presence of setup times), sequencing aspects that have to be taken into account because they affect the capacity utilization.

The purpose of this lecture is precisely to describe and study more global production planning models addressing these complicating factors. The final aim of such modelling approaches is to provide tools allowing one to better plan and control the flow of materials and information within the firms.

2.2 Production Planning Models

We describe in this Subsection the main modelling elements defining production planning models, and we provide examples of mixed integer programming formulations corresponding to the MRP-I, MPS and integrated MPS/MRP decision problems.

Modelling Elements. There are a number of modelling elements present in many or most production planning problems.

- Production planning problems deal with sizing and timing decisions for production lots,
- defining the availability of resources (machine hours, workforce, subcontracting, ...),
- allocating the resources to the production lots,
- meeting forecasted demand (in a make to stosk environment) and/or customer orders (in a make to order environment)
- and maximizing performance, expressed in terms of production costs, inventory costs, and customer service level.
- over a finite planning horizon.

There are also some complicating modelling elements that are not present in all models.

- Multiple items interacting through shared resources.
  In this case, there is no material flow between the items, but there are capacity restrictions linking the items and coming from the shared resources. A typical example of such a model is the MPS model described in Subsection 2.2.
- Multiple items interacting through multi-level product structures.
  In the case of multi-level product structures, there are additional material flow restrictions because a product can be an input of some production stage and also an output of some other production stage, or it may be delivered from an external supplier. This creates some precedence constraints between the supply and the consumption of that product. These restrictions are usually modelled through inventory balance constraints. A typical example of such a model is the MPS/MRP intergated model described in Subsection 2.2.
- Demand backlogging.
  In this case, it is possible – but penalized because it has a negative impact on customer satisfaction – to deliver a customer later than required. This occurs for example when a factory does not have enough capacity to deliver all customers on time. The single item model with demand backlogging will be studied in Section 3.5.

– Startup or switching capacity utilization.
  In many cases, it is necessary to model accurately the capacity utilization in order to obtain feasible production plans. This requires sometimes to model the capacity consumed when a machine starts a production batch, or when a machine switches from one product to another. In these cases, we obtain setup or startup times models, change over times models, or models with sequencing restrictions. These models will be described in Section 4.
  In some other cases, the model are too complex to be solved with such setup or startup times restrictions, and usually simpler models involving setup or startup costs are considered. Such models can be seen as obtained by relaxing the setup or startup time restrictions. Models with setup and startup costs will be studied respectively in Sections 3 and 3.5.

Models Classification. The classification of deterministic production planning models given in Table 1 is adapted from Kuik et al [35]. In each model class, we cite seminal or important work related to the optimization of some corresponding models. Some more references will be given in the text. We refer to the above paper for a more comprehensive classification of the related literature.

The models are classified along three criteria: capacitated or uncapacitated, constant or variable demand, single or multiple item.

Uncapacitated Lot-Sizing Model. The first example model is the uncapacitated single item, single level, uncapacitated lot-sizing problem that we will study in detail in Section 3. This model is the core subproblem in production planning because it is the problem solved repeatedly for each item (from end products to raw materials) in the material requirements (MRP-I) sequential planning system.

We define the index $t = 1, \cdots, n$ to represent the discrete time periods, and n is the final period at the end of the planning horizon. The purpose is to plan the production over the planning horizon (i.e. fix the lot size in each period) in order to satisfy demand, and to minimize the sum of production and inventory costs. Classically, the production costs exhibit some economies of scale that are are modelled through a fixed charge cost function. That is, the production cost of a lot in decomposed into a fixed cost independent of the lot size, and a unit cost incurred for each unit produced in the lot. The inventory costs are modelled by charging an inventory cost per unit held in inventory at the end of each period. Any demand in a period can be satisfied by production or inventory, and backlogging is not allowed. The production capacity in each period is not considered in the model, and therefore assumed to be infinite.

For each period $t = 1, \cdots, n$, the decision variables are x_t , y_t and s_t . They represent respectively the production lot size in period t, the binary variable indicating whether or not there is a positive production in period t ($y_t = 1$ if

Table 1. Production planning models

		Uncapacitated (fixed lead times)	Capacitated (variable lead times)
Stationary or Constant demand	Single Item	Inventory control EOQ (raw material) EPQ (production) [Harris [28], Wilson [80]]	
Stationary or Constant demand	Multi Item	Serial lot sizing Assembly lot sizing (multi-level) [Crowston, W., W. [15]]	Economic lot scheduling (single-level) (shared resources) [Elmaghraby [18]]
Dynamic demand	Single Item	Uncapacitated lot sizing (ULS) [Wagner, Whitin [78]]	Capacitated lot sizing (CLS) [Florian, Klein [23]]
Dynamic demand	Multi Item	Serial lot sizing Assembly lot sizing General lot sizing (multi-level) [Veinott [75]] [Zangwill [85]] [Love [41]] [Crowston, W., [14]] [Afentakis, G., K. [1]] [Afentakis, G., [2]]	Big buckets Small buckets Discrete LS (single-level) (shared resources) [Eppen, Martin, [19]] [Trigeiro, [68] [Karmarkar, Schrage, [31] [Lasdon, Terjung, [36]] [Fleischmann, [21], [22]] [Salomon, K., K., V.W., [63]]

$x_t > 0$), and the inventory at the end of period t. The data are p_t, f_t, h_t and d_t modelling respectively, and for each period t, the unit production cost, the fixed production cost, the unit inventory cost, and the demand to be satisfied. For simplicity we suppose that $d_t \geq 0$ for all periods t.

The natural formulation of this uncapacitated lot-sizing problem (ULS) can be written as follows.

$$\min \sum_{t=1}^{n} (p_t x_t + f_t y_t + h_t s_t) \tag{1}$$

$$s_{t-1} + x_t = d_t + s_t \quad \text{for all } t \tag{2}$$

$$s_0 = s_n = 0 \tag{3}$$

$$x_t \leq M y_t \quad \text{for all } t \tag{4}$$

$$x_t, s_t \geq 0,\ y_t \in \{0,1\} \quad \text{for all } t \tag{5}$$

where M is a large positive number. Constraint (2) expresses the demand satisfaction in each period, and is called the flow balance or flow conservation constraint. This is because every feasible solution of ULS corresponds to a flow in the network shown in Figure 3, where $d_{14} = \sum_{i=1}^{4} d_i$ is the total demand. Constraint (3) says there is no initial and no final inventory. Constraint (4) forces the setup variable in period t to be 1 when there is positive production (i.e. $x_t > 0$) in period t. Constraint (5) imposes the nonnegativity and binary restrictions on the variables. The objective function defined by (1) is simply the sum of unit production, fixed production and unit inventory costs.

The set of feasible solutions to (2)-(5) is called X^{ULS}.

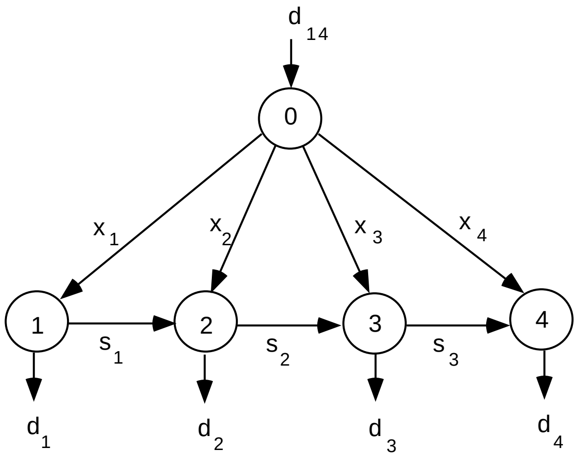

Fig. 3. Uncapacitated lot-sizing network $(n = 4)$

Master Production Scheduling Model. The next example is known as the multi item (single level) capacitated lot-sizing model. This model corresponds typically to the problem solved at the master production scheduling level in an MPC system. The purpose is to plan the production of a set of items, usually finished products, over a short term horizon corresponding to the total production cycle of these items. For each item, the model is the same as the ULS model in terms of costs and demand satisfaction. In addition, the production plans of the different items are linked through capacity restrictions coming from the common resources used to produce the items.

We define the indices $i = 1, \cdots, I$ to represent the set of items whose production has to be planned, $k = 1, \cdots, K$ to represent the set of shared resources with limited capacity, and $t = 1, \cdots, n$ to represent the time periods. The variables

x, y, s, and the data p, f, h, d, have the same meaning for each item i as in the model ULS. A superscript i has been added to represent the item i for which they are each defined.

The data L_t^k represents the available capacity of resource k during period t. The data α^{ik} and β^{ik} represent the amount of capacity of resource k consumed respectively per unit of item i produced, and for a setup of item i. The coefficient β^{ik} is often called the setup time of item i on resource k, and represents the time spent to prepare the resource k just before the production of a lot of item i. Together with α^{ik}, it may also be used to represent some economies of scale in the productivity factor of item i on resource k.

The natural formulation of this multi item capacitated lot-sizing model, or basic MPS model, can be written as follows.

$$\min \sum_i \sum_t \{p_t^i x_t^i + f_t^i y_t^i + h_t^i s_t^i\} \tag{6}$$

$$s_{t-1}^i + x_t^i = d_t^i + s_t^i \quad \text{for all } i, t \tag{7}$$

$$x_t^i \leq M y_t^i \quad \text{for all } i, t \tag{8}$$

$$\sum_i \alpha^{ik} x_t^i + \sum_i \beta^{ik} y_t^i \leq L_t^k \quad \text{for all } t, k \tag{9}$$

$$x_t^i, s_t^i \geq 0, \ y_t^i \in \{0, 1\} \quad \text{for all } i, t \tag{10}$$

where constraints (6)-(8) and (10) are the same as for the ULS model, and constraint (9) expresses the capacity restriction on each resource k in each period t.

Material Requirements Planning Model. As a last example model, we describe the multi-item multi-level capacitated lot-sizing model, that can be seen as the integration of the previous MPS model for finished products, and the ULS models for all intermediate products ans raw materials, into a single monolithic model. The purpose of this model is to optimize simultaneously the production and purchase of all items –from raw materials to finished products–, in order to satisfy for each item the external or independent demand coming from customers and the internal or dependent demand coming from the production of other items, over a short term horizon.

The dependency between items is modelled through the definition of the product structure, also called the bill of materials (BOM). The product structures are usually classified into Series, Assembly or General structures, see Figure 4.

The indices, variables and data are the same as before, except that, for simplicity, we also use the index $j = 1, \cdots, I$ to indentify items. For item i, we use the additional notation $S(i)$ to represent the set of successor items of i, i.e. the items consuming directly some amount of item i when they are produced. Note

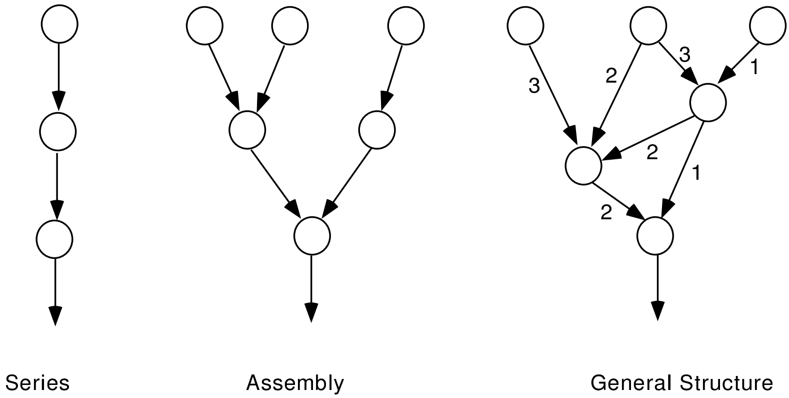

Fig. 4. Types of product structures in multi-level models

that for series and assembly structures, these sets $S(i)$ are singleton for all items i, and for a finished product i, we always have $S(i) = \emptyset$. And for $j \in S(i)$, we denote by r^{ij} the amount of item i required to make one unit of item j. These r^{ij} values are indicated along the edges (i, j) in Figure 4. This parameter r is used to identify the dependent demand, whereas d_t^i corresponds to the independent demand. For each item i, we denote by γ^i the lead time to produce or deliver any lot of i. More precisely, x_t^i represents the size of a production or purchase order of item i launched in period t, and delivered in period $t + \gamma^i$.

The natural formulation for the general product structure capacitated multi-level lot-sizing model, or the monolithic MRP model, is

$$\min \sum_i \sum_t \{p_t^i x_t^i + f_t^i y_t^i + h_t^i s_t^i\} \tag{11}$$

$$s_{t-1}^i + x_{t-\gamma^i}^i = [d_t^i + \sum_{j \in S(i)} r^{ij} x_t^j] + s_t^i \quad \text{for all } i, t \tag{12}$$

$$x_t^i \leq M y_t^i \quad \text{for all } i, t \tag{13}$$

$$\sum_i \alpha^{ik} x_t^i + \sum_i \beta^{ik} y_t^i \leq L_t^k \quad \text{for all } t, k \tag{14}$$

$$x_t^i,\ s_t^i \geq 0,\ y_t^i \in \{0, 1\} \quad \text{for all } i, t \tag{15}$$

where the only difference with respect to the previous MPS model resides in the form of the flow conservation constraint (12). For each item i in each period t, the amount delivered from production or vendors is $x_{t-\gamma^i}^i$ ordered in period $t - \gamma^i$, and the demand to be satisfied is the sum the independent demand d_t^i and the dependent demand $\sum_{j \in S(i)} r^{ij} x_t^j$ implied by the production of immediate successors $j \in S(i)$.

Because of the multi-level structure, the presence of single item ULS models as submodels is less obvious, but we will show in Section 5.1 how to reformulate this model in the form of single item ULS models linked by capacity and product structure restrictions.

2.3 Optimization Methods

We are interested in this lecture in the optimization approaches used to solve production planning problems. This means that we want either to find provably optimal solutions, or to find near-optimal solutions with a performance guarantee, expressed usually in terms of a percentage of deviation of the objective value from the optimal value.

As we have seen in the examples from Section 2.2, many or most multi item production planning problems can be modelled as the following mixed integer programming (MIP) model

$$MIP = \min \sum_i \sum_t \{p_t^i x_t^i + h_t^i s_t^i + f_t^i y_t^i\} \tag{16}$$

$$(x^i, s^i, y^i) \in W^i \quad \text{for all } i \tag{17}$$

$$(x^i, s^i, y^i)_{i=1,\cdots,I} \in P \tag{18}$$

$$\tag{19}$$

where the index t is used to represent time periods, the index i is used to represent the different items, W^i represents the set of feasible solutions (i.e. lot sizes, setups and inventory levels) to the item i lot-sizing problem – like ULS, or some of its variants studied in Section 3.5–, and P represents the set of solutions satisfying a set of coupling linear constraints – like the capacity constraints (9) in Section 2.2, among others–.

