MANAGEMENT SCIENCE
Vol. 33, No. 4, April 1987
Printed in U.S.A.

A MATHEMATICAL PROGRAMMING APPROACH TO A DETERMINISTIC KANBAN SYSTEM*

GABRIEL R. BITRAN AND LI CHANG
Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
IBM Corporation, Poughkeepsie, New York 12602

In this paper we present a mathematical programming model for the Kanban system in a deterministic multi-stage capacitated assembly-tree-structure production setting. We discuss solution procedures to the problem and address three special cases of practical interest.
(JUST-IN-TIME MANUFACTURING—KANBAN; INVENTORY/PRODUCTION—DETERMINISTIC MODELS; PROGRAMMING)

1. Introduction

The Kanban system is a multi-stage production scheduling and inventory control system. It is motivated by the concept of just-in-time production and aims at reducing the level of inventory to a minimum. Briefly speaking, the concept of just-in-time production is that materials should flow through the entire production sequence without being stopped or accumulated in an intermediate stage. Under this concept, no inventory of any kind is viewed as an absolute necessity.

Obviously, in many instances inventories are justified because of the important role they play. For example, cycle stock is carried due to the trade-off that has to be made between setup cost and inventory holding cost; and safety stock is accumulated to protect against various uncertainties. Unfortunately, the basic concepts that justify the existence of inventories have been abused over the years. Managers very often accept the existence of setup work without looking into the possibility of reducing it, which could lead to a down-sized cycle stock. Similarly, instead of improving the accuracy of forecasts of demand and lead times and ameliorating preventive maintenance procedures, managers often choose to increase safety stock. In short, inventory has become more of a cover-up of production problems than of a solution to them.

The Kanban system, originally designed by Toyota to realize just-in-time production, is intended to keep a tight control over inventory and force the hidden problems to surface so that they can be identified and addressed directly.

1.1. *Summary of the Operating Procedures of the Kanban System*

For the purpose of this paper, we present a brief description of how the Kanban system operates. For more details, the reader is referred to Kimura and Terada 1981, Monden 1981a, b, c and Sugimori et al. 1977. "Kanban", in the Japanese language, refers to a card or tag. It can serve as a production, delivery, or purchase order. In the system, items are put into containers and different types of items are held in different containers. Once a container is full, a Kanban is attached to it. A Kanban usually carries the following information: (1) item name, (2) item number, (3) description of the item, (4) container type, (5) container capacity, (6) Kanban identification number, (7) preceding stage, and (8) succeeding stage.

* Accepted by David G. Dannenbring; received March 22, 1985. This paper has been with the authors 4 months for 2 revisions.

0025-1909/87/3304/0427$01.25

In Figure 1, stage n represents an intermediate stage in a production setting. It encompasses a production process P^n and a subsequent inventory point I^n. The type of production process involved can be fabrication, subassembly, delivery, or purchase. Using as inputs the items stored in the inventory point of the immediate predecessor, process P^n produces its own items to fill a container and then stores the full container in I^n with a Kanban attached to it. When the first piece of a full container in I^n is used by the production process of the immediate successor, the Kanban originally attached to the container is detached and kept aside. At the end of each time period (for example, at the end of every half-shift), all the Kanbans detached in I^n during the time period are collected and sent back to P^n. These Kanbans then serve as new production orders for P^n. Generally P^n uses a first-in-first-out rule to process these orders. Once P^n produces a full container (i.e., P^n fills an order), the Kanban which ordered the full container is attached to it and the container is sent to I^n.