Most optimization methods are based on easy to solve relaxations of the initial problem, either to prove optimality, or to provide a performance guarantee of some near-optimal solution. For example, the above problem can be solved by some standard MIP software using a branch and bound approach based on the following linear relaxation

$$LR = \min \sum_i \sum_t \{p_t^i x_t^i + h_t^i s_t^i + f_t^i y_t^i\} \tag{20}$$

$$(x^i, s^i, y^i) \in W^i_{LR} \quad \text{for all } i \tag{21}$$

$$(x^i, s^i, y^i)_{i=1,\cdots,I} \in P \tag{22}$$

$$\tag{23}$$

where W^i_{LR} represents the linear relaxation of the set W^i obtained by relaxing the constraints $y_t^i \in \{0, 1\}$ into $0 \leq y_t^i \leq 1$, for all i, t.

Unfortunately, this direct MIP approach can only be used for the solution of small size problems. In order to solve or to find good solutions for more realistic or real size problems, one has to work with better or tighter relaxations. Because of the structure of the intial problem, most efficient solution approaches are based on the following relaxation

$$LB = \min \sum_i \sum_t \{p_t^i x_t^i + h_t^i s_t^i + f_t^i y_t^i\} \tag{24}$$

$$(x^i, s^i, y^i) \in conv(W^i) \quad \text{for all } i \tag{25}$$

$$(x^i, s^i, y^i)_{i=1,\cdots,I} \in P \tag{26}$$

$$\tag{27}$$

where $conv(W^i)$ represents the convex hull of the solutions of W^i.

This bound LB is never worse, and typically much tighter than the bound LR, and is precisely the relaxation bound exploited in many different methods.

– Lagrangean relaxation methods and Dantzig-Wolfe or column generation methods give rise to schemes where the bound LB is iteratively computed by relaxing or eliminating the coupling constraints (26), and solving separately for each item i the resulting single item subproblems. These subproblems are optimization problems defined over the sets W^i. See, for instance, [1], [2], [10], [17], [66], [67], [68] for Lagrangean relaxation approaches, and [30], [69] for column generation approaches. See also Lemarechal [⟶ Lemaréchal] for a general presentation of Lagrangean relaxation.
– If a compact (small size) linear description of $conv(W^i)$ is available for each item i, then by adding the coupling constraints one gets a formulation whose linear relaxation gives the bound LB. Then a direct branch and bound approach can be used with this formulation. See, for instance, [19], [54] and [81].
– If a linear but non compact description of $conv(W^i)$ is available for each item i, using the corresponding formulation directly to compute LB takes too much time. Then the classical approach is to add (some of) the constraints defining $conv(W^i)$ in the course of optimization, instead of a priori, by solving a so-called separation problem over the set $conv(W^i)$. This gives branch and cut type methods. See, for instance, [7], [12] and [55].

Therefore, in order to design optimization methods for solving complex multi item production planning problems, one has to design algorithms to solve the single item subproblem, or one has to find a compact complete linear description for the subproblem, or one has to find a complete linear description with an efficient separation algorithm for the subproblem.

3 The CORE Subproblem: Single Item Uncapacitated Lot-Sizing

First, we study in detail in this Section the uncapacitated lot-sizing problem (ULS) already introduced in Section 2.2. Next, we will give the main optimization and reformulation results known for the main variants of this single item problem. A number of formulations and the general structure of this Section are imported from the more technical paper by Pochet and Wolsey [58], and from Chapter 13 of Wolsey [83].

3.1 Basic Formulation and Motivation

For the ease of reading, we have tried to separate as much as possible the previous modelling Section from this more technical Section. Therefore, we recall here the basic formulation of the ULS problem.

$$\min \sum_{t=1}^{n} (p_t x_t + f_t y_t + h_t s_t) \tag{28}$$

$$s_{t-1} + x_t = d_t + s_t \quad \text{for all } t \tag{29}$$

$$s_0 = s_n = 0 \tag{30}$$

$$x_t \leq d_{tn} y_t \quad \text{for all } t \tag{31}$$

$$x_t, s_t \geq 0, \ y_t \in \{0,1\} \quad \text{for all } t \tag{32}$$

where the large upper bound on x_t in constraint (31) has been replaced by its true upper bound d_{tn} equal to the sum of the demands from period t up to end of the horizon. In general, we define d_{kl} by $d_{kl} = \sum_{t=k}^{l} d_t$ for all $k \leq l$.

The set of feasible solutions to (29)-(32) is called X^{ULS}.

The first motivation to study this specific ULS problem is that ULS is solved repeatedly in MRP systems for each component, backward in the product structure from end products to raw materials. This approach, to define the production plans using a sequential item by item and uncapacitated model, has important shortcomings. By solving the capacity problems only after the solution of ULS for all items, and by neglecting the multi-level product structure, the production plans obtained are suboptimal in terms of cost and flexibility. The extension to capacitated models and multi-level models are discussed in Sections 4 and 5, respectively.

The other motivation to study ULS comes from the fact that it is the most common subproblem arising in complex multi-item production planning problems, and we have already seen in Section 2.3 the central role of single item subproblems in designing decomposition or branch and cut solution approaches for multi item problems.

3.2 Dynamic Programming Algorithm

The first natural question to ask in studying a subproblem is the following.

Question 1: *Is optimization over X^{ULS} polynomially solvable ?*

The answer is yes, and comes from a well known decomposition property of optimal solutions (see, for example, Zangwill [84]). This property means that there always exists an optimal solution where production only occurs in a period t when the entering stock s_{t-1} is zero.

Theorem 1 *There exists an optimal solution to problem ULS (28)-(32) where $s_{t-1}x_t = 0$ for all t*

Proof. Production may only occur in periods t with $y_t = 1$. For any selection of such production periods, the problem reduces to a minimum cost flow problem through the network shown in Figure 3. In any extreme solution, the arcs with positive flow form an acyclic subgraph and any acyclic subgraph satisfies $s_{t-1}x_t = 0$ for all t. ∎

There are a number of algorithms based on this decomposition property. The seminal paper of Wagner and Whitin [78] describes an $O(n^2)$ dynamic programming algorithm solving ULS. More efficient implementations of this algorithm running in $O(n \log n)$ have been proposed by Federgrun, Tsur [20], Wagelmans, van Hoesel, Kolen [77], and Aggarwal, Park [3].

We describe here a simple forward dynamic program with a trivial implementation in $O(n^2)$. For simplicity of notation, we eliminate first the stock cost coefficients from the objective function using the flow balance constraints (29). That is, we replace $h_t s_t$ by $h_t[\sum_{k=1}^t x_k \; - d_{1,t}]$ in the objective (28), for all t (Recall that $d_{1,t} = \sum_{k=1}^t d_k$). Observe also that if $f_t < 0$ for some t in the objective (28), then $y_t = 1$ in all optimal solutions to (28)-(32). Therefore, problem ULS can be solved using the following objective function.

$$\min \sum_{t=1}^{n} (c_t x_t + f_t^+ y_t) \tag{33}$$

with $c_t = p_t + \sum_{i=t}^n h_i$ and $f^+ = \max\{f, 0\}$. In other words, to obtain the optimal solution to ULS ((28)-(32)) is suffices to add $\sum_{t=1}^n f_t^-$, with $f^- = \min\{f, 0\}$, to the optimal solution of the transformed problem (33), (29)-(32). From now on, and unless otherwise mentioned, we will work with this transformed objective function.

Now, let $H(k)$ be the cost of the minimum cost solution for problem ULS restricted to periods 1 up to k. From Theorem 1, there exists a minimum cost solution of value $H(k)$ such that, if period $t \leq k$ is the last production period before period k, then $x_t = d_{tk}$, $x_l = 0$ for $l = t+1, \cdots, k$, and the cost of the solution for periods 1 to $t-1$ must be $H(t-1)$. This holds by the optimality principle, and the fact that $s_{t-1} = 0$ when $x_t > 0$. This allows us to define the following recursion to compute $H(k)$.

$$H(k) = \min_{1 \le t \le k} \{H(t-1) + f_t^+ + c_t \; d_{tk}\} \tag{34}$$

Starting from $H(0) = 0$, and computing $H(k)$ using (34) for $k = 1, \cdots, n$ leads to the value

$$H(n) \; + \; \sum_{t=1}^{n} f_t^- \tag{35}$$

of an optimal solution to ULS. It is easy to check that a direct implementation of this recursion leads to an $O(n^2)$ algorithm for ULS.

3.3 Extended Reformulations

Given the equivalence between optimization and separation (see Grotschel, Lovasz and Schrijver [25]), and given the fact that ULS is polynomially solvable and appears as a subproblem in many complex multi-item lotsizing problems, the next question to ask in order to solve these problems by branch and bound is now.

Question 2: *Is there a compact linear description for ULS ?*

The answer is again yes for such a simple lot-sizing problem, and we describe three of the popular extended reformulations for ULS.

Facility Location. The first reformulation takes the form of a simple facility location model, defined for facilities and customers located on a line, and introduced by Krarup, Bilde [34]. It consists in locating facilities ($y_t = 1$ if a facility is located at location t) to serve customer demands on a one way road (d_t is the demand to meet at location t) and minimize installation costs (f_t at location t), unit production costs (p_t at location t) and unit transportation costs (h_t from location t to location $t+1$).

If we define the variable w_{st} as the fraction of the demand d_t served from a production in period s (i.e. a facility located in s), for all $1 \le s \le t \le n$, and y_s as the usual 0/1 setup variable for period $s = 1, \cdots, n$, then, by analogy to the location model, ULS can be reformulated as

$$\text{(FL)} \qquad \min \sum_{t=1}^{n} f_t y_t + \sum_{s=1}^{n} \sum_{t=s}^{n} c_s d_t w_{st} \tag{36}$$

$$\sum_{s=1}^{t} w_{st} = 1 \quad \text{for all } t \tag{37}$$

$$w_{st} \leq y_s \quad \text{for all } 1 \leq s \leq t \leq n \tag{38}$$

$$w_{st}, y_t \geq 0 \quad \text{for all } 1 \leq s \leq t \leq n \tag{39}$$

$$y_t \leq 1 \quad \text{for all } t \tag{40}$$

$$y_t \text{ integer} \quad \text{for all } t \tag{41}$$

where constraint (37) expresses the demand satisfaction in period t from productions in periods $s \leq t$, constraint (38) imposes a setup in period s when there is production in s to satisfy some demand in period $t \geq s$, constraints (39)-(41) impose the nonnegativity and binary restrictions on the variables, the objective function (36) expresses the cost of the production plan in the form (33).

Formulation (36)-(41) is a valid reformulation for ULS, but due to the particular structure of the objective function, we have the following stronger result from Krarup, Bilde [34] (see also Barany, Van Roy and Wolsey [6] for a primal-dual proof).

Theorem 2 *The linear programming relaxation (36)-(40)of FL always has an optimal solution with y integer, and solves ULS.*

Note that the polyhedron (37)-(40) has fractional vertices. However, as there is always an optimal integral solution to (36)-(40), this linear formulation suffices to solve ULS. In order to obtain an integral polyhedron corresponding to the convex hull of solutions to ULS (i.e. a linear formulation with all integral vertices), one needs to add the constraints $w_{st} \geq w_{s,t+1}$ for $1 \leq s \leq t < n$ to (37)-(40).

Multicommodity. A classical way to tighten the formulation of fixed charge network flow problems is to decompose the flow along each arc of the network as a function of its destination. This defines a so-called multicommodity formulation by assigning a different commodity to each destination node, see Rardin and Choe [60].

We define the variable x_{it} (resp. s_{it}) as the production of commodity t (resp. inventory of commodity t) in period i. Commodity t corresponds to the demand delivered in period t. By decomposing the flow by commodity, ULS can be reformulated as

$$\text{(MC)} \qquad \min \sum_{i=1}^{n} \sum_{t=i}^{n} c_i x_{it} + \sum_{i=1}^{n} f_i y_i \tag{42}$$

$$s_{i-1,t} + x_{it} = \delta_{it} d_t + s_{it} \quad \text{for } 1 \leq i \leq t \leq n \tag{43}$$

$$s_{0t} = s_{tt} = 0 \quad \text{for } 1 \leq t \leq n \tag{44}$$

$$x_{it} \leq d_t y_i \quad \text{for } 1 \leq i \leq t \leq n \tag{45}$$

$$s_{it}, x_{it}, y_{it} \geq 0,\ y_{it} \leq 1 \quad \text{for } 1 \leq i \leq t \leq n \tag{46}$$

$$y_{it} \text{ integer} \quad \text{for } 1 \leq i \leq t \leq n \tag{47}$$

where the notation δ_{it} denotes 1 if $i = t$, and 0 otherwise. Constraint (43) is the flow conservation constraint of commodity t in period $i \leq t$. Constraints (44) impose there is no intial and no final inventory (end of period t) of commodity t. Constraint (45) forces the setup variable y_i to be 1 when there is production in period i, and using the decomposition of the flow the upper bound on x_{it} is d_t. Constraints (46)-(47) impose the nonnegativity and binary restrictions on the variables and constraint (42) expresses the cost of the production plan.

Formulation (42)-(47) is a valid reformulation for ULS. Using the equations (43) to eliminate the inventory variables from formulation (MC), and using the substitution $w_{s,t} = x_{s,t}/d_t$ between formulations MC and FL, it is easy to prove the following result.

Theorem 3 *The linear relaxations of MC ((42)-(46)) and FL ((36)-(40)) are equivalent. Therefore,the linear programming relaxation of MC always has an optimal solution with y integer, and solves ULS.*

Shortest Path. The last reformulation technique that we illustrate here is due to Martin [44], and first applied to lot-sizing problems by Eppen and Martin [19]. It can be used to transform the dynamic programming algorithm for ULS into a pair of primal-dual linear formulations.

The dynamic programming algorithm (34) constructs a least cost succession of so-called (disjoint) regeneration intervals $[t, k]$, with $t \leq k$, covering the planning horizon from period 1 to period n. Each interval $[t, k]$ represents the satisfaction of the demands in periods t up to k by the production of d_{tk} in period t. This is illustrated in Figure 5 with $n = 7$, where the entire production plan consists in the succession of 3 regeneration intervals: $[1, 2], [3, 6], [7, 7]$.

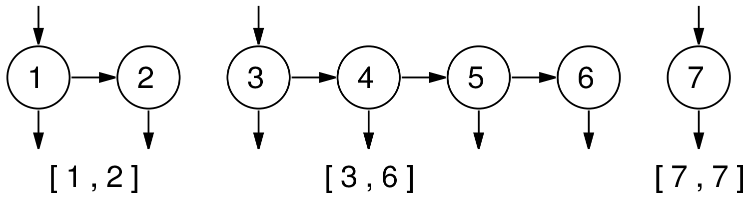

Fig. 5. An extreme solution as a succession of regeneration intervals

Using as variables $H(k)$, for $k = 0, \cdots, n$, the following linear program computes the optimal value $H(n) + \sum_{t=1}^{n} f_t^-$ of the dynamic program (34) solving ULS.

$$\max \quad H(n) - H(0) \quad + \quad \sum_{t=1}^{n} f_t^- \tag{48}$$

$$H(k) \leq H(t-1) + f_t^+ + c_t \, d_{tk} \quad \text{for all } 1 \leq t \leq k \leq n \tag{49}$$

$$H(0) = 0 \tag{50}$$

This claim is easy to prove by observing that the constraints (49) are equivalent to $H(k) \leq \min_{1 \leq t \leq k}\{H(t-1) + f_t^+ + c_t \, d_{tk}\}$, for all k. The maximization direction in the objective function (48) suffices to guarantee that the values $H(k)$, for all k, computed in the dynamic program (34) define an optimal solution to (48)-(50).