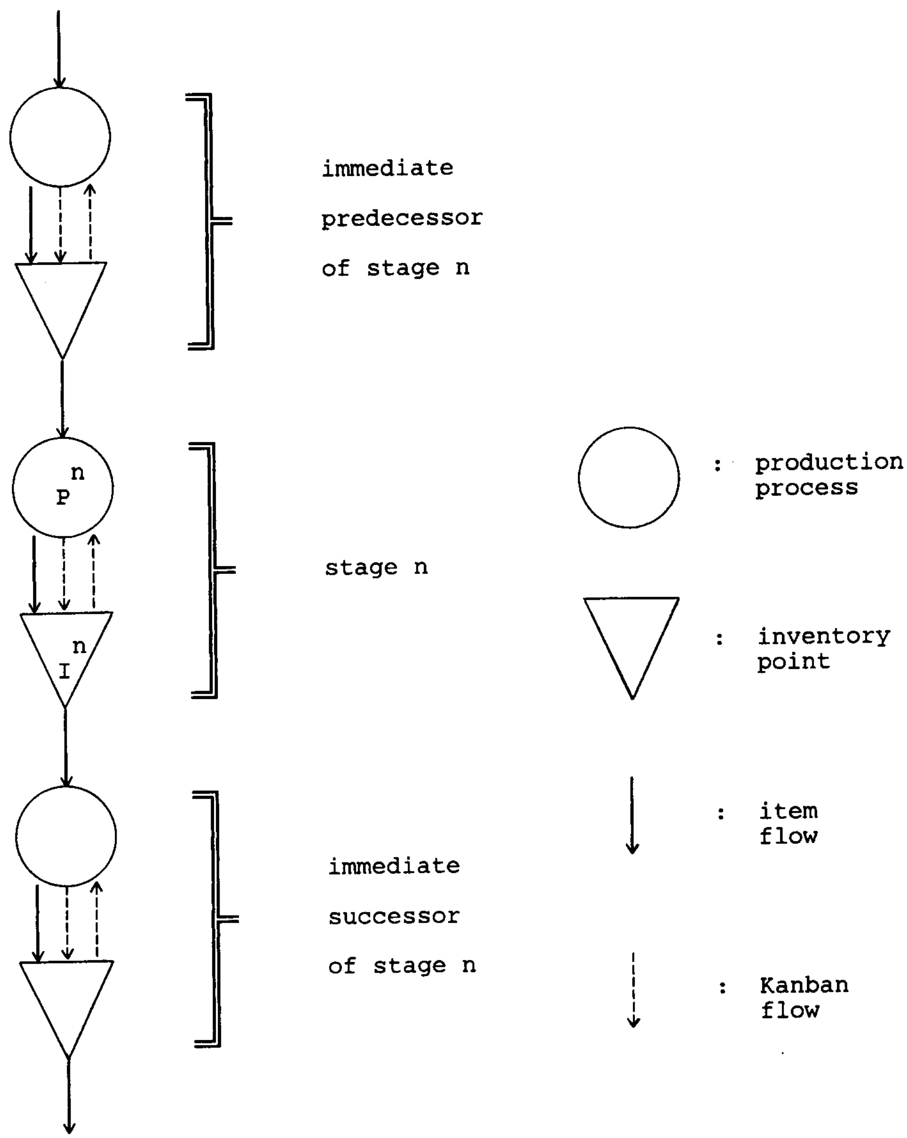

FIGURE 1. Flows of Items and Kanbans.

Below, we outline four important observations regarding the system. First, the total number of Kanbans circulating between P^n and I^n is unchanged over time, unless management interferes to drain Kanbans from, or to inject more Kanbans into, stage n. Second, the maximum inventory buildup in I^n is limited by the number of Kanbans circulating between P^n and I^n. Consequently, by controlling the number of circulating Kanbans and requiring that every full container have a Kanban attached to it, managers can be assured that the inventory buildup will not exceed a certain limit. Third, the movement of Kanbans between P^n and I^n is triggered by the inventory withdrawal from I^n by the immediate successor. In other words, P^n will produce to replenish what has been withdrawn from I^n by the immediate successor. Fourth, by circulating Kanbans within every stage, all the stages in a production setting are chained together. Therefore, the production schedule of the final stage is transmitted back to all the upstream stages. Since a detached Kanban automatically becomes a new order, managers need not issue any other document to trigger an order in an upstream stage. The upstream stages can actually be self-operated.

These features of the Kanban system reduce significantly the paperwork and the overhead to run the facilities and control the inventory. They also make the Kanban system robust in the sense that it tends to absorb and adapt to uncertainties, in demand and production, without requiring continuous management intervention. For instance, whenever a stage ceases to produce due to machine breakdown, it automatically stops sending Kanbans to its predecessors, and hence prevents the buildup of unwanted inventory. On the other hand, whenever the demand fluctuates markedly over the planning horizon the fixed number of Kanbans, released to a stage, will cause an early production and force the facility to carry more inventory than the one required by a traditional capacitated lot-size model.

Therefore, when choosing a Kanban system, managers need to consider the tradeoffs among the length of the planning horizon, the fluctuation of the demand pattern, the degree of overhead and management intervention, and the amount of extra inventory that might be implied by an easy-to-manage system. These trade-offs are of course situation dependent and should be made considering, in addition to the issues mentioned above, the culture of the firm, relations to suppliers, and the competing environment. Our experience indicates that the benefits associated with systems that are easy to manage have been often underestimated because they are frequently not easy to quantify.

Figure 1 depicts a serial production setting. Nevertheless, the reader can easily observe that the operating procedure, as described above, will also work with an assembly-type, a distribution-type, or a mixed-type production setting.

1.2. *Purposes of the Paper*

The Kanban system, enabling Toyota to drastically cut its inventory investment, has attracted much attention from production professionals worldwide. Most research efforts to date have focused on the comparison of the Kanban system, or Japanese production methods in general, and Western production methods. Rice and Yoshikawa (1982) contrasted Kanban with MRP (Material Requirements Planning). Schonberger (1982) provided nine lessons on Japanese manufacturing techniques from which Western companies could learn in order to simplify their production problems. Most recently, Krajewski, King, Ritzman, and Wong (1983) conducted a simulation experiment to identify the critical technical factors in the Japanese and U.S. manufacturing environments, represented by Kanban and MRP, respectively. With the exception of Kimura and Terada (1981), efforts have not been made to develop mathematical models for the Kanban system. In 1981, Kimura and Terada provided several basic equations for the Kanban system in a multi-stage serial production setting to show how the fluctuation of final demand influences the fluctuation of production and inventory volumes at upstream stages. In their theoretical analysis, they assumed small container size and unlimited production capacity.

The purpose of this paper is two-fold. First, it provides a mathematical programming formulation for the Kanban system in a deterministic multi-stage assembly production setting. The model assists managers in determining the number of circulating Kanbans, and hence the inventory level, at each stage. Contrasting with Kimura and Terada (1981), we make no assumptions on the container size (except for three special cases in which we make assumption on the *relative* container size between stages); in addition, we allow limited production capacity. As a result, our model should be applicable in more general manufacturing situations. Second, the paper investigates solution procedures, for the resulting Kanban model, that will make it usable in practice. To this end, the initial model, which is nonlinear integer in nature, is transformed into an integer linear program. The integer linear program presents the following advantages: (1) it is more tractable than the nonlinear model; and (2) it provides the same set of feasible solutions and the same set of optimal solutions as the nonlinear model in terms of the decision variables controlled by managers. To the same end, three special cases of practical interest are constructed on the basis of the relative container size between stages. In one case, the integer linear program is converted into a mixed integer linear program with the number of integer variables greatly reduced. In the other two cases, the linear programming (LP) technique is used and the relative error due to the LP approximation is shown to approach zero asymptotically.

The readers should note that the models that we analyze do not incorporate uncertainties. Therefore managers should adjust the number of Kanbans obtained from the models to take into consideration the potential uncertainties in demand and machine breakdowns. In addition, managers may wish to periodically resolve the models to incorporate additional information like revised demand forecast.

2. Model Description

The model that we present deals with a multi-stage capacitated assembly-tree-structure production setting with each stage producing one type of item. There are $N + 1$ stages in the setting. Let $n \in \{0, 1, \ldots, N\}$ index the stages with the understanding that $n_1 < n_2$ if stage n_1 succeeds stage n_2. We also denote an item by the index of the stage producing it. The final stage, stage 0, includes only the final assembly operation P^0, while every upstream stage $n \in \{1, 2, \ldots, N\}$ includes a production process P^n and an immediately succeeding inventory point I^n. An example of indexing is provided in Figure 2. Let $t \in \{0, 1, \ldots, T\}$ index the time periods with the understanding that the planning horizon starts at the beginning of period 1 and finishes at the end of period T. For the final stage, a time-phased production schedule is given and must be met. For each upstream stage, a production quota for the whole planning horizon is given and the quota is determined by the effective demand imposed upon the stage. Once an upstream P^n has reached its production quota, all the detached Kanbans remaining at P^n or to be sent to P^n in the future from I^n stop triggering any further production and are drained from the system by management at the end of the planning horizon.

Throughout the paper, we shall use $\lceil Z \rceil$ to denote the smallest integer which is larger than or equal to Z and $\lfloor Z \rfloor$ the largest integer which is smaller than or equal to Z. Proofs of propositions are omitted whenever they are not particularly difficult to reproduce.

Parameters

α^n = number of units of item n in a full container; $\alpha^n \in \{1, 2, \cdots\}$ ($n = 0, 1, \ldots, N$). These parameters represent container capacities.

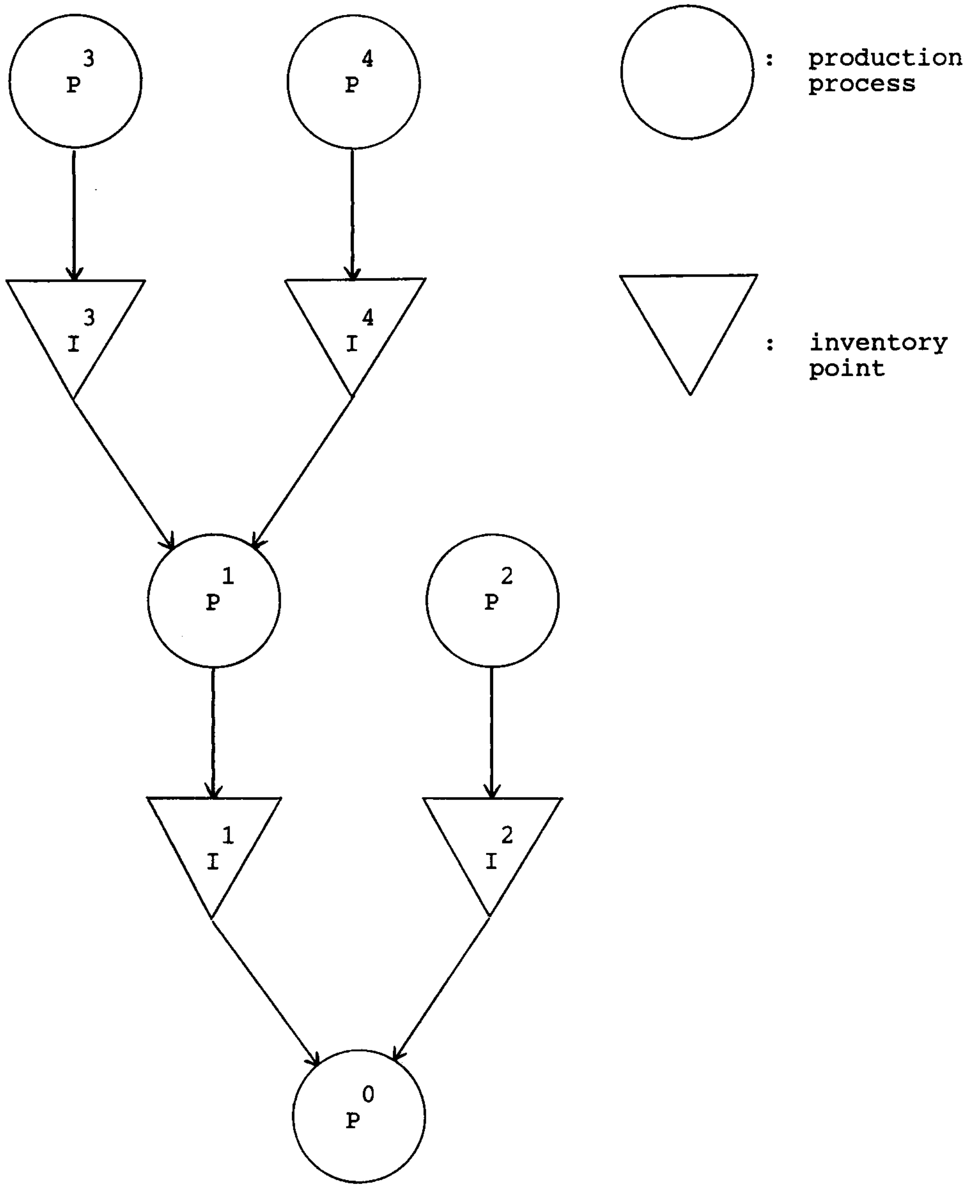

FIGURE 2. An Example of Indexing.

β_t^n = production capacity, in terms of the number of full containers of item n, at P^n in period t; $\beta_t^n \in \{0, 1, \cdots\}$ $(n = 1, \ldots, N;\ t = 1, \ldots, T)$.

$s(n)$ = immediately succeeding stage of stage n $(n = 1, \ldots, N)$.

$P(n)$ = set of immediately preceding stages of stage n $(n = 0, 1, \ldots, N)$.