Associating the dual variable $\phi_{t,k}$ to each constraint in (49), the dual of (48)-(50) can be written as the minimum cost network flow problem

$$\min \sum_{t=1}^{n} \sum_{k=t}^{n} [c_t d_{tk} + f_t^+]\phi_{tk} \ + \ \sum_{t=1}^{n} f_t^- \tag{51}$$

$$\sum_{t=1}^{n} \phi_{tn} = 1 \tag{52}$$

$$\sum_{i=1}^{t} \phi_{it} - \sum_{l=t+1}^{n} \phi_{t+1,l} = 0 \quad \text{for } 1 \leq t \leq n-1 \tag{53}$$

$$-\sum_{l=1}^{n} \phi_{1l} = -1 \tag{54}$$

$$\phi_{tk} \geq 0 \quad \text{for } 1 \leq t \leq k \leq n \tag{55}$$

where the constraints (52)-(54) correspond to the flow conservation constraints of the network flow problem, with one flow constraint for each time period, and one unit of flow sent from the single source node (52) to the single sink node (54). Therefore, the linear program (51)-(55) defines a shortest path problem from source to sink, and it has 0-1 extreme points. It is easy to see that this formulation models the possible sequences of regeneration intervals, with $\phi_{t,k} = 1$ if the regeneration interval $[t, k]$ is part of the solution at cost $[c_t d_{tk} + f_t^+]$. This is illustrated in Figure 6 for the case $n = 3$.

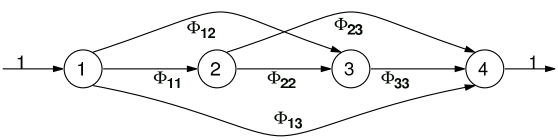

Fig. 6. Shortest Path Formulation of ULS ($n = 3$)

A final reformulation step can be used to obtain an equivalent formulation where the setup decisions are made explicit.

$$\text{(SP)} \qquad \min \sum_{t=1}^{n}\sum_{k=t}^{n} c_t d_{tk}\phi_{tk} + \sum_{t=1}^{n} f_t y_t \tag{56}$$

$$\sum_{t=1}^{n} \phi_{tn} = 1 \tag{57}$$

$$\sum_{i=1}^{t} \phi_{it} - \sum_{l=t+1}^{n} \phi_{t+1,l} = 0 \quad \text{for } 1 \leq t \leq n-1 \tag{58}$$

$$-\sum_{l=1}^{n} \phi_{1l} = -1 \tag{59}$$

$$\sum_{k=t}^{n} \phi_{tk} \leq y_t \quad \text{for all } t \tag{60}$$

$$\phi_{tk}, y_t \geq 0,\ y_t \leq 1 \quad \text{for } 1 \leq t \leq k \leq n \tag{61}$$

In order to obtain this formulation, we have first replaced the objective term $\sum_{k=t}^{n}[f_t^+]\phi_{tk}$ by $[f_t^+]y_t$ and added the setup defining equalities $\sum_{k=t}^{n} \phi_{tk} = y_t$, for all t. Next we have replaced $\sum_{t=1}^{n} f_t^+ y_t + \sum_{t=1}^{n} f_t^-$ by $\sum_{t=1}^{n}[f_t^+ + f_t^-]y_t = \sum_{t=1}^{n} f_t y_t$, and relaxed $\sum_{k=t}^{n} \phi_{tk} = y_t$ into $\sum_{k=t}^{n} \phi_{tk} \leq y_t$ to be able to force $y_t = 1$ when $f_t < 0$ without perturbing the flow constraints.

The above discussion has shown the following.

Theorem 4 *The linear program SP ((56)-(61)) always has an optimal solution with y integer, and solves ULS.*

Note that here the polyhedron (57)-(61) has no fractional vertices. To relate this shortest path formulation to the other formulations, one can show that formulation (57)-(61) is equivalent to the facility location formulation (37)-(40), augmented with $w_{st} \geq w_{s,t+1}$ for $1 \leq s \leq t < n$, using the substitution $\phi_{tk} = w_{tk} - w_{t,k+1}$ for $1 \leq t \leq k \leq n$.

3.4 Complete Linear Description in the Initial Space

We have described in Section 3.3 several linear formulations for the ULS problem with n periods, but they involve $O(n^2)$ variables and $O(n^2)$ or $O(n)$ constraints.

Although these formulations are as tight as possible, to be able to solve large size multi item production planning problems – with many items and many periods –, we need to reduce the size of the formulation used for the single item subproblems. One way to achieve this is to look for a complete linear description of X^{ULS} in the initial variable space with only $O(n)$ variables. In that case, even if this complete linear description needs an exponential number of constraints, we do not need to add them all (a priori) to solve a particular instance. If we add the constraints using a separation algorithm, then at most $O(n)$ variables and constraints are needed to describe any particular extreme point of X^{ULS}.

This is typically what happens when branch and cut is used instead of branch and bound.

In order to identify a complete linear description of $conv(X^{ULS})$ in the initial space of variables, we first need to answer to the following question.

Question 3: *How to find valid inequalities for the set X^{ULS}?*

To answer to that question, we describe the typical way a researcher could think of in order to identify classes of valid inequalities.

One approach is to find numerically the complete description of small instances, and to generalize the results to classes of valid inequalities for arbitrary data and problem dimension.

Another approach is to solve the LP relaxations of some instances, and to find valid inequalities cutting off the non-integral optimal solutions so obtained, and then to generalize the results to classes of valid inequalities for arbitrary data and problem dimension. This approach gives also some intuition for the separation problem.

The example and the procedure used to find valid inequalities are taken from Wolsey [83]. Consider the fractional solution pictured in Figure 7 and obtained by solving the linear programming relaxation of the natural formulation (29)-(32) of some specific instance of ULS.

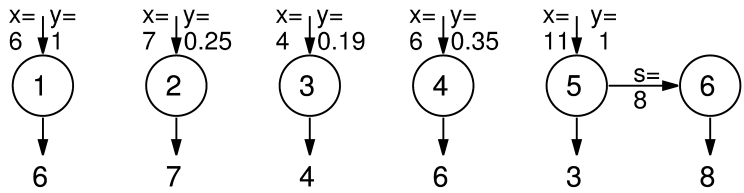

Fig. 7. A fractional solution($n = 6$)

Observation: As in the dynamic program, the solution decomposes into intervals of periods between successive points with $s_t = 0$. Intervals in which all the y variables are integral, like interval $[5, 6]$ in Figure 7, cannot be cut off because they correspond to integer (partial) solutions in $conv(X^{ULS})$. Therefore, we look at intervals in which the corresponding y variables are fractional, like interval $[2, 2]$ in Figure 7, and try to cut-off the corresponding point.

This can be done either by generating valid inequalities using known cutting planes, or by generating new and specialized valid inequalities for the problem at hand. We illustrate these two approach types in turn on this very simple problem and on the fractional interval $[2, 2]$ from Figure 7.

Valid Inequalities Using Known Cutting Planes. The fractional interval $[2, 2]$ from Figure 7 corresponds to a fractional solution of the single node or single period flow model defined by

$$\{(x_2, y_2, s_1, s_2) \in R^1_+ \times B^1 \times R^2_+ : s_1 + x_2 = d_2 + s_2,\ x_2 \leq d_{2,n} y_2\} \tag{62}$$

where $d_{2,n} = 28$ and $d_2 = 7$ in our toy example.

The well known flow cover inequalities have been defined for such single node flow models, see Padberg, Van Roy and Wolsey [50]. Let the cover be $C = \{x_2\}$, with excess capacity $\lambda = d_{2,n} - d_2$. The resulting flow cover inequality is

$$x_2 \leq d_2 \ - (d_{2,n} - \lambda)(1 - y_2) \ + s_2 \quad \text{, or equivalently} \tag{63}$$

$$x_2 \leq d_2 \ y_2 \ + s_2. \tag{64}$$

which gives $x_2 \leq 7 \ y_2 \ + s_2$ in our example.

This inequality is violated by the current fractional point from Figure 7, and it can be easily generalized to

$$x_t \leq d_t y_t \ + s_t \ \text{ forall } t \tag{65}$$

for arbitrary demand data and time period.

New Valid Inequalities. Consider again the fractional interval $[2, 2]$ from Figure 7. In any feasible solution, if there is no production in period 2 (which surely occurs if $y_2 = 0$), then the demand d_2 for period 2 must be produced earlier and must be contained in the entering stock s_1.

Therefore, a logical implication is $y_2 = 0$ implies $s_1 \geq d_2$. This implication can be converted into the valid linear inequality $s_1 \geq d_2(1 - y_2)$, which can be interpreted as if $y_2 = 0$ then $s_1 \geq d_2$, and if $y_2 = 1$ then $s_1 \geq 0$.

This inequality is violated by the current fractional point from Figure 7, and can be easily generalized to

$$s_{t-1} \geq d_t(1 - y_t) \ \text{ forall } t \tag{66}$$

for arbitrary demand data and time period.

Reoptimization and Generalization. After some violated valid inequalities have been identified, they can be added to the initial formulation, and the resulting tighter linear formulation solved.

Using the flow balance equality, $s_{t-1} + x_t = d_t + s_t$, one observes that the 2 classes of valid inequalities found (65) and (66) are equivalent. We add these inequalities for $t = 2,\ 3,\ 4$ to the formulation, reoptimize, and get the new non-integral solution pictured in Figure 8. Again, to find new violated valid inequalities, it suffices to look at the solution for periods 1 up to 3 because the partial solution for periods 4 to 6 is integral.

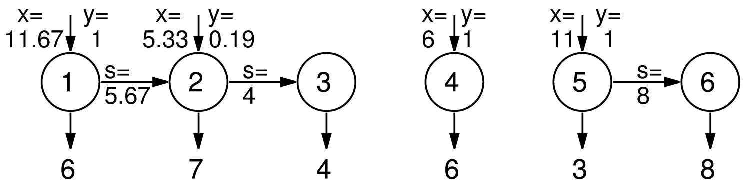

Fig. 8. A fractional solution after some cuts ($n = 6$)

By the same type of logical arguments, one can derive the valid inequality $s_1 \geq d_2(1-y_2) + d_3(1-y_2-y_3)$, which says that s_1 must carry over d_2 if $y_2 = 0$ and must also carry over d_3 if both $y_2 = 0$ and $y_3 = 0$.

This inequality is violated by the current fractional point from Figure 8, and it can be generalized to

$$s_{t-1} \geq \sum_{k=t}^{l} d_k(1 - y_t - \cdots - y_k) \ \text{ for } 2 \leq t \leq l \leq n \tag{67}$$

Complete Linear Description. Once classes of valid inequalities have been identified, the next natural questions to ask are, in sequence.

Question 4: *Are the valid inequalities (67) facet defining for conv*(X^{ULS})*?*
Question 5: *Do we have a complete linear description of conv*(X^{ULS})*?*

We will not describe here the various techniques that can be used to prove that some valid inequalites are facet defining, and that some valid inequalities suffice to describe the convex hull of a problem. For a general presentation of these topics, we refer the reader to Nemhauser and Wolsey [47]. We will rather state the main results for the ULS problem, with adequate references for their proofs.

Using the flow balance constraints to eliminate the stock variables (i.e. $s_t = [\sum_{k=1}^{t} x_k \ - d_{1,t}]$), the class of valid inequalities (67) is rewritten as

$$\sum_{k=1}^{t-1} x_k \;+\; \sum_{k=t}^{l} d_{k,l}\; y_k \geq d_{1,l} \;\text{ for } 2 \leq t \leq l \leq n \tag{68}$$

These inequalities (68) do not suffice to describe $conv(X^{ULS})$, but they suffice to solve most practical instances without any branching, i.e. the solution of the linear relaxation of (28)-(32), augmented with the inequalities (68), is almost always integral. We will see later why, and in which sense, this is the case (see Section 3.5). To obtain a linear description of $conv(X^{ULS})$, the inequalities (68) have to be further generalized, as shown in the following Theorem.

Theorem 5 *Assuming $d_t > 0$ for all t, $conv(X^{ULS})$ is described by*

$$\sum_{i \in L \setminus S} x_i + \sum_{i \in S} d_{il} y_i \geq d_{1l} \quad \text{for } 1 \leq l \leq n, S \subseteq L = \{1, \cdots, l\} \tag{69}$$

$$y_1 = 1 \tag{70}$$

$$\sum_{i=1}^{n} x_i = d_{1n} \tag{71}$$

$$x_i \geq 0,\; 0 \leq y_i \leq 1 \quad \text{for all } i \tag{72}$$

Barany, Van Roy, Wolsey [6] give a primal-dual proof of Theorem 5, Pochet and Wolsey [58] give a direct proof using a technique due to Lovasz [40].

Separation. We have obtained a complete linear programming formulation in the original variables (x, y) that solves ULS. Unfortunately this formulation contains an exponential number of so-called (l, S) inequalities (69). In a branch and cut or cutting plane approach, instead of adding all these inequalities a priori to the formulation, they are added in the course of optimization when they are needed. We then need to solve the following separation problem.

We define X^{ULS}_{LR} as the linear relaxation of the natural formulation (29)-(32) of ULS.

Separation Given $(x^*, y^*) \in X^{ULS}_{LR}$:

- Either we find an (l, S) inequality violated by (x^*, y^*),
- or we prove that all (l, S) inequalities are satisfied by (x^*, y^*).

To find the most violated (l, S) inequality for fixed $l \in \{1, \cdots, n\}$, it suffices to test whether or not $\sum_{i=1}^{l} \min(x_i^*, d_{il} y_i^*) < d_{1l}$.

- If this holds, then the (l, S^*) inequality with $S^* = \{i \in L : d_{il} y_i^* \leq x_i^*\}$ is the most violated inequality for the given value of l.
- Otherwise, there is no violated (l, S) inequality for the given value of l.

By enumerating all possible values of l, we obtain an $O(n^2)$ separation algorithm for the (l, S) inequalities.

3.5 Single Item Variants of Uncapacitated Lot-Sizing

Using the same approach as for ULS, we describe here very briefly several ULS variants that have been studied in the literature, and for which complexity and reformulation results are available.