$e^{n,s(n)}$ = number of units of item n which are required to make one unit of item $s(n)$; $e^{n,s(n)} \in \{1, 2, \cdots\}$ $(n = 1, \ldots, N)$.

V_0^n = number of full containers of item n available in I^n at the end of period 0; $V_0^n \in \{0, 1, 2, \cdots\}$ $(n = 1, \ldots, N)$. Note that each of these full containers has a Kanban attached to it.

W_0^n = number of units of item n remaining in a partially filled container, whose Kanban has been detached, in I^n at the end of period 0; $W_0^n \in \{0, 1, \ldots, \alpha^n - 1\}$ $(n = 1, \ldots, N)$.

X_t^0 = production requirement, in terms of the number of full containers of item 0 (i.e., the final product), at stage 0 in period t; $X_t^0 \in \{0, 1, 2, \cdots\}$ ($t = 1, \ldots, T$).

$Q^n = \max\{0, \lceil (e^{n,s(n)}\alpha^{s(n)}/\alpha^n)Q^{s(n)} - V_0^n - (W_0^n/\alpha^n)\rceil\}$
= production quota or effective demand, in terms of the number of full containers of item n, imposed upon stage n for the whole planning horizon; $Q^n \in \{0, 1, 2, \cdots\}$ ($n = 1, \ldots, N$). Q^0 is defined as $\sum_{t=1}^T X_t^0$.

To lessen the burden of notation we assume that the production lead time is zero, and that the Kanbans detached in I^n in period t are available to serve as production orders in P^n in period $t + 1$. This simplification will have no impact on the results of the paper. Note that at the beginning of the planning horizon, the initial inventory at stage n is composed of V_0^n full containers and W_0^n units of item n ($n = 1, \ldots, N$). Also note that both X_t^0 and β_t^n are allowed to vary from period to period in order to give management more flexibility in scheduling final assembly operations and shifting resources.

Variables

X_t^n = number of detached Kanbans of item n which respectively trigger the production of a full container in P^n in period t ($n = 1, \ldots, N$; $t = 1, \ldots, T$).

Y_t^n = number of Kanbans of item n which are detached from their associated containers in I^n in period t ($n = 1, \ldots, N$; $t = 1, \ldots, T$).

U_t^n = number of detached Kanbans of item n which are available in P^n at the end of period t and have not triggered any production yet ($n = 1, \ldots, N$; $t = 1, \ldots, T$).

V_t^n = number of full containers of item n which are available in I^n at the end of period t ($n = 1, \ldots, N$; $t = 1, \ldots, T$).

W_t^n = number of units of item n remaining in a partially filled container, whose Kanban has been detached, in I^n at the end of period t ($n = 1, \ldots, N$; $t = 1, \ldots, T$).

U_0^n = number of detached Kanbans of item n which are injected into P^n by management at the beginning of the planning horizon ($n = 1, \ldots, N$).

We shall use the following abbreviations for variables:

(1) $\langle U, V, W, X, Y\rangle$ stands for all the variables involved, which includes $N(T + 1)$ U-type variables, NT V-type variables, NT W-type variables, NT X-type variables, and NT Y-type variables.

(2) $\langle U_0, X\rangle$ stands for N U-type variables with $t = 0$ and NT X-type variables.

(3) $\langle U_0\rangle$ stands for $\langle U_0^1, U_0^2, \ldots, U_0^N\rangle$.

(4) $\langle X^n\rangle$ stands for $\langle X_1^n, X_2^n, \ldots, X_T^n\rangle$.

We describe mathematically the Kanban system as follows:

$$U_{t-1}^n + Y_{t-1}^n - X_t^n - U_t^n = 0, \qquad n = 1, \ldots, N; t = 1, \ldots, T, \tag{2.1}$$

$$V_{t-1}^n + X_t^n - Y_t^n - V_t^n = 0, \qquad n = 1, \ldots, N; t = 1, \ldots, T, \tag{2.2}$$

$$X_t^n = \min \begin{cases} U_{t-1}^n + Y_{t-1}^n, \\ \beta_t^n, \\ \lfloor (\alpha^k V_{t-1}^k + W_{t-1}^k + \alpha^k X_t^k)/(e^{k,n}\alpha^n)\rfloor \quad \text{all } k \in P(n), \\ Q^n - \sum_{\tau=1}^{t-1} X_\tau^n, \qquad n = 1, \ldots, N; t = 1, \ldots, T, \end{cases} \tag{2.3}$$

$$\alpha^k V_{t-1}^k + W_{t-1}^k + \alpha^k X_t^k \geq e^{k,0}\alpha^0 X_t^0, \qquad \text{all } k \in P(0); t = 1, \ldots, T, \tag{2.4}$$

$$\begin{cases} Y_t^n = \lceil (e^{n,s(n)}\alpha^{s(n)} X_t^{s(n)} - W_{t-1}^n)/\alpha^n \rceil, & n = 1, \ldots, N; t = 1, \ldots, T, \\ Y_0^n = 0, & n = 1, \ldots, N, \end{cases} \tag{2.5}$$

$$W^n_{t-1} + \alpha^n Y^n_t - e^{n,s(n)}\alpha^{s(n)} X^{s(n)}_t - W^n_t = 0, \qquad n = 1, \ldots, N; t = 1, \ldots, T, \quad (2.6)$$

$$U^n_0 \text{ nonnegative integer}, \qquad n = 1, \ldots, N. \quad (2.7)$$

Constraints (2.1) and (2.2) describe the conservation of flow in P^n and I^n, respectively, in terms of Kanbans. Constraints (2.3) indicate that the number of full containers put into production in P^n in period t is determined by the available detached Kanbans, production capacity, available inventories in the previous stages, and remaining production quota. Constraints (2.4) ensure that the production schedule of the final stage can be carried out. Constraints (2.5) indicate the number of Kanbans which are detached from their associated containers in I^n in period t. Constraints (2.6) describe the conservation of flow in I^n for the number of units of item n remaining in a partially filled container. The nonnegative integrality of U^n_0 is enforced by (2.7). No setup is involved explicitly in (2.1)–(2.7). If an upstream stage P^n needs a setup in a particular period t (due, for example, to the fact that the machinery in P^n is scheduled for other purposes in the previous period, i.e., $\beta^n_{t-1} = 0$), the setup is assumed to be executed externally to the model and the value of β^n_t is determined after making allowance for the setup.