Start-up Costs. We consider first the single item lot-sizing problem with start-up costs. This model is identical to ULS, with an additional startup cost incurred in the first period of a production sequence or batch. More precisely, the startup cost is incurred in period t, if there is a setup in period t, but not in period $t-1$. To introduce this startup cost model into the formulation, we define additional binary variables z_t taking the value 1 when there is a startup in period t. This leads to the formulation

$$\min \sum_t \{p_t x_t + f_t y_t + h_t s_t + g_t z_t\} \tag{73}$$

$$s_{t-1} + x_t = d_t + s_t \quad \text{for all } t \tag{74}$$

$$s_0 = s_n = 0 \tag{75}$$

$$x_t \leq M y_t \quad \text{for all } t \tag{76}$$

$$y_t \geq z_t \geq y_t - y_{t-1} \quad \text{for all } t \tag{77}$$

$$x_t, s_t \geq 0,\ y_t, z_t \in \{0,1\} \quad \text{for all } t \tag{78}$$

where constraints (77) impose a startup in period t if there is a setup in t, but not in $t-1$, and also force a setup in period t if there is a startup in period t. We define ULSS as the optimization problem (73)-(78).

Magnanti and Vachani [42] show that the decomposition property of optimal solutions given in Theorem 1 still holds. This allows one to define a dynamic programming algorithm for solving ULSS, as in van Hoesel [71], with a direct implementation running in $O(n^2)$. Again, more effcient implementations run in $O(n\ log\ n)$, see van Hoesel [71] and Aggarwal, Park [3].

The techniques used to generate compact linear formulations for ULS in Section 3.3 can be extended to ULSS. In particular, Wolsey [81] provides a facility location type formulation and van Hoesel, Wagelmans and Wolsey [72] provide a proof that its linear relaxation solves ULSS, Rardin and Wolsey [61] prove that the muticommodity linear formulation solves ULSS, and the technique of Martin [44] can be used to generate a shortest path linear formulation from the dynamic program solving ULSS. All these compact linear formulations involve $O(n^2)$ variables and constraints.

A class of valid inequalities can easily be obtained starting from the (l,S) inequalities (69) of ULS. For example, starting with the (l,S) inequality $s_1 \geq d_2(1-y_2) + d_3(1-y_2-y_3) + d_4(1-y_2-y_3-y_4)$, which is also valid in the presence of startup costs, one can derive the stronger valid inequality $s_1 \geq$

$d_2(1-y_2) + d_3(1-y_2-z_3) + d_4(1-y_2-z_3-z_4)$. This inequality says that s_1 must carry over d_2 if $y_2 = 0$, must also carry over d_3 if $y_2 = z_3 = 0$ (which implies that $y_3 = 0$, and thus there is no production in periods 2 and 3), and must carry over d_4 when $y_2 = z_3 = z_4 = 0$ (which implies $y_3 = y_4 = 0$).

This inequality can be generalized to

$$s_{t-1} \geq d_t(1-y_t) + \sum_{k=t+1}^{l} d_k(1-y_t-z_{t+1}-\cdots-z_k) \quad \text{for } 2 \leq t < l \leq n. \tag{79}$$

However, as for ULS, the inequalities (79) do not suffice to obtain a linear description of the convex hull of solutions, but they suffice to solve most practical instances by linear programming, without any branching (see Section 3.5). A further generalization of inequalities (79) proposed by van Hoesel, Wagelmans and Wolsey [72] is proved to define the convex hull of feasible solutions to ULSS in the initial variable space. This formulation contains an exponential number of constraints that can be separated in time of $O(n^3)$.

Constant Capacity. We consider now the single item constant capacity lot-sizing problem (CCLS) defined by (80)-(84). This constant capacity model is quite realistic and occurs often in practice when the production, inventory and demand are expressed in equivalent-time units.

$$\min \sum_t \{p_t x_t + f_t y_t + h_t s_t\} \tag{80}$$

$$s_{t-1} + x_t = d_t + s_t \quad \text{for all } t \tag{81}$$

$$s_0 = s_n = 0 \tag{82}$$

$$x_t \leq C y_t \quad \text{for all } t \tag{83}$$

$$x_t, s_t \geq 0,\ y_t \in \{0,1\} \quad \text{for all } t \tag{84}$$

The formulation of CCLS is identical to the natural formulation of ULS, except that constraint (83) expresses now a limited capacity when $C < d_{1n}$. The capacity limit C is constant over time.

Florian and Klein [23] propose a dynamic programming programming algorithm for CCLS running in $O(n^4)$ time, and van Hoesel, Wagelmans [73] reduce this time to $O(n^3)$. The technique of Martin [44] can be used to generate a shortest path linear formulation from the dynamic program solving CCLS with $O(n^4)$ variables and constraints. Pochet and Wolsey [56] give a linear programming formulation with $O(n^3)$ variables and constraints solving CCLS that corresponds to a shortest path sequence of regeneration intervals, but cannot be derived from the dynamic program solving CCLS.

Pochet and Wolsey [56] describe also a new class of valid inequalities for CCLS. This class was generalized to other problems than CCLS, and a new procedure to obtain these inequalities by mixing the mixed-integer rounding inequalities of Nemhauser and Wolsey [48] is described in Gunlunk and Pochet [26]. We illustrate here this class of inequalities with an example. The reader is referred to the above papers for a formal definition of these inequalities.

Consider the instance of CCLS with $C = 10$, $n = 5$ and demand $d = (1, 7, 4, 2, 3)$ represented in Figure 9. First, by summing the flow conservation constraints (81) from period 2 to some period $l \geq 2$, by replacing the production variables x_t by their variable upper bounds Cy_t, one gets the following starting or base (valid but redundant) inequalities

$$s_1 + 10y_2 \geq 7, \tag{85}$$
$$s_1 + 10y_2 + 10y_3 \geq 11, \tag{86}$$
$$s_1 + 10y_2 + 10y_3 + 10y_4 + 10y_5 \geq 16. \tag{87}$$

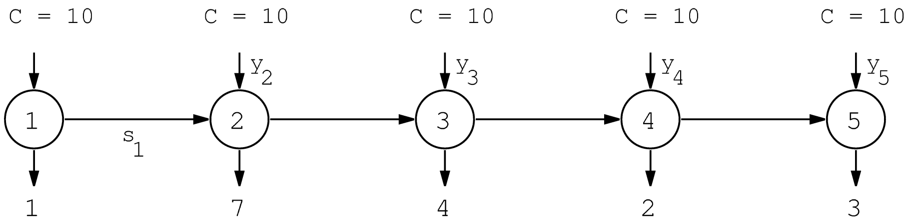

Fig. 9. A CCLS example($n = 5$)

Then, standard mixed-integer rounding (MIR) inequalities [$\longleftarrow$ Martin] for a general description of MIR inequalities) can be constructed from these base inequalities. This gives

$$s_1 \geq 7(1 - y_2), \tag{88}$$
$$s_1 \geq 1(2 - y_2 - y_3), \tag{89}$$
$$s_1 \geq 6(2 - y_2 - y_3 - y_4 - y_5). \tag{90}$$

Finally, the MIR inequalities can be mixed by taking differences of terms at the right hand side. This gives, for example, the new inequalities

$$s_1 \geq 1(2 - y_2 - y_3) + (6 - 1)(2 - y_2 - y_3 - y_4 - y_5) + (7 - 6)(1 - y_2), \tag{91}$$
$$s_1 \geq 1(2 - y_2 - y_3) + (7 - 1)(1 - y_2), \tag{92}$$
$$s_1 \geq 6(2 - y_2 - y_3 - y_4 - y_5) + (7 - 6)(1 - y_2). \tag{93}$$

For example, the second of these inequalities can be interpreted as follows. As $d_{23} = 11$, if there is only one production period in periods 2 and 3 (i.e. $y_2+y_3 = 1$, so a capacity of 10 is available in periods 2 and 3), then s_1 must carry over at least one unit in order to satisfy demand in periods 2 and 3. But if, in addition, $y_2 = 0$, then s_1 must carry over the whole demand $d_2 = 7$ (instead of only 1 unit when $y_2 = 1$ and $y_3 = 0$).

Such mixed MIR inequalities do not suffice to describe the convex hull of solutions for CCLS by linear inequalites only, but they suffice to solve most practical instances without branching (see Section 3.5). Examples of facet defining inequalities that are not of the type (91)-(93) are given in Pochet and Wolsey [56]. However, in this case, no complete linear description of the convex hull of solutions is known for CCLS.

Leung, Magnanti and Vachani [37], Pochet [53], Loparic, Marchand and Wolsey [39] define and test valid inequalities for the case where the capacities may vary over time.

Backlogging. Another variant of ULS is to allow for backlogging. To model this situation, we introduce a new variable r_t, for each period t, representing the backlog of demand at the end of period t. So, r_t is the demand from periods $\{1, \cdots, t\}$ that will be delivered late (in some period in $\{t+1, \cdots, n\}$). A classical formulation for the uncapacitated lot-sizing problem with backlogging (BLS) is (94)-(98)

$$\min \quad \sum_t \{p_t x_t + f_t y_t + h_t s_t + g_t r_t\} \tag{94}$$

$$(s_{t-1} - r_{t-1}) + x_t = d_t + (s_t - r_t) \quad \text{for all } t \tag{95}$$

$$s_0 = r_0 = s_n = r_n = 0 \tag{96}$$

$$x_t \leq M y_t \quad \text{for all } t \tag{97}$$

$$x_t, s_t, r_t \geq 0, \ \ y_t \in \{0,1\} \quad \text{for all } t \tag{98}$$

where the flow conservation constraints (95) are adapted to take into account these backlogging variables. In each period t, r_t represents a flow coming from period $t+1$ to period t (to satisfy artificially some demand in periods at or before t). Constraints (96) impose that there is no initial and final inventory and backlog. It is assumed in 94 that each unit of product backlogged at the end of period t costs g_t.

Zangwill [85] describes a dynamic programming programming algorithm for BLS running in $O(n^2)$ time, and Wagelmans, van Hoesel, Kolen [77] obtain an implementation in $O(n \ log \ n)$ time. Pochet and Wolsey [54] provide facility location, multicommodity and shortest path linear formulations involving $O(n^2)$ variables and constraints.

Pochet and Wolsey [54] describe a new class of valid inequalities for BLS. This class, that can also be applied to other problems than BLS, gives a way to generalize the class of valid inequalities defined in Van Roy and Wolsey [74] for uncapacitated fixed charge network flow problems. We illustrate this class of inequalities with a simple example. The reader is referred to the above papers for a formal definition of these inequalities.

Consider an instance of BLS with $n \geq 5$, and represented in Figure 10. First we define a subset of the inventory and backlog variables, for example $\{s_1, s_3, r_2, r_4\}$, and then we find logical conditions under which some demand has to flow through the selected variables. For example, if $y_2 = 0$ then $s_1 + r_2 \geq d_2$, giving the base valid inequality $s_1 + r_2 \geq d_2(1 - y_2)$. Similarly, one can obtain the base valid inequalities $s_1 + r_4 \geq d_3(1 - y_2 - y_3 - y_4)$ and $s_3 + r_4 \geq d_4(1 - y_4)$. Finally we put all these inequalities together to obtain the valid inequality

$$s_1 + s_3 + r_2 + r_4 \geq d_2(1 - y_2) \tag{99}$$
$$+ d_3(1 - y_2 - y_3 - y_4) \tag{100}$$
$$+ d_4(1 - y_4) \tag{101}$$

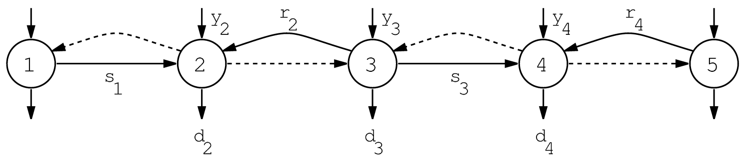

Fig. 10. A BLS example ($n \geq 5$)

Again, the inequalities illustrated above do not suffice to obtain a linear description of the convex hull of solutions to BLS (such a description is unknown for general objective functions), but they suffice to solve most practical instances of BLS without branching (see Section 3.5).

Wagner-Whitin Costs. We analyze in this Section the special case of the previously studied lot-sizing problems where the production, inventory and backlogging costs satisfy the Wagner-Whitin condition. An instance of ULS, ULSS, CCLS or BLS is said to have Wagner-Whitin costs if $p_t + h_t \geq p_{t+1}$ and $p_{t+1} + g_t \geq p_t$, for all t. This means that the unit production, unit inventory and unit backlogging costs are non speculative in the sense that,

- if a demand d_t is satisfied from production in periods t or before, then this production occurs as late before t as possible (with respect to setup decisions and capacity available) because producing and stocking in earlier periods costs more than producing later ($p_t + h_t \geq p_{t+1}$),

– if a demand d_t is satisfied from production in periods t or after, then this production occurs as early after t as possible (with respect to setup decisions and capacity available) because producing and backlogging from later periods costs more than producing earlier ($p_{t+1} + g_t \geq p_t$).

The implication of this non speculative cost condition is that

– the complexity bound for the optimization of the subproblems ULS, ULSS, CCLS and BLS are reduced,
– new and smaller (in number of variables and constraints) linear reformulations can be derived based on the fact that there always exists stock minimal and backlog minimal solutions (i.e. each stock and backlog variable is set to its minimal minimal possible value with respect to setup decisions and capacity constraints) to these problems,
– fewer constraints are needed in the complete linear description of the convex hull of feasible solutions in the initial space, and the complexity bounds of the separation algorithms are reduced.

Federgrun, Tsur [20], Wagelmans, van Hoesel, Kolen [77], and Aggarwal, Park [3]. Furthermore have shown that the dynamic programs solving ULS and ULSS can be implemented in $O(n)$ when the cost function satisfies the Wagner-Whitin assumption.

For the linear reformulation, in the case of ULS, we have seen that inequalities (68) do not suffice to obtain a linear description of $conv(X^{ULS})$, but they suffice to solve ULS by linear programming with a Wagner-Whitin objective function. This means that in this case there always exists an optimal solution with y integer to the linear relaxation of (28)-(32) augmented by (68). Moreover, this Wagner-Whitin condition is a very natural assumption that is almost always satisfied by the objective functions encountered in practice. This is the reason why it generally suffices to add the inequalities (68) to solve most ULS problems by LP.

Similarly, there always exists an optimal solution with y integer to the linear relaxation of problem (73)-(78), plus (79). Therefore, adding inequalities (79) suffices to solve ULSS by LP. Similar results hold for problems CCLS and BLS.

We summarize in Table 2 the results about the linear reformulations of problems ULS, ULSS, CCLS and BLS. See Pochet and Wolsey [57] for a detailed description of the formulations and separation algorithms under the Wagner-Whitin cost assumption. We indicate in the table the sizes and complexity bounds for general cost functions, and the size/complexity reduction obtained, if any, with Wagner-Whitin costs (after the $\rightarrow_{WW}$ sign). In the table, ?? indicates that no result is known, to the best of our knowledge. These results are extended to the uncapacitated lot-sizing problem with backlogging and startup costs in Agra and Constantino [4].

Table 2. Size/complexity reductions with Wagner-Whitin costs

	Extended linear formulation	Initial space linear formulation
Uncapacitated (ULS)		
number of variables	$O(n^2)$	$O(n)$
number of constraints	$O(n^2)$	$O(2^n) \rightarrow_{WW} O(n^2)$
separation complexity		$O(n^2)$
Startup costs (ULSS)		
number of variables	$O(n^2)$	$O(n)$
number of constraints	$O(n^2)$	$O(2^n) \rightarrow_{WW} O(n^2)$
separation complexity		$O(n^3) \rightarrow_{WW} O(n^2)$
Constant capacity (CCLS)		
number of variables	$O(n^3) \rightarrow_{WW} O(n^2)$	$O(n)$
number of constraints	$O(n^3) \rightarrow_{WW} O(n^2)$	?? $\rightarrow_{WW} O(2^n)$
separation complexity		?? $\rightarrow_{WW} O(n^2 \, log \, n)$
Backlogging (BLS)		
number of variables	$O(n^2) \rightarrow_{WW} O(n)$	$O(n)$
number of constraints	$O(n^2)$	?? $\rightarrow_{WW} O(2^n)$
separation complexity		?? $\rightarrow_{WW} O(n^3)$

Profit Maximization. As a last variant of the ULS problem, we consider the profit maximization problem PLS formulated as

$$\max \sum_t \pi_t v_t - \sum_t \{p_t x_t + f_t y_t + h_t s_t\} \tag{102}$$

$$s_{t-1} + x_t = v_t + s_t \quad \text{for all } t \tag{103}$$

$$s_0 = s_n = 0 \tag{104}$$

$$x_t \leq M y_t \quad \text{for all } t \tag{105}$$

$$0 \leq v_t \leq U_t \quad \text{for all } t \tag{106}$$

$$x_t, s_t \geq 0, \; y_t \in \{0,1\} \quad \text{for all } t \tag{107}$$

where the variable v_t, representing the amount sold in period t, replaces the fixed demand in the flow balance constraint (103), and is bounded from above by the maximum forecasted market potential U_t in constraint (106). The objective function (102) represents the profit contribution to be maximized and defined as the product of sales at a price π_t per unit sold in period t, minus production and inventory costs.

Such a problem arises typically as a subproblem in profit maximization multi item production planning problems, where there are capacity linking constraints between the items. For each single item subproblem, the production costs in each period contain implicitely, or even explicitely when the problem is solved by decomposition or Lagrangean relaxation, the opportunity cost of the capacity used for production. The problem consists in allocating the capacity to the items in order to maximize overall profit contribution.

Loparic, Pochet and Wolsey [38] analyze the optimization algorithms and linear formulations for problem PLS with additional lower bounds on stocks to model required safety stock levels. They show that the problem can be solved in $O(n^4)$ by dynamic programming, they propose linear formulations with $O(n^2)$ variables and constraints, and they provide a complete linear description with an exponential $O(2^n)$ number of constraints of the convex hull of solutions in the initial variable space. Again the problem can be solved by LP with only a subset of these constraints – but still exponentially many – when the cost function satisfies the Wagner-Whitin assumption.

We only illustrate here the type of stock minimal inequalities needed to solve PLS by LP in the presence of such Wagner-Whitin costs. The inequality

$$s_1 \geq (v_2 - U_2 y_2) \ + \ (v_4 - U_4(y_2 + y_3 + y_4)) \tag{108}$$

is valid for the convex hull of solutions to (103)-(107) and it is a slight generalisation of the inequality

$$s_1 \geq (d_2 - d_2 y_2) \ + \ (d_4 - d_4(y_2 + y_3 + y_4)) \tag{109}$$

needed to describe $conv(X^{ULS})$. The inequality (108) expresses that if $y_2 = 0$, then $x_2 = 0$ and s_1 must cover the sales level v_2, and if $y_2 = 1$ then there is no restriction on s_1 because $v_2 \leq U_2$ can be covered by x_2. Therefore, $s_1 \geq (v_2 - U_2 y_2)$ is valid. Similarly, $s_1 \geq v_4$ when there is no production in periods 2, 3 and 4, that is when $y_2 + y_3 + y_4 = 0$.

4 Formulations of Capacitated Models

We analyze in this section the main classes of capacitated production planning models. We first classify the models, and then describe in more details the basic mathematical programming models and formulations for the two main model classes, the big buckets and small buckets models.

4.1 Classes of Capacitated Models

The first and important shortcoming of the classical MRP approach in discrete manufacturing is to decompose the planning procedure in two phases. First the uncapacitated planning problem is solved, to obtain production or purchase plans for all items in the product structure. Then the capacity problems (overloads) created by the production plans are resolved by smoothing the load profiles, i.e. by shifting some of the production lots forward or backward in time. This lack of integration between the uncapacitated planning phase and the capacitated post-optimization phase explains that there is no guarantee to obtain a feasible production plan if one exists, and explains why the production plans obtained are suboptimal in terms of cost. There is thus a need to develop capacitated models allowing one to obtain feasible and optimized production plans, and providing the needed flexibility and tools to adapt the production plans to demand and to market requirements.

There are two main classes of capacitated models that can be distinguished based on the size, or time length, of the production periods. These model classes are known as the big buckets and small buckets models, and correspond to quite different planning environments and problems that we describe now.

Discrete time planning models are models where the planning horizon is divided into discrete time periods, and where all events – production, sales, ... – occuring during a period are aggregated in time as if they occur all at the same time. All the models that we have studied so far are of this type.

Big buckets planning models are discrete time models with large time periods where the typical aim is to take global decisions about the resources to be used and to acquire, and about the assignment of products to facilities or departments, rather than to plan the production to meet some precise and firm short term customer orders, or to control the short term output from the shop floor. Therefore, there is usually no need for a detailed representation of time. Moreover, and very often, a more detailed time representation would not make sense because the required detailed information (e.g. detailed sales forecasts) would not be available on time.

Such models are used for solving tactical medium term resource planning problems. The size of the time period is usually taken as one week, or one month, depending on the particular problem, and the planning horizon ranges from a few weeks to months, up to one year.

Small buckets planning models are discrete time models with small time periods where the aim is to take detailed decisions about the materials flow in the shop floor, the sequence of production lots on machines to control short term output and capacity usage, to verify that technical precedence constraints between lots can be satisfied. A detailed time representation is needed to represent accurately these events.

Such models are used for solving operational short term materials planning problems. The size of the time period is usually taken as one week, or one day, down to one hour, depending on the particular problem, and the planning horizon ranges from a few weeks down to one day. In some cases, similar models have to be used for longer term problems with larger time periods when, for instance, the detailed sequence of products has a significant impact on the medium term capacity utilization. This is for example often the case in the process or chemical industry where change-over times may represent days of production lost.

For simplicity, we will consider that big buckets models are discrete time models where there is no need to represent the sequence of events inside a period, and between the different periods. Small buckets models are models where it is required to represent the sequence of events from one period to the other.

4.2 Big Buckets Models: Capacitated Lot-Sizing

Basic Big Buckets Multi-item Model. The master production scheduling model (6)-(10) formulated in Section 2.2, also known as the capacitated multi-item capacitated lot-sizing problem, is the basic example of big bucket model. We formulate here the simplified version without setup times, called MICLS, that has been widely studied in the literature.

$$\min \quad \sum_i \sum_t \{p_t^i x_t^i + f_t^i y_t^i + h_t^i s_t^i\} \tag{110}$$

$$s_{t-1}^i + x_t^i = d_t^i + s_t^i \quad \text{for all } i, t \tag{111}$$

$$x_t^i \leq C_t^i y_t^i \quad \text{for all } i, t \tag{112}$$

$$\sum_i \alpha^i x_t^i \leq L_t \quad \text{for all } t \tag{113}$$

$$x_t^i, s_t^i \geq 0, \; y_t^i \in \{0, 1\} \quad \text{for all } i, t \tag{114}$$

where constraint (112) represents a single item capacity constraint and constraint (113) is the multi-item capacity linking constraint. Note that if $C_t^i = L_t/\alpha^i$, then constraint (112) is just used to define the setup variable y_t^i without imposing another capacity constraint than the linking constraint (113).

Several optimization approaches have been proposed and tested for problem MICLS. They are based on the single item uncapacitated ULS subproblems, and use the optimization and reformulation results surveyed in Section 3. Thizy and Van Wassenhove [67], Chen and Thizy [10] test and compare several Lagrangean relaxation schemes, Eppen and Martin [19] solve the problem by branch and bound using extendend linear formulations for the ULS subproblems, Barany, Van Roy and Wolsey [7] and Pochet, Wolsey [55] use a cut and branch approach (cuts are added only before the branch and bound enumeration), Constantino [13] solve MICLS by a branch and cut approach.

Valid Inequalities Using Known Cutting Planes. As we did in Section 3.4, we illustrate how to identify classes of valid inequalities for problem MICLS using known cutting planes. This approach can be used to tighten the formulation and solve such problems using branch and cut approaches.

The first idea is to use the mixed integer rounding (MIR) inequalities from Nemhauser and Wolsey [48] [$\longrightarrow$ Martin]. Marchand and Wolsey [43] show how to generate MIR inequalities from 0-1 knapsack problems with continuous variables. To use this idea, it suffices to construct relaxations of MICLS in the form of 0-1 knapsacks with continuous variables.

We illustrate this approach with an example. First we construct a single item knapsack relaxation for item i by summing the flow conservation constraints (111) over periods $t = k, \cdots, l$, replacing s_l^i by its lower bound 0, and replacing x_t^i by its upper bound $\min[C_t^i, d_{tl}^i]y_t^i$, for $t = k, \cdots, l$. This gives the following knapsack relaxation with one continuous variable.

$$s_{k-1}^i \; + \; \sum_{t=k}^{l} \min[C_t^i, d_{tl}^i] y_t^i \; \geq \; d_{kl}^i \tag{115}$$

$$s_{k-1}^i \; \geq 0, \; y_t^i \in \{0,1\} \quad \text{for} \;\; t = k, \cdots, l \tag{116}$$

where as before we use the notation $d_{tl}^i = \sum_{j=t}^{l} d_j^i$.

Suppose now that we have solved the linear relaxation of MICLS and we want to cut off the fractional solution obtained and illustrated in Figure 11 for one item i and for periods 2 up to 5.

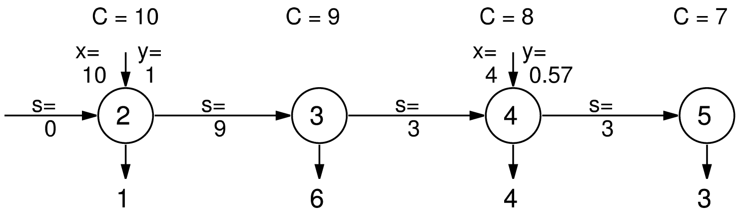

Fig. 11. A fractional solution for one item in MICLS

The corresponding knapsack relaxation is (where we have dropped the item i superscript for simplicity)

$$s_1 \; + 10y_2 + 9y_3 + 7y_4 + 3y_5 \geq \; 14, \tag{117}$$

which can be relaxed into (where we consider variable y_5 as continuous)

$$(s_1 + 3y_5) \; + 10(y_2 + y_3 + y_4) \geq \; 14. \tag{118}$$

We obtain the classical MIR inequality

$$\frac{s_1}{4} + \frac{3}{4} y_5 \; + y_2 + y_3 + y_4 \; \geq \; 2 \tag{119}$$

by dividing (118) by 10 and rounding. The inequality (119) is violated by the fractional point in Figure 11.

Another idea is to use the well known flow cover inequalities that have been defined for single node flow models in Padberg, Van Roy and Wolsey [50]

[$\longrightarrow$ Martin] for a general description of flow cover inequalities). Again, it suffices to construct single node flow relaxations, and to generate the corresponding flow cover inequalities.

Consider first the single node flow relaxation defined for a single item i over several consecutive periods, and obtained by summing the flow conservation constraints (111) over periods $t = k, \cdots, l$, replacing s^i_{k-1} by its lower bound 0. This gives the following single node flow model.

$$\sum_{t=k}^{l} x^i_t \;\leq\; d^i_{kl} \;+\; s^i_l \tag{120}$$

$$x^i_t \leq \min[C^i_t, d^i_{tl}]y^i_t + s^i_l \quad \text{for} \quad t = k, \cdots, l \tag{121}$$

$$s^i_l \;\geq 0,\; x^i_t \geq 0,\; y^i_t \in \{0,1\} \quad \text{for} \quad t = k, \cdots, l \tag{122}$$

where we have changed the variable upper bound constraint in (121) by adding the flow s^i_l at the right hand side, allowing us to reduce the coefficient of y^i_t.

Again, suppose that we are given the fractional solution illustrated in Figure 11. The corresponding single node flow relaxation is

$$x_2 + x_3 + x_4 + x_5 \leq 14 \tag{123}$$

$$x_2 \leq 10y_2 + s_5,\; x_3 \leq 9y_3 + s_5,\; x_4 \leq 7y_4 + s_5,\; x_5 \leq 3y_5 + s_5 \tag{124}$$

By projecting s_5 to zero, we obtain the classical flow cover inequality

$$x_2 + x_4 \leq 14 \;- (10-3)(1-y_2) - (7-3)(1-y_4) \tag{125}$$

using the cover $C = \{2,4\}$ and the excess capacity of the cover $\lambda = 3$. We obtain finally the valid inequality

$$x_2 + x_4 \leq 14 \;- (10-3)(1-y_2) - (7-3)(1-y_4) \;+ s_5 \tag{126}$$

by adding the flow s_5 at the right hand side of the inequality (125) to lift back the variable s_5, i.e. to make it valid even when $s_5 > 0$. The inequality (125) is violated by the fractional point in Figure 11.

Consider next another single node flow relaxation obtained by taking the linking capacity constraint (113) for a given period t, and the single item capacity constraints (113) for all items in the same period t. This gives the following single node flow relaxation

$$\sum_i (\alpha^i x^i_t) \leq L_t \tag{127}$$

$$(\alpha^i x^i_t) \leq (\alpha^i C^i_t) y^i_t \quad \text{for all } i \tag{128}$$

$$(\alpha^i x^i_t) \geq 0,\; y^i_t \in \{0,1\} \quad \text{for all } i \tag{129}$$

where the continuous flow variables are the variables $(\alpha^i x^i_t)$ with an indivudual upper bound of $(\alpha^i C^i_t)$.

Suppose now that we have solved the linear relaxation of MICLS and we want to cut off the fractional solution obtained and illustrated in Figure 12 for all items and for a given single period.