THEOREM 2.1. *If* $\langle U, V, W, X, Y\rangle$ *satisfies* (2.1)–(2.7), *then*

(a) $U^n_t, V^n_t, W^n_t, X^n_t$, *and* Y^n_t *are nonnegative integers*, $n = 1, \ldots, N; t = 1, \ldots, T$,

(b) $Y^n_t \le V^n_{t-1} + X^n_t$, $n = 1, \ldots, N; t = 1, \ldots, T$,

(c) $W^n_t \le \alpha^n - 1$, $n = 1, \ldots, N; t = 1, \ldots, T$,

(d) $U^n_0 + V^n_0 = U^n_t + V^n_t + Y^n_t$, $n = 1, \ldots, N; t = 1, \ldots, T$,

(e) $\sum^T_{t=1} X^n_t = Q^n$, $n = 1, \ldots, N$.

The above theorem shows that (2.1)–(2.7) implicitly enforce the following properties: (a) In addition to U^n_0 ($n = 1, \ldots, N$), the rest of the variables are also nonnegative integers. (b) The number of Kanbans detached from their associated containers in I^n in period t is smaller than or equal to the number of full containers available in I^n at the end of period $t - 1$ plus the number of full containers received by I^n in period t. (c) The number of units of item n remaining in a partially filled container in I^n at the end of period t is smaller than the container size α^n. (d) The number of Kanbans circulating in each upstream stage is unchanged and equal to $U^n_0 + V^n_0$ during the planning horizon. (e) The production quota imposed on each upstream stage is met exactly.

We propose the following optimization model, henceforth referred to as model (M), for the Kanban system:

$$\text{Minimize} \quad \sum_{n=1}^{N} C^n[U^n_0 + V^n_0 + 1 - (1/\alpha^n)] \quad (2.8)$$

s.t. (2.1)–(2.7),

where C^n is the accumulated value of one full container of item n; in other words, C^n represents the sum of material, labor and all other manufacturing costs which have been accumulated by the system in a full container of item n. The cost objective (2.8) can be interpreted in two different ways. One interpretation is that it represents an upper bound on the value tied up in inventory in the system at any point in time. It is not difficult to construct examples where this bound is attained in some period of the planning horizon. Clearly, the multiplication of (2.8) by the inventory holding charge over the planning horizon results in an upper bound for the inventory holding cost over the same period. The other interpretation of (2.8) is that it attempts to minimize a weighted combination of the number of Kanbans in circulation. A smaller number of Kanbans circulating at a stage reflects higher operating efficiency at that stage, and hence it is perceived as a desirable goal by workers and management.

3. Model Solution

Model (M) is a complex integer problem. The nonlinear constraints (2.3) and linear constraints (2.5), when re-expressed in more operational forms, greatly increase the number of integer variables and the number of constraints.

3.1. *Transformation*

In this section, we shall transform model (M) into a simpler model such that both have the same set of feasible, and optimal, solutions in terms of $\langle U_0 \rangle$, and the same optimal value. The transformation is motivated by the observation that if $\langle U, V, W, X, Y \rangle$ and $\langle \bar{U}, \bar{V}, \bar{W}, \bar{X}, \bar{Y} \rangle$ are two feasible solutions to (2.1)–(2.7) and $\langle U_0 \rangle = \langle \bar{U}_0 \rangle$, then $\langle U, V, W, X, Y \rangle = \langle \bar{U}, \bar{V}, \bar{W}, \bar{X}, \bar{Y} \rangle$. In other words, once U_0^n $(n = 1, \ldots, N)$ assume their specific values, all other variables in (2.1)–(2.7) are uniquely determined. This observation corresponds to the characteristic of the Kanban system that it is self-operational once the Kanbans have been distributed to the stages. The key decision variables that need to be controlled by management are the U_0^n $(n = 1, \ldots, N)$.

For future discussion, the following nomenclature will be adopted. A partial solution is said to satisfy, or to be feasible in, a set of constraints if there exists a complement to it such that the whole solution, i.e., the partial one together with its complement, satisfies all the constraints. For example, the partial solution $\langle X, Y \rangle$ satisfies, or is feasible in, (2.1)–(2.7) if there exists $\langle U, V, W \rangle$ such that the whole solution $\langle U, V, W, X, Y \rangle$ satisfies (2.1)–(2.7). Similarly, a partial solution is said to be feasible (optimal) in an optimization model if there exists a complement to it such that the whole solution is feasible (optimal) in the model. Two optimization models are said to have the same feasible (optimal) partial solution if there exist two complements, which may or may not be different from each other, such that the two resulting whole solutions are feasible (optimal) in the two models, respectively.