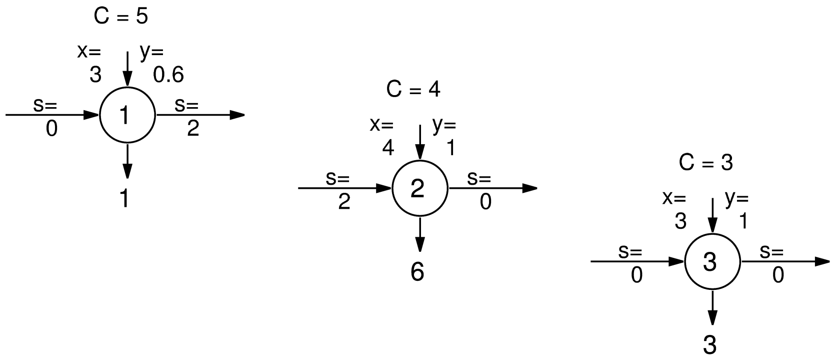

Fig. 12. A fractional solution for one period, 3 items in MICLS

The corresponding single node flow model is (where we have dropped the time index t for simplicity)

$$x^1 + 2x^2 + 3x^3 \leq 20 \tag{130}$$
$$x^1 \leq 5y^1, \ 2x^2 \leq 2(4y^2), \ 3x^3 \leq 3(3y^3). \tag{131}$$

We obtain the classical flow cover inequality

$$\begin{aligned} x^1 + 2x^2 + 3x^3 \leq 20 \ & -(5-2)(1-y^1) \\ & -(8-2)(1-y^2) - (9-2)(1-y^3) \end{aligned} \tag{132}$$

using the cover $C = \{1, 2, 3\}$ and the excess capacity of the cover $\lambda = 2$. This inequality is violated by the fractional point in Figure 12.

Big Buckets Multi-item Model with Setup Times. The classical and most studied MICLS problem as formulated in (110)-(114) does not contain any setup time, but setups are only taken into account through a penalty cost in the objective (110).

Therefore, this setup cost f^i_t must contain the opportunity cost of the setup activity for item i in period t. The opportunity cost is defined as the additional or marginal cost incurred because of the setup activity. It is very difficult to estimate a priori because it really depends on problem data, on capacity utilization and capacity available and productivity in period t. Hence, because of the difficulty to estimate the setup cost , the true optimality of the solution to MICLS in not guaranteed.

Moreover, the setup activity consumes part of the capacity available, and influences the productivity. Therefore, it is not even guaranteed that the solution obtained from MICLS is really feasible because the capacity consumed by the setup activity is neglected.

In order to improve the reliability of the solution from MICLS, both in terms of feasibility and optimality, it is important to model more accurately the productivity, i.e. the capacity used as a function of the production lot sizes. In the presence of economies of scale in production, a better model is to introduce directly in the model the capacity used by the setup activity – and not only implicitely through the setup cost –. The linking capacity constraint in period t should then be modelled as

$$\sum_i (\alpha^i x_t^i) \ + \ \sum_i (\beta^i y_t^i) \leq L_t \tag{133}$$

where β^i represents the capacity used by the setup of item i, and is often called the setup time of item i. We call MICLSS the multi item model MICLS where the capacity constraint with setup times (133) replaces the original capacity constraint (113).

Very few optimzation methods have been proposed and tested for MICLSS. Trigeiro, Thomas, McClain [68] and Diaby, Bahl, Karwan, Zionts [17] test Lagrangean relaxation heuristic approaches – with performance guarantee – based on the solution of the ULS subproblems and the relaxation of the capacity linking constraints (133).

To tighten the basic formulation of MICLSS beyond the complete description of the ULS subproblems, other relaxations or subproblems have to be studied. Goemans [24] proposes generalized flow cover valid inequalities to tighten the formulation of the single period t multi item relaxation defined by

$$\sum_i (\alpha^i x_t^i) \ + \ \sum_i (\beta^i y_t^i) \leq L_t \tag{134}$$

$$0 \leq x_t^i \leq M y_t^i \quad \text{for all } i \tag{135}$$

$$y_t^i \in \{0, 1\} \quad \text{for all } i \tag{136}$$

Constantino [13] provides a complete linear description of this single period relaxation when the setup times are constant for all items, i.e. $\beta^i = \beta$ for all i. These tightened formulations have been tested on multi-item single-period subproblems, but no tests of these improved formulations on multi-item problems with setup times (MICLSS) have been reported on in the literature.

Another single period relaxation of MICLSS that has been studied tries to combine the linking capacity constraints and the demand satisfaction for each item. It is formulated as

$$\sum_i(\alpha^i x_t^i) \ + \ \sum_i(\beta^i y_t^i) \leq L_t \tag{137}$$

$$s_{t-1}^i + x_t^i \geq d_t^i \quad \text{for all } i \tag{138}$$

$$0 \leq x_t^i \leq M y_t^i \quad \text{for all } i \tag{139}$$

$$s_{t-1}^i \geq 0,\ y_t^i \in \{0,1\} \quad \text{for all } i \tag{140}$$

Miller, Nemhauser and Savelsbergh [46] define cover and reverse cover valid inequalities for the one period subproblem (137)-(140), and tests their effectiveness in solving MICLSS instances by branch and cut.

4.3 Small Buckets Models

Small buckets models are used when changing the setup of the machines has a significant impact on the capacity used and/or on the production costs. Typically in these models, on each machine, we keep on producing the same item for a number of periods, and we only incur a setup cost/time when the setup of the machine is changed from one item to another.

When the capacity used or the cost incurred during the setup modification do not depend on the sequence of items, but on the fact that the machine setup is changed, we speak about startup time or startup cost models. When the cost or time lost during the setup depend on the sequence of products, i.e. on the products produced before and after the setup modification, we speak about change over costs and times.

We describe here the basic models, formulations and applications for such small buckets planning problems with startup costs and times, or change over costs and times.

Basic Small Buckets Multi-item Model, One Setup per Period. The natural and classical single machine formulation to represent the setup, startup and change over decisions in a small buckets planning model allows one to setup only one item per period. If the setup of the machine is changed from period $t-1$ to period t, this change is supposed to occur at the beginning of period t. This means that only one item can be produced in any period, the one for which the machine is setup in the period.

The variable y_t^i is used to represent the setup decision, it takes the value 1 when the machine is setup to produce item i during the whole period t, and 0 otherwise.

The variable z_t^i is the startup variable and takes the value 1 when the machine starts producing item i in period t, and was not setup for item i in period $t-1$, and 0 otherwise.

The variable w_t^{ij} is the change over variable and takes the value 1 when the machine strats producing item j in period t and was setup for item i in period $t-1$, and 0 otherwise.

Independently on the production levels and the demand satisfaction part of the problem, the natural formulation used to represent these setup, startup and change over decisions is the following

$$\sum_i y_t^i = 1 \quad \text{for all } t \tag{141}$$

$$z_t^i \geq y_t^i - y_{t-1}^i \quad \text{for all } i, t \tag{142}$$

$$w_t^{ij} \geq y_{t-1}^i + y_t^j - 1 \quad \text{for all } i, j, t \tag{143}$$

$$z_t^i,\ w_t^{ij},\ y_t^i \in \{0, 1\} \quad \text{for all } i, j, t \tag{144}$$

where constraint (141) imposes to setup the machine for exactly one item per period, constraint (142) forces the startup of item i in period t when $y_t^i = 1$ and $y_{t-1}^i = 0$, constraint (143) forces the change over from item i in period $t-1$ to item j in period t when $y_{t-1}^i = y_t^j = 1$.

To obtain a complete formulation of the small buckets planning problem, classical flow conservation constraints are used to model the satisfaction of demand, as well as capacity constraints. In a model with startup time γ^i to start producing item i, the production capacity constraints are formulated as

$$x_t^i \leq U_t^i y_t^i \ - \gamma^i z_t^i \quad \text{for all } i, t. \tag{145}$$

In a model with change over time γ^{ji} for switching from j to i, the production capacity constraint are formulated as

$$x_t^i \leq U_t^i y_t^i \ - \sum_{j:j \neq i} \gamma^{ji} w_t^{ji} \quad \text{for all } i, t. \tag{146}$$

Several optimization approaches have been proposed in the literature for solving small buckets constant capacity over time (but different for each item) multi item lot sizing problems. Karmarkar and Scharge [31] propose, and test on small instances, a Lagrangean relaxation and branch and bound approach for the model with startup costs. Wolsey [81] solve problems with change over costs by branch and bound using the extended unit flow reformulation (see below) to model the setup decisions. Constantino [12] solve larger instances, with up to 5 items and 36 periods, of problems with startup costs using a branch and cut approach. Vanderbeck [69] solve problems with startup times using a column generation approach combined with cut generation.

Unit Flow Reformulation. The formulation (141)-(144) used to model the steup decisions can be tightened and reformulated as a network flow problem consisting of sending one unit of flow from period 1 to period n. The flow models the status of the machine through the time priods. This formulation is the tightest possible to represent the machine status decisions in the sense that the constraint matrix is totally unimodular. This formulation proposed in Karmarkar and Schrage [31] has been tested in Wolsey [81]. It is defined by

$$\sum_i y_0^i = 1 \tag{147}$$

$$\sum_j w_t^{ij} = y_{t-1}^i \quad \text{for all } i, t \tag{148}$$

$$z_t^j + w_t^{jj} = y_t^j \quad \text{for all } j, t \tag{149}$$

$$\sum_{i:i \neq j} w_t^{ij} - z_t^j = 0 \quad \text{for all } j, t \tag{150}$$

$$z_t^j,\ w_t^{jj},\ y_t^j\ \in \{0, 1\} \quad \text{for all } i, j, t \tag{151}$$

where constraint (147) defines the initial (period 0) status of the machine – usually these variables have a fixed value corresponding to the current status of the machine–. Constraint (148) stipulates that if the machine was setup for i in period $t-1$, then it must switch from i to some j (possibly equal to i) in period t. Constraint (149) expresses that if the machine is setup for j in period t, then either it was already setup for j in period $t-1$, or it starts producing item j in period t. Constraint (150) says that if the machine starts producing item j in period t, it must switch from a different product in period $t-1$.

Alternative formulations for modelling changeovers in production planning and scheduling problems are surveyed in Wolsey [82].

Models with Two Setups per Period. Haase [27] has proposed a more flexible interpretation of the setup decisions. Suppose that $y_t^i = 1$ if the machine is setup for item i at the end of period t, and 0 otherwise. The natural formulation (141)-(144), or the unit flow formulation (147)-(151), can still be used to model the sequence of admissible setup decisions. But with this interpretation of the setup variables, it is now possible to produce two items in each period t. These are the items for which the machine is setup at the end of period $t-1$ and t. If any, the setup modification occurs now during period t between the production lots of the two products. So we only need to reformulate the capacity constraints. For instance, in the case of startup times, this gives

$$x_t^i \leq U_t^i(y_{t-1}^i + y_t^i - w_t^{ii})\ - \gamma^i z_t^i \quad \text{for all } i, t \tag{152}$$

$$\sum_i (\alpha^i x_t^i)\ +\ \sum_i (\gamma^i z_t^i) \leq L_t \tag{153}$$

Kimms [32] solves multi item small buckets problems with two setups per period using a Lagrangean relaxtion approach.

Sequencing Models or m-Setups per Period. As a final extension of the small buckets models, we can mention the pure batch sequencing problem, or problems with many setups per period. In this case, we still need to know the sequence of products in order to compute the change over times, but we may produce in each period as many items or batches as feasible given the available capacity.

The natural formulation (141)-(144), or the unit flow formulation (147)-(151), can still be used to model the sequence of admissible setup decisions inside each period, but with a different interpretation of the index t as the batch number rather than the period number. For instance, $y_t^i = 1$ means that item i is the item produced during the t^{th} batch.

An alternative formulation in this case is to use the asymmetric traveling salesman problem (ATSP) to formulate the sequence of setup decisions inside each period.

Several papers in the literature report on such applications and propose optimization approaches for similar problems. Kang, Malik and Thomas [30] propose and test a column generation approach to solve a multi period small buckets sequencing problem with change over costs, and compare with a direct branch and bound approach using the ATSP formulation for the setup decisions. Batta and Teghem [8] solve a multi item single period problem with change over times by branch and bound, again using an ATSP formulation.

5 Formulations of Multi-level Models

We study in this section a general class of multi level production planning models. First we recall the description of the multi level lot sizing problem from Section 2.2, and analyze the classical level by level decompostion approach to show that it is suboptimal. Then we survey the optimzation approaches from the literature. Finally we present the echelon stock reformulation that plays an important role in all optimization approaches published so far.

5.1 Multi-level Models and the Decomposition Approach

We recall first the formulation of the general product structure capacitated multi-level lot-sizing model from Section 2.2 (see for instance Billington, McClain and Thomas [9]). The product structures are classified into Series, Assembly or General structures, as represented in Figure 4.

$$\min \quad \sum_i \sum_t \{p_t^i x_t^i + f_t^i y_t^i + h_t^i s_t^i\} \tag{154}$$

$$s_{t-1}^i + x_{t-\gamma^i}^i = [d_t^i + \sum_{j \in S(i)} r^{ij} x_t^j] + s_t^i \ \text{ for all } i, t \tag{155}$$

$$x_t^i \le M y_t^i \ \text{ for all } i, t \tag{156}$$

$$\sum_i \alpha^{ik} x_t^i + \sum_i \beta^{ik} y_t^i \le L_t^k \quad \text{for all } t, k \tag{157}$$

$$x_t^i,\ s_t^i \ge 0,\ y_t^i \in \{0, 1\} \ \text{ for all } i, t \tag{158}$$

A classical approach used for solving multi level planning problems is to decompose the problem level by level into single level subproblems, solved sequentially from end-products to raw materials. The production plans at each level define the demand at subsequent levels. This is the approach used in MRP-type systems. This separate optimization at each level yields suboptimal production plans. This sequential planning procedure, and its suboptimality, are illustrated in the following simple example.

Suppose that we have a 3 periods lot-sizing problem with 2 levels, one item at each level, and a serial product structure where one unit of the raw material ($i = 2$) is required to produce one unit of the finished product($i = 1$). For simplicity, we assume that the lead time γ^i in zero for each item. The external or independent demand for the finished product is $d^1 = (10, 15, 20)$. There is no external demand for the raw material. There is a fixed ordering cost of $f_t^2 = 200$ for the raw material, and fixed production cost of $f_t^1 = 100$ for the finished product, for all t. There is no unit production cost. The inventory cots is $h_t^i = 5$, for all i, t. This planning problems corresponds to the fixed charge minimum cost network flow problem represented in Figure 13.

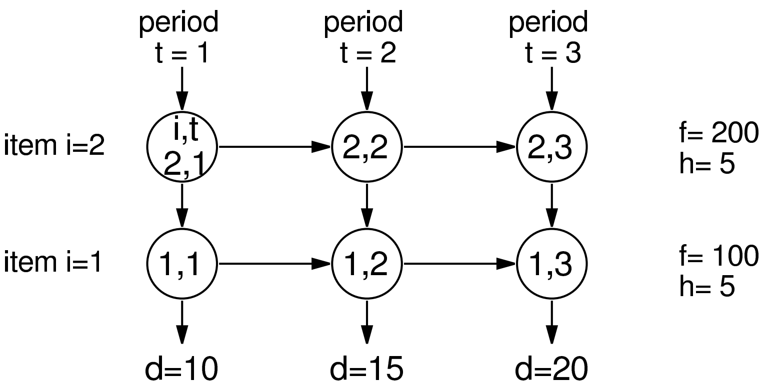

Fig. 13. A 2-level serial production planning example

In the classical decomposition approach, we plan first the production of the finished product in order to satisfy its external demand, by solving the corresponding single item ULS problem. The optimal solution with a cost of 275 is to produce 25 units in period 1, stock 15 from period 1 to period 2, and produce 20 units in period 3.

This production plan defines the internal or dependent demand for the raw material, 25 units of raw material have to be available in period 1, and 20 units have to be available in period 3. We then solve the ULS subproblem for the raw material, and find that it is optimal at a cost of 400 to order 45 units in period 1, and stock 20 units from period 1 to period 3. Note that an alternate optimal solution for the raw material is to order twice (25 units in period 1 and 20 units in period 3), and avoid the inventory costs.

So, we have obtained a global (finished product and raw material) production plan with a cost of 675. The optimal solution with a cost of 575 of this network flow problem is to order 45 units of raw material in period 1, transform these units directly into finished product in period 1, and stock the finished product to satisfy demand in periods 2 and 3. Furthermore, at the first step of the sequential procedure, if you enumerate all extreme solutions (those satisfying satisfying the decomposition property in Theorem 1) of the ULS subproblem for the finished product, then the worst extreme solution (which is to produce the 45 units in period 1, and to stock to periods 2 and 3) of this subproblem should be selected in order to find the globally optimal solution.

We have illustrated with this naive example, that it seems very difficult to optimize the production plans by solving the single level subproblems sequentially. Therefore, we need to solve integrated models, as (154)-(158), taking the planning decisions at all levels simultaneously, in order to optimize the quality of the production plans.

5.2 Optimization of Multi-level Models

We review briefly the complexity results known, and the optimization approaches used to solve various multi level planning problems.

The uncapacitated problem (without constraint (157)) with a serial product structure (see Figure 4) is the simplest problem in the class. It can be solved in polynomial time by dynamic programming (Zangwill [85] and Love [41]) using the same decomposition property as for ULS, and the technique of Martin [44] can be used to transform this dynamic program into a compact (i.e. polynomial number of variables and constraints, in the number of items and time periods) linear extended reformulation of the problem (see Pochet [52]). The solution of this problem using a branch and cut approach, and a partial description of the convex hull of solutions, is presented in Pochet [52].

The next problem studied in the literature is the uncapacitated problem with an assembly product structure (see Figure 4). The computational complexity of this problem is still an open question. Afentakis, Gavish and Karmarkar [1] have solved problems with up to 50 items using a Lagrangean relaxation approach based on the echelon stock reformulation and combined with a specialized branch and bound algorithm.

The uncapacitated problem with general product structure has been tackled by Afentakis and Gavish [2] using a Lagrangean relaxation approach, and by Pochet and Wolsey [55] using a cut and branch approach (i.e. cuts are added only at the first node of the branch and bound tree, before any branching).

The full problem (154)-(158) with general product structure, capacities and setup times, has been addressed by Tempelmeier and Derstroff [66] to obtain heuristic solutions with performance guarantee using a Lagrangean relaxation approach, and by Stadtler [65] using a branch and bound approach and the extended linear reformulations for the ULS subproblems.

All the optimization applications mentioned for assembly and general product structures, using Lagrangean relaxation, branch and bound or branch and cut approaches, are based on the same reformulation of the problem. First, problem (154)-(158) is reformulated using the echelon stock concept. This transforms the initial formulation into a series of ULS subproblems linked by capacity constraints and by product structure constraints (see Section 5.3 below). Then, the ULS part of the problem is reformulated in the branch and bound or branch and cut approaches, or the linking constraints are relaxed in the Lagrangean relaxation approach.

Better relaxations are available, but not yet used for solving complex capacitated multi-level problems. For instance, the reformulations and valid inequalities found for big buckets capacitated models with setup times could be used to tighten the formulations further. Valid inequalities for fixed charge network flow problems could be generated on the multi-level network flow problem (see Van Roy and Wolsey [74]).

5.3 The Echelon Stock Reformulation

The formulation (154)-(158) of the general multi-level planning problem does not contain explicitely the single item ULS subproblems (except for finished products) because of the presence of both dependent and independent demand for the items. By using the concept of echelon stock, introduced by Clark and Scarf [11], one can obtain a reformulation that contains explicitely the ULS subproblems. This allows one then to design decomposition optimization algorithms based on our knowledge on how to solve and reformulate ULS.

For simplicity, we assume that the lead times γ^i are equal to zero in (155). The echelon stock E^i_t of item i in period t is the total stock of component i in the production system, as item i or in items appearing further on in the production process. It can be defined as

$$E^i_t = s^i_t + \sum_{j \in S(i)} r^{ij} E^j_t \tag{159}$$

because, in addition to the stock s_t^i of items i as items i, there are r^{ij} units of items i present in each unit of stock of item j.

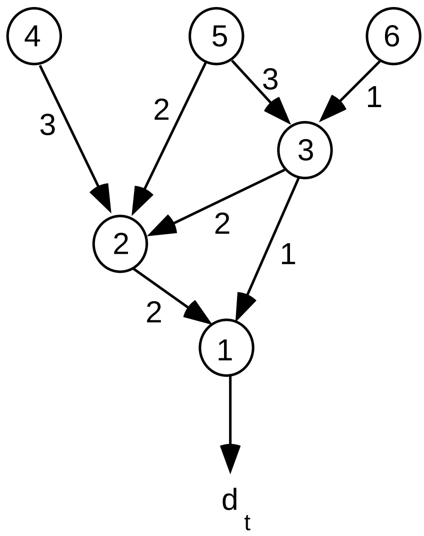

Fig. 14. A general product structure

Figure 14 gives an example of general product structure. For this example, the echelon stocks of the different items are

$$E_t^1 = s_t^1 \tag{160}$$
$$E_t^2 = s_t^2 + 2E_t^1 = s_t^2 + 2s_t^1 \tag{161}$$
$$E_t^3 = s_t^3 + 2E_t^2 + E_t^1 = s_t^3 + 2s_t^2 + 5s_t^1 \tag{162}$$
$$E_t^4 = s_t^4 + 3E_t^2 = s_t^4 + 3s_t^2 + 6s_t^1 \tag{163}$$
$$E_t^5 = s_t^5 + 2E_t^2 + 3E_t^3 = s_t^5 + 3s_t^3 + 8s_t^2 + 19s_t^1 \tag{164}$$
$$E_t^6 = s_t^6 + E_t^3 = s_t^6 + s_t^3 + 2s_t^2 + 5s_t^1 \tag{165}$$

We can use equation (159) to define k^{ij} as the number of items i present in each unit of item j, for all i, j (i.e. $E_t^i = \sum_j k^{ij} s_t^j$), and to replace the natural stock variables s_t^i by the echelon stock variables E_t^i in formulation (154)-(158). This gives the formulation

$$E_{t-1}^i + x_t^i = \sum_j k^{ij} d_t^j + E_t^i \text{ for all } i, t \tag{166}$$

$$x_t^i \leq M y_t^i \;\text{ for all } i,t \tag{167}$$

$$E_t^i - \sum_{j \in S(i)} r^{ij} E_t^j \geq 0 \;\text{ for all } i,t \tag{168}$$

$$\sum_i \alpha^{ik} x_t^i + \sum_i \beta^{ik} y_t^i \leq L_t^k \quad \text{for all } t,k \tag{169}$$

$$x_t^i,\; E_t^i \geq 0,\; y_t^i \in \{0,1\} \;\text{ for all } i,t \tag{170}$$

where (168) is the product structure linking constraint coming from the non negativity of the natural stock variables, (169) is the capacity linking constraint, (166)-(167) define independent ULS subproblems, one for each item.

6 New Directions

We conclude in this Section by giving a personal view on some new directions to be investigated in modeling production planning problems. These include better productivity models, and in particular better models for capacity utilization and setup times, new models to represent the product structure – or recipes – in process industries, and the study of continuous time planning and scheduling models as opposed to the discrete time models studied in this review. As pointed out in Kuik et al [35], such extensions are crucial for the vitality of the production planning and lot-sizing research.

We also define some challenges for the future of this research field.

6.1 Setup Times, Capacity Utilization, and Productivity Models

We have already mentioned in Section 4.2 that the optimal solutions of multi-item capacitated lot-sizing models (MICLS) with setup cost may well correspond to infeasible and/or suboptimal in reality. The main difficulty there was to estimate a priori the setup cost as the opportunity cost of the capacity used during setup. This was the main reason to study similar capacitated models, but with setup times instead of setup costs, using constraint 133 to model capacity utilization.

For such capacitated models with setup times, a first direction of research is to investigate the polyhedral structure of multi-item subproblems involving both capacity utilization and demand satisfaction, such as Miller [46]. This will allow one to obtain new classes of valid inequalities and to solve problems using better relaxations than the one obtained from the reformulation of the simple uncapacitated subproblems.

Another extension is to develop new capacity utilization or productivity models, and to study the corresponding formulations.

– In classical MRP databases, the routing of an item is defined as a fixed sequence of processing or production steps, each performed in a specific workcenter (i.e. on a specific resource), requiring a fixed setup time and a variable processing time per unit produced. This capacity utilization model can be introduced in the capacitated multi-level lot-sizing formulation (154)-(158) using $\alpha^{ik\Delta}$ and $\beta^{ik\Delta}$ to represent respectively the unit processing time and fixed setup time for item i, on resource k, in the Δ^{th} processing period, for $\Delta = 1, \cdots, \gamma^i$. The capacity constraint (157) becomes then

$$\sum_i \sum_{l=t-\gamma^i+1}^{t} \alpha^{ik,t-l+1} x_l^i + \sum_i \sum_{l=t-\gamma^i+1}^{t} \beta^{ik,t-l+1} y_l^i \leq L_t^k \quad \text{for all } t, k \quad (171)$$

– Product similarities or product families also have some influence on the setup and change over times. On a specific resource or machine, switching from one product family to another requires more time than switching from one product to another product of the same family. Van de Velde [70] suggested new capacity utilization models to better reflect such situations, see Simpson, Erenguc [64].
– In activity based costing and management systems (ABC/ABM), hierarchical models for activities and costs are used to better approximate the (production) cost structure. These can be modelled using joint or hierarchical setup cost models, as in Degraeve and Roodhooft [16].

6.2 Process Industry: Products and Operations Structure

The multi-level MRP-II model, and its typical formulation (154)-(158), should be generalized in order to be applicable to other processes than discrete parts manufacturing. In particular, for chemical or process industries, production planning and sequencing problems require some generalization of BOM structures to allow the modelling of **OPERATIONS** consuming items as **INPUTS** producing items as **OUTPUTS** in fixed or variable proportions and requiring some amount of **RESOURCES**. The main difference with respect to discrete BOMs is that several output items or co-products are generally produced simultaneously in a single operation, see Westenberger, and Kallrath [79], Kondili, Pantelides and Sargent [33].

The basic example –and formulation– of an operations model is

$$s_{t-1}^i + \sum_o \rho_{OUT}^{oi} \, x_{t-\gamma^{oi}}^o = [d_t^i + \sum_o \rho_{IN}^{oi} \, x_t^o] + s_t^i \quad \text{for all } i, t \quad (172)$$

$$LB_t^o y_t^o \leq x_t^o \leq UB_t^o y_t^o \quad \text{for all } o, t \quad (173)$$

$$x_t^o, \; s_t^i \geq 0, \; y_t^o \in \{0, 1\} \quad \text{for all } i, o, t \quad (174)$$

where the index o is used to model operations, x_t^o represents the amount of time spent on performing operation o in period t, and $y_t^o = 1$ indicates that operation

o is carried out in period t. The extended BOM structure is defined by ρ_{OUT}^{oi} and ρ_{IN}^{oi} representing respectively the number of units of item i produced and consumed per time unit of operation o, for all operations o and items i. The data γ^{oi} represents the lead time needed for operation o to produce the output item i.

Many of the reformulation techniques presented in this paper can be used to tighten the formulation of this operations planning model.

6.3 Continuous Time Scheduling Models

In production planning, small buckets models are used when there is a need (cost, capacity, ...) to represent the detailed sequence and timing of events (such as production start, production end, empty or full storage tank, ...) through time. Very often in such models, the period time lengths are very small in order to represent these events at precise moments in time, but the number of events to occur in the planning horizon is small. To avoid to have too many time periods, and too large models, continuous time models are used. They use variable period lengths in order to represent the events in time, see Zhang and Sargent [86], Pinto and Grossman [51], Pochet, Tahmassebi, Wolsey [59].

Consider an example planning problem where a number of items – indexed by i or j – have to be produced on a set of independent parallel production lines – indexed by k. There is a demand of d^i units of item i to meet, and the production rate of item i on machine k is ρ^{ik} units of i per unit production time. When there is a change over on machine k from item i to item j, c^{ijk} units of time are lost, and L^k units of time are available on line k during the planning horizon.

A continuous time formulation for this model is

$$\sum_k \rho^{ik}\,(b^{ik} - a^{ik}) \;=\; d^i \;\text{ for all } i \tag{175}$$

$$\sum_i (b^{ik} - a^{ik}) \;+\; \sum_{i,j} c^{ijk}\, w^{ijk} \le L^k \;\text{ for all } k \tag{176}$$

$$a^{jk} \ge b^{ik} + c^{ijk}\, w^{ijk} \;-\; M(y^{ik} - w^{ijk}) \;\text{ for all } i,j,k \tag{177}$$

$$0 \le (b^{ik} - a^{ik}) \le [d^i/\rho^{ik}]\; y^{ik} \;\text{ for all } i,k \tag{178}$$

$$\sum_j w^{0jk} = 1 \;\text{ for all } k \tag{179}$$

$$\sum_i w^{i0k} = 1 \;\text{ for all } k \tag{180}$$

$$w^{0jk} + \sum_i w^{ijk} = y^{jk} \;\text{ for all } j,k \tag{181}$$

$$w^{i0k} + \sum_j w^{ijk} = y^{ik} \;\text{ for all } i,k \tag{182}$$

$$a^{ik}, b^{ik} \ge 0, \quad y^{ik}, w^{ijk} \in \{0,1\} \;\text{ for all } i,j,k \tag{183}$$

where variable a^{ik} is the start time of item i on line k, variable b^{ik} is the end time of item i on line k, variable y^{ik} takes the value 1 if item i is produced on line k, and variable w^{ijk} equals 1 if there is a change over from item i to item j on line k. Item 0 is a fictive item used to start and finish the sequence on each machine. Constraint (175) imposes demand satisfaction, constraint (176) is the line capacity constraint, constraint (177) imposes the starting time of item j when it follows directly item i on line k, constraint (178) fixes the value of the setup variables, (179)–(182) model the sequencing constraints (i.e. each line must have a starting item, a finishing item, each item must have a predecessor, and a successor).