Let $E^{n,s(n)} = e^{n,s(n)}\alpha^{s(n)}/\alpha^n$ for $n = 1, \ldots, N$. The parameter $E^{n,s(n)}$ represents the number of full containers of item n required to make one full container of item $s(n)$. Depending upon the values of $e^{n,s(n)}$, $\alpha^{s(n)}$, and α^n, the value of $E^{n,s(n)}$ may or may not be integral. Also let $0 < \epsilon < \min\{1/\alpha^n | n = 1, \ldots, N\} \leq 1$. Consider the following optimization model:

minimize (2.8)

s.t.

$$(W_0^n/\alpha^n) + V_0^n + \sum_{\tau=1}^{t} X_\tau^n - E^{n,s(n)} \sum_{\tau=1}^{t} X_\tau^{s(n)} \geq 0, \qquad n = 1, \ldots, N; t = 1, \ldots, T, \quad (3.1)$$

$$U_0^n - \sum_{\tau=1}^{t} X_\tau^n + E^{n,s(n)} \sum_{\tau=1}^{t-1} X_\tau^{s(n)} - (W_0^n/\alpha^n) + 1 - \epsilon \geq 0, \qquad n = 1, \ldots, N; t = 1, \ldots, T, \quad (3.2)$$

$$X_t^n \in \{0, 1, \ldots, \beta_t^n\}, \qquad n = 1, \ldots, N; t = 1, \ldots, T, \quad (3.3)$$

$$U_0^n \in \{0, 1, 2, \cdots\}, \qquad n = 1, \ldots, N. \quad (3.4)$$

We refer to the above model as model (M0). The relation between model (M) and model (M0) is summarized in the next theorem.

THEOREM 3.1. *(M) is feasible if and only if (M0) is feasible. The two models have the same set of feasible partial solutions $\langle U_0 \rangle$, the same set of optimal partial solutions $\langle U_0 \rangle$, and the same optimal value.*

(M0) is an integer linear program which has $2NT$ constraints, excluding (3.3)–(3.4), and $NT + N$ integral variables. The configuration of (M0) is computationally more favorable than that of the nonlinear integer problem (M). However, it should be pointed out that the constraints of (M0) do not describe the operating procedure of the Kanban system while those of (M) do. The link between (M) and (M0) hinges on U_0^n ($n = 1, \ldots, N$), as shown in Theorem 3.1. Since U_0^n ($n = 1, \ldots, N$) are the only decision variables controlled by management, we can solve (M0) and still obtain a relevant feasible or optimal partial solution $\langle U_0 \rangle$ to (M).

The proof of Theorem 3.1 follows directly from Lemmas A.1 and A.2 in Appendix 1. Note that the constraints of model (M0), i.e., (3.1)–(3.4) do not specifically involve the production quota Q^n. It can be easily shown that $\sum_{\tau=1}^{T} X_\tau^n \geq Q^n$ for $n = 1, \ldots, N$ if $\langle X \rangle$ satisfies (3.1) and (3.3). For any $\langle U_0, \bar{X} \rangle$ satisfying (3.1)–(3.4), the proof of Lemma A.1 provides a way for constructing $\langle X \rangle$ such that $\sum_{\tau=1}^{T} X_\tau^n \leq Q^n$ for $n = 1, \ldots, N$ and $\langle U_0, X \rangle$ still satisfies (3.1)–(3.4). Obviously, the newly constructed $\langle X \rangle$ does satisfy the constraints $\sum_{\tau=1}^{T} X_\tau^n = Q^n$ for $n = 1, \ldots, N$.

3.2. *Feasibility Test*

Before solving model (M0), it is important to know whether it is feasible or not. Let $\Omega = \{\langle X \rangle | \langle X \rangle$ satisfies (3.1) and (3.3)$\}$. It is obvious that $\Omega \neq \varnothing$ if and only if (M0) is feasible. Gabbay (1979) devised a method to test if a multi-stage serial production setting is feasible. A similar feasibility test can be applied to Ω, which represents a multi-stage assembly production setting. The algorithm of the feasibility test is given below. The procedure attempts to compute a feasible solution to Ω such that each stage satisfies the effective demand of the succeeding stage by producing as late as possible.

(*Step* 0)

$$\bar{X}_t^0 \leftarrow X_t^0 \quad \text{for} \quad t = 1, \ldots, T,$$

$$n \leftarrow 1.$$

(*Step* 1)

$$Q_1^n \leftarrow \max \{0, \lceil E^{n,s(n)} \bar{X}_1^{s(n)} - V_0^n - (W_0^n/\alpha^n) \rceil\},$$

$$Q_t^n \leftarrow \max \{0, \lceil E^{n,s(n)} \sum_{\tau=1}^{t} \bar{X}_\tau^{s(n)} - V_0^n - (W_0^n/\alpha^n) \rceil\} - \sum_{\tau=1}^{t-1} Q_\tau^n \quad \text{for} \quad t = 2, \ldots, T,$$

$$t \leftarrow T.$$

(*Step* 2) $\bar{X}_t^n \leftarrow \min \{Q_t^n, \beta_t^n\}$; if $t = 1$, then go to Step 3; else,

$$Q_{t-1}^n \leftarrow Q_{t-1}^n + \max \{0, Q_t^n - \beta_t^n\},$$

$$t \leftarrow t - 1,$$

return to the beginning of Step 2.

(*Step* 3) If $\beta_1^n < Q_1^n$, then $\Omega = \varnothing$ and exit; else, if $n = N$, then $\Omega \neq \varnothing$ and exit; else, $n \leftarrow n + 1$ and return to Step 1.

Through Steps 1 and 2, the algorithm computes

$$\langle \bar{X}^n \rangle \in \Omega(n; \bar{X}^{s(n)}) = \{\langle X^n \rangle | (W_0^n/\alpha^n) + V_0^n + \sum_{\tau=1}^{t} X_\tau^n - E^{n,s(n)} \sum_{\tau=1}^{t} \bar{X}_\tau^{s(n)} \geq 0,$$

$$X_t^n \in \{0, 1, \ldots, \beta_t^n\} \quad \text{for} \quad t = 1, \ldots, T\}$$

such that $\sum_{\tau=1}^{t} \bar{X}_\tau^n \leq \sum_{\tau=1}^{t} X_\tau^n$ ($t = 1, \ldots, T$) for every $\langle X^n \rangle \in \Omega(n; \bar{X}^{s(n)})$. If $\beta_1^n < Q_1^n$ in Step 3, then such $\langle \bar{X}^n \rangle$ does not exist and consequently $\Omega(n; \bar{X}^{s(n)}) = \Omega = \varnothing$. This feasibility test is necessary and sufficient in the sense that it generates a vector $\langle X \rangle$ if and only if (M0) is feasible.