6.4 Challenges

As a conclusion, we summarize some key challenges for the future of this research field

- The multi-item production planning model (154)-(158) with general product structure, capacities and setup times must be solved in practice. This means that specific optimization codes have to be developed, that heuristic algorithms based on linear programming relaxations must be designed in order to produce good solutions for large instances, and fast and specialized algorithms for solving the large size LP relaxations have to be developed.
- Other more general models have to be constructed in order to come closer to interesting and real planning problems. This includes new capacity utilization models, joint setup times and costs models, operations planning and scheduling models.
- The models have to be changed when the limits of such deterministic discrete time planning problems have been reached. This includes the development and resolution of continuous time scheduling models, as well as of stochastic models.

References

1. P. Afentakis, B. Gavish and U. Karmarkar, “Computationally efficient optimal solutions to the lot-sizing problem in multistage assembly systems”, Management Science 30(2), 222–239, 1984.
2. P. Afentakis and B. Gavish, “Optimal lot-sizing algorithms for complex product structures”, Operations Research 34, 237–249, 1986.
3. A. Aggarwal and J. Park, “Improved algorithms for economic lot-size problems”, Operations Research 41, 549–571, 1993.
4. A. Agra and M. Constantino, “Lotsizing with backlogging and start-ups: the case of Wagner-Whitin costs”, Operations Research Letters 25 (2), 81–88, 1999.
5. R.N. Anthony, “Planning and control systems: a framework for analysis”, Harvard University Press, Cambridge, Mass., 1965.
6. I. Barany, T.J. Van Roy and L.A. Wolsey, “Uncapacitated lot sizing: the convex hull of solutions”, Mathematical Programming Study 22, 32–43, 1984.

7. I. Barany, T.J. Van Roy and L.A. Wolsey, "Strong formulations for multi-item capacitated lot-sizing", Management Science 30, 1255–1261, 1984.
8. C. Batta and J. Teghem, "Optimization of production scheduling in plastics processing industry", Jorbel (Belgian Journal of operations research, statistics and computer science) 34 (2), 55–78, 1994.
9. P.J. Billington, J.O. McClain, L.J. Thomas, "áMathematical Programming approaches to capacity constrained MRP systemsá: review, formulation and problem reductioná", Management Science 29 (10), 1126–1141, 1983.
10. W.-H. Chen and J.-M. Thizy, "Analysis of relaxations for the multi-item capacitated lot-sizing problem", Annals of Operations Research 26, 29–72, 1990.
11. A.J. Clark and H. Scarf, "Optimal policies for multi echelon inventory problems, Management Science 6, 475–490, 1960.
12. M. Constantino, "A cutting plane approach to capacitated lot-sizing with start-up costs", Mathematical Programming 75 (3), 353–376, 1996.
13. M. Constantino, "Lower bounds in lot-sizing models: A polyhedral study", Mathematics of Operations Research 23 (1), 101–118, 1998.
14. W.B. Crowston, M.H. Wagner, "Dynamic lot size models for multi stage assembly systems", Management Science 20(1), 14–21, 1973.
15. W.B. Crowston, M.H. Wagner, J.F. Williams, "Economic lot size determination in multi stage assembly systems", Management Science 19(5), 517–527, 1973.
16. Z. Degraeve, F. Roodhooft, "Improving the efficiency of the purchasing process using total cost of ownership informationá: The case of heating electrodes at Cockerill-Sambre S.A.á", European Journal of Operational Research 112, 42–53, 1999.
17. M. Diaby, H.C. Bahl, M.H. Karwan and S. Zionts, "A Lagrangean relaxation approach to very large scale capacitated lot-sizing", Management Science 38 (9), 1329–1340, 1992.
18. S.E. Elmaghraby, "The economic lot-scheduling problem (ELSP): reviews and extensions", Management Science 24, 587–598, 1978.
19. G.D. Eppen and R.K. Martin, Solving multi-item lot-sizing problems using variable definition, Operations Research 35, 832–848, 1987.
20. A. Federgrun and M. Tsur, "A simple forward algorithm to solve general dynamic lot-size models with n periods in $O(nlogn)$ or $O(n)$ time", Management Science 37, 909–925, 1991.
21. B. Fleischmann, "The discrete lotsizing and scheduling problem", European Journal of Operational Research 44(3), 337–348, 1990.
22. B. Fleischmann, "The discrete lotsizing and scheduling problem with sequence-dependent setup costs", European Journal of Operational Research,1994.
23. M. Florian and M. Klein, "Deterministic production planning with concave costs and capacity constraints", Management Science 18, 12–20, 1971.
24. M.X. Goemans, "Valid inequalities and separation for mixed 0-1 constraints with variable upper bounds", Operations Research Letters 8, 315–322, 1989.
25. M. Grotschel, L. Lovasz and A. Schrijver, "The ellipsoid method and its consequences in combinatorial optimization", Combinatorica 1, 169–197, 1981.
26. O. Gunluk and Y. Pochet, "Mixing mixed integer rounding inequalities", CORE discussion paper 9811, Université catholique de Louvain, Belgium, 1998. (to appear in Mathematical Programming)
27. Haase, "Lotsizing and scheduling for production planning", Lecture notes in economics and mathematical systems 408, Springer, Berlin, 1994.
28. F.W. Harris, "How many parts to make at once", Factory, the Magazine of Management 10(2), 1913.

29. A.C. Hax and H.C. Meal, “Hierarchical integration of production planning and scheduling”, in M. Geisler editor, TIMS studies in Management Science, chapter 1, North Holland/American Elsevier, New York, 1975.
30. S. Kang, K. Malik, L.J. Thomas, “Lotsizing and scheduling in parallel machines with sequence dependent setup costs”, Working paper 97-07, Johnson Graduate school of Management, Cornell university, 1997.
31. U.S. Karmarkar and L. Schrage, “The deterministic dynamic product cycling problem”, Operations Research 33, 326–345, 1985.
32. A. Kimms, “Multi-level lot sizing and scheduling: methods for capacitated, dynamic and deterministic models”, Physica-Verlag (production and logistics series), Heidelberg, 1997.
33. E. Kondili, C.C. Pantelides, R.W.H. Sargent, “A general algorithm for short-term scheduling of batch operations - 1. MILP formulation”, Computers Chemical Engineering 17, 211–227, 1993.
34. J. Krarup and O. Bilde, “Plant location, set covering and economic lot sizes: an $O(mn)$ algorithm for structured problems”, in “Optimierung bei Graphentheoretischen und Ganzzahligen Probleme”, L. Collatz et al. eds, Birkhauser Verlag, Basel, 155–180, 1977.
35. R. Kuik, M. Salomon and L.N. van Wassenhove, “Batching decisions: structure and models”, European Journal of Operational Research 75, 243–263, 1994.
36. L.S. Lasdon and R.C. Terjung, “An efficient algorithm for multi-item scheduling”, Operations Research 19, 946–969, 1971.
37. J. Leung, T.M. Magnanti and R. Vachani, “Facets and algorithms for capacitated lot-sizing”, Mathematical Programming 45, 331–359, 1989.
38. M. Loparic, Y.Pochet and L.A. Wolsey, “Uncapacitated lot-sizing with sales and safety stocks”, Mathematical Programming 89 (3), 487–504, 2001.
39. M. Loparic, H. Marchand and L.A. Wolsey, “Dynamic knapsack sets and capacitated lot-sizing”, CORE discussion paper 2000/47, Université catholique de Louvain, Louvain-la-Neuve, Belgium, 2000.
40. L. Lovasz, “Graph theory and integer programming”, Annals of Discrete Mathematics 4, 141–158, 1979.
41. S.F. Love, “A facilities in series inventory model with nested schedules”, Management Science 18(5), 327–338, 1972.
42. T.M. Magnanti and R. Vachani, “A strong cutting plane algorithm for production scheduling with changeover costs”, Operations Research 38, 456–473, 1990.
43. H. Marchand and L.A. Wolsey, “The 0-1 knapsack problem with a single continuous variable”, Mathematical Programming 85, 15–33, 1999.
44. R.K. Martin, “Generating alternative mixed-integer programming models using variable redefinition”, Operations Research 35, 331–359, 1987.
45. R.K. Martin, “Using separation algorithms to generate mixed integer model reformulations”, Operations Research Letters 10, 119–128, 1991.
46. A.J. Miller, G.L. Nemhauser and M.W.P. Savelsbergh, “On the polyhedral structure of a multi-item production planning model with setup times”, CORE discussion paper 2000/52, Université catholique de Louvain, Louvain-la-Neuve, Belgium, 2000.
47. G.L. Nemhauser and L.A. Wolsey, “Integer and combinatorial optimization”, Wiley, New York, 1988.
48. G.L. Nemhauser and L.A. Wolsey, “A recursive procedure for generating all cuts for 0-1 mixed integer programs”, Mathematical Programming 46, 379–390, 1990.

49. J. Orlicky, “Material Requirements planning”, McGraw-Hill, New York, 1975.
50. M.W. Padberg, T.J. Van Roy and L.A. Wolsey, “Valid inequalities for fixed charge problems”, Operations Research 33, 842–861 , 1985.
51. Pinto and Grossmann, “A logic-based approach to scheduling problems with resource constraints”, Computers and Chemical Engineering 21 (8), 801–818, 1997.
52. Y. Pochet, “Lot-sizing problems : reformulations and cutting plane algorithms”, PhD Thesis, Université Catholique de Louvain, Belgium, 1987.
53. Y.Pochet, “Valid inequalities and separation for capacitated economic lot-sizing”, Operations Research Letters 7, 109–116, 1988.
54. Y.Pochet and L.A. Wolsey, “Lot-size models with backlogging: Strong formulations and cutting planes”, Mathematical Programming 40, 317–335, 1988.
55. Y.Pochet and L.A. Wolsey, “Solving multi-item lot sizing problems using strong cutting planes”, Management Science 37, 53–67, 1991.
56. Y.Pochet and L.A. Wolsey, “Lot-sizing with constant batches: Formulation and valid inequalities”, Mathematics of Operations Research 18, 767–785, 1993.
57. Y.Pochet and L.A. Wolsey, “Polyhedra for lot-sizing with Wagner-Whitin costs”, Mathematical Programming 67, 297–323, 1994.
58. Y.Pochet and L.A. Wolsey, “Algorithms and reformulations for lot-sizing problems”, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, 20, 245-293, 1995.
59. Y. Pochet, T. Tahmassebi and L.A. Wolsey, “Reformulation of a single stage packing model”, Report, Memips: Esprit project 20118, June 1996.
60. R.L. Rardin and U. Choe, “Tighter relaxations of fixed charge network flow problems”, report J-79-18, Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia, 1979.
61. R. Rardin and L.A. Wolsey, “Valid inequalities and projecting the multicommodity extended formulation for uncapacitated fixed charge network flow problems”, European Journal of Operations Research, Nov 1993.
62. M. Salomon, “Deterministic lotsizing models for production planning”, PhD. Thesis, Erasmus Universiteit Rotterdam, The Netherlands, 1990.
63. M. Salomon, L.G. Kroon, R. Kuik, L.N. Van Wassenhove, “Some extensions of the discrete lotsizing and scheduling problem”, Management Science 37(7), 801–812, 1991.
64. N.C. Simpson, S.S. Erenguc, “Production planning in multiple stage manufacturing environments with joint costs, limited resources nad set-up times”, Technical report, Department of Management Science and Systems, University of Buffalo, 1998.
65. H. Stadtler, “Mixed integer programming model formulations for dynamic multi-item multi-level capacitated lotsizing”, European Journal of Operational Research 94 (3), 561–581, 1996.
66. H. Tempelmeier and M. Derstroff, “A Lagrangean-based heuristic for dynamic multilevel multiitem constrained lotsizing with setup times”, Management Science 42 (5), 738–757, 1996.
67. J.M. Thizy and L.N. Van Wassenhove “Lagrangean relaxation for the multi-item capacitated lot-sizing problem: a heuristic implementation”, IIE Transactions 17 (4), 308–313, 1985.
68. W. Trigeiro, L.J. Thomas and J.O. McClain, “Capacitated lot sizing with setup times”, Management Science 35(3), 353–366, 1989.
69. F. Vanderbeck, “Lot-sizing with start up times”, Management Science 44 (10), 1409–1425, 1998.

70. W. Van de Velde, Private communication, 1997.
71. C.P.M. van Hoesel, "Models and algorithms for single-item lot sizing problems", Ph.D. Thesis, Erasmus Universiteit, Rotterdam, 1991.
72. S. van Hoesel, A. Wagelmans and L.A. Wolsey, "Economic lot-sizing with start-up costs: the convex hull", SIAM Journal of Discrete Mathematics, 1994.
73. S. van Hoesel and A. Wagelmans , "An O(T-3) algorithm for the economic lot-sizing problem with constant capacities", Management Science 42 (1), 142–150, 1996.
74. T.J. Van Roy and L.A. Wolsey, "Valid inequalities and separation for uncapacitated fixed charge networks", Operations Research Letters 4, 105–112,1985.
75. A.F. Veinott, "Minimum concave cost solution of Leontief substitution models of multi-facility inventory systems", Operations Research 17(2), 262–291, 1969.
76. T.E. Vollman, W.L. Berry, and D.C. Whybark, "Manufacturing Planning and Control Systems", Third Edition, Richard D. Irwin., 1997.
77. A.P.M. Wagelmans, C.P.M. van Hoesel and A.W.J. Kolen, "Economic lot-sizing: an $O(nlogn)$ algorithm that runs in linear time in the Wagner-Whitin case", Operations Research 40, Supplement 1, 145–156, 1992.
78. H.M. Wagner and T.M. Whitin, "Dynamic version of the economic lot size model", Management Science 5, 89–96, 1958.
79. H. Westenberger and J. Kallrath, "Formulation of a jobshop problem in process industry", Preprint, Bayer, Leverkusen, January 1995.
80. R.H. Wilson, "A scientific routine for stock control", Harvard Business Review 13, 1934.
81. L.A. Wolsey, "Uncapacitated lot-sizing problems with start-up costs", Operations Research 37, 741–747, 1989.
82. L.A. Wolsey, "MIP modelling of changeovers in production planning and scheduling problems", European Journal of Operational Research 99, 154–165, 1997.
83. L.A. Wolsey, "Integer programming", Wiley, New York, 1999.
84. W.I. Zangwill, "Minimum concave cost flows in certain networks", Management Science 14, 429–450, 1968.
85. W.I. Zangwill, "A backlogging model and a multi-echelon model of a dynamic economic lot size production system – a network approach", Management Science 15(9), 506–527, 1969.
86. X. Zhang and R.W.H. Sargent, "A new unified formulation for process scheduling", AIChE annual meeting, Paper 144c, St Louis, Missouri, 1993.