4. Model Specialization

In the previous discussion, we developed a general approach to the Kanban system. In practice, managers may find certain choices of container size desirable. In this section, we shall investigate the solution procedures to the Kanban model for three particular choices of container size.

4.1. *A Container-for-Container Mode*

By container-for-container mode we mean that exactly one full container of item n is required to make one full container of the subsequent item $s(n)$ for all $n \in \{1, \ldots, N\}$. In other words, under this mode container sizes must be adjusted properly so that $e^{n,s(n)}\alpha^{s(n)} = \alpha^n$, or $E^{n,s(n)} = 1$, for all $n \in \{1, \ldots, N\}$. Note that $e^{n,s(n)}$ is given by the product specification. The parameters to adjust are α^n for $n = 1, \ldots, N$. The one-for-one scheme, whenever possible, is a convenient one and it is supported by the philosophy of just-in-time production. It does not need handling multiple containers from one stage to the next as would be the situation for the case where $E^{n,s(n)}$ is greater than one; neither does it tend to accumulate the inventory of work-in-process as would be the situation for the case where $E^{n,s(n)}$ is smaller than one.

Consider the following optimization model:

$$\text{minimize (2.8)}$$

s.t.

$$V_0^n + \sum_{\tau=1}^{t} X_\tau^n - \sum_{\tau=1}^{t} X_\tau^{s(n)} \geq 0, \qquad n = 1, \ldots, N; t = 1, \ldots, T, \tag{4.1}$$

$$U_0^n - \sum_{\tau=1}^{t} X_\tau^n + \sum_{\tau=1}^{t-1} X_\tau^{s(n)} \geq 0, \qquad n = 1, \ldots, N; t = 1, \ldots, T, \tag{4.2}$$

$$0 \leq X_t^n \leq \beta_t^n, \qquad n = 1, \ldots, N; t = 1, \ldots, T, \tag{4.3}$$

$$U_0^n \in \{0, 1, 2, \cdots\}, \qquad n = 1, \ldots, N. \tag{4.4}$$

We refer to the above model as (M1). We summarize in the next theorem the relation between models (M0) and (M1).

THEOREM 4.1. *Assume that $E^{n,s(n)} = 1$ for all $n \in \{1, \ldots, N\}$. (M0) is feasible if and only if (M1) is feasible. The two models have the same set of feasible partial solutions $\langle U_0 \rangle$, the same set of optimal partial solutions $\langle U_0 \rangle$, and the same optimal value.*

(The proof of Theorem 4.1 is provided in Appendix 2.)

(M1) is a mixed integer linear program which has $2NT$ constraints, excluding (4.3)–(4.4), and $NT + N$ variables, of which N are required to be integral. Recall that $NT + N$ variables are required to be integral in (M0). Therefore, we can expect that it is easier to solve (M1) than to solve (M0); and Theorem 4.1 ensures that we can still obtain a relevant optimal partial solution $\langle U_0 \rangle$ for (M0), and hence for (M), by solving (M1). Note that for any solution $\langle U_0, \bar{X} \rangle$ to model (M1), the proof of Theorem 4.1 provides a way for constructing $\langle X \rangle$ such that $\langle U_0, X \rangle$ is a feasible solution to model (M0) under the container-for-container mode.

4.2. *A One-Container-for-Multiple-Containers Mode*

By one-container-for-multiple-container mode, we mean that one full container of item n is required to make an integral number of full containers of the subsequent item $s(n)$ for all $n \in \{1, \ldots, N\}$. Under this mode, the container sizes must be adjusted properly so that the inverse of $E^{n,s(n)}$ is an integer for all $n \in \{1, \ldots, N\}$.

We impose the following conditions: (a) the inverse of $E^{n,s(n)}$ is an integer for all $n \in \{1, \ldots, N\}$; and (b) W_0^n is an integral multiple of $e^{n,s(n)}\alpha^{s(n)}$, or equivalently (W_0^n/α^n) is an integral multiple of $E^{n,s(n)}$, for all $n \in \{1, \ldots, N\}$. Condition (a) is the definition of one-container-for-multiple-containers mode. Condition (b) deals with the partially filled containers which exist at the beginning of the planning horizon. As shown in the next theorem, condition (b) can always be enforced under condition (a).

THEOREM 4.2. *Assume that condition* (a) *holds. Let* $\bar{W}_0^n \in \{0, 1, \ldots, \alpha^n - 1\}$ *and* $\bar{\bar{W}}_0^n = e^{n,s(n)}\alpha^{s(n)} * \lfloor \bar{W}_0^n/(e^{n,s(n)}\alpha^{s(n)}) \rfloor$ *for* $n = 1, \ldots, N$. *Then,* $\langle U_0, X \rangle$ *satisfies* (3.1)–(3.4) *with* $\langle W_0 \rangle = \langle \bar{W}_0 \rangle$ *if and only if* $\langle U_0, X \rangle$ *satisfies* (3.1)–(3.4) *with* $\langle W_0 \rangle = \langle \bar{\bar{W}}_0 \rangle$.

(The proof of Theorem 4.2 is provided in Appendix 3.)

According to Theorem 4.2, we can always replace W_0^n by $e^{n,s(n)}\alpha^{s(n)} * \lfloor W_0^n/(e^{n,s(n)}\alpha^{s(n)}) \rfloor$ in model (M0) without altering the feasible region of the model under the one-container-for-multiple-containers mode. As a result, condition (b) does not really impose any additional restriction.

Consider the following optimization model:

$$\text{minimize } (2.8)$$

s.t. (3.1), (3.2)

$$0 \le X_t^n \le \beta_t^n, \qquad n = 1, \ldots, N; t = 1, \ldots, T, \tag{3.3'}$$

$$U_0^n \ge 0, \qquad n = 1, \ldots, N. \tag{3.4'}$$

We refer to the above model as model (M2). The model is a linear programming relaxation of (M0). We summarize their relation in the next theorem.

THEOREM 4.3. *Assume that conditions* (a) *and* (b) *hold.* (*M0*) *is feasible if and only if* (*M2*) *is feasible. If* $\langle U_0, X \rangle$ *is feasible for* (*M0*), *it is also feasible for* (*M2*). *If* $\langle U_0, X \rangle$ *is feasible for* (*M2*), *then* $\langle \lceil U_0^1 \rceil + 1, \ldots, \lceil U_0^N \rceil + 1 \rangle$ *is a feasible partial solution to* (*M0*).

(The proof of Theorem 4.3 is provided in Appendix 4.)

It is easier to solve (M2), a linear program, than to solve (M0), an integer linear program. Theorem 4.3 provides a way to approximate the optimal partial solution $\langle U_0 \rangle$ of (M0) by solving (M2). We are therefore interested in knowing the performance of the approximation. Let Z_0 be the optimal objective value of (M0) and Z_2 the objective value given by $\langle \lceil U_0^1 \rceil + 1, \ldots, \lceil U_0^N \rceil + 1 \rangle$ where $\langle U_0 \rangle$ is an optimal partial solution to (M2). The next theorem shows that the relative error $(Z_2 - Z_0)/Z_0$ caused by the LP approximation approaches zero asymptotically.

THEOREM 4.4. *Assume that conditions* (a) *and* (b) *hold. Also assume that* (*M0*) *and* (*M2*) *are feasible. Then* $(Z_2 - Z_0)/Z_0 \to 0$ *as* $Q^0/T \to \infty$.

(The proof of Theorem 4.4 is provided in Appendix 5.)

In the above theorem, Q^0/T represents the average production requirement (in terms of the number of containers) per period at the final stage. When Q^0/T becomes very large, the relative error due to the approximation of (M0) by (M2) becomes negligible.

EXAMPLE 1. Let $N = 1$, $T = 2$, $C^1 = 1$, $\alpha^1 = 5$, $e^{1,0} = 1$, $\alpha^0 = 1$, $E^{1,0} = 0.2$, $\beta_1^1 = 0$, $\beta_2^1 = \infty$, $\epsilon = 0.1$, $V_0^1 = 0$, $W_0^1 = 4$, $X_1^0 = 4$, and $X_2^0 = 26$. At optimality, $U_0^1 = 6$ for (M0) while $U_0^1 = 4.3$ for (M2). Since $\lceil 4.3 \rceil + 1 = 6$, the approximation method actually finds the optimal U_0^1 for (M0) in this example. □

EXAMPLE 2. Let $N = 1$, $T = 2$, $C^1 = 1$, $\alpha^1 = 4$, $e^{1,0} = 1$, $\alpha^0 = 1$, $E^{1,0} = 0.25$, $\beta_1^1 = \beta_2^1 = \infty$, $\epsilon = 0.1$, $V_0^1 = 0$, $W_0^1 = 0$, and $X_1^0 = X_2^0 = 8$. At optimality, $U_0^1 = 2$ for (M0) while $U_0^1 = 1.1$ for (M2). Since $\lceil 1.1 \rceil + 1 = 3 > 2$, the approximation method finds a non-optimal feasible U_0^1 for (M0) in this case. With $Q^0/T = 8$, the relative error $(Z_2 - Z_0)/Z_0 = 4/11$.

Now let $X_1^0 = X_2^0 = 800$ while the other parameters remain unchanged. At optimality, $U_0^1 = 200$ for (M0) while $U_0^1 = 199.1$ for (M2). With $Q^0/T = 800$, the relative error $(Z_2 - Z_0)/Z_0 = 4/803$. □

4.3. *A Multiple-Containers-for-One-Container Mode*

In this section we analyze the case of multiple-containers-for-one-container. That is, the case where an integral number of full containers of item n are required to produce one full container of subsequent item $s(n)$ for $n \in \{1, \ldots, N\}$. Under this mode the container sizes must be adjusted properly so that $E^{n,s(n)}$ is an integer for $n \in \{1, \ldots, N\}$. This mode is in line with the philosophy of just-in-time production because it discourages the use of large-sized containers in the upstream stages and tends not to create partially filled containers. For future discussion, let $M^0 = 1$ and $M^n = E^{n,s(n)}M^{s(n)}$ for $n = 1, \ldots, N$. The parameter M^n represents the number of full containers of item n which are required to make one full container of final product in stage 0. We will also let $\lceil x \rceil_k$ denote the smallest integer which is greater than or equal to x and is an integral multiple of k.

In this section, we assume the following conditions: (1) $E^{n,s(n)}$ is a nonnegative integer for $n \in \{1, \ldots, N\}$; (2) V_0^n is an integral multiple of M^n for $n \in \{1, \ldots, N\}$ and (3) β_t^n $(t = 1, \ldots, T)$ are integral multiples of M^n for all $n \in \{1, \ldots, N\}$.

Consider the following optimization model:

minimize (2.8)

s.t.

$$V_0^n + \sum_{\tau=1}^{t} X_\tau^n - E^{n,s(n)} \sum_{\tau=1}^{t} X_\tau^{s(n)} \geq 0, \qquad n = 1, \ldots, N; t = 1, \ldots, T, \tag{4.5}$$

$$U_0^n - \sum_{\tau=1}^{t} X_\tau^n + E^{n,s(n)} \sum_{\tau=1}^{t-1} X_\tau^{s(n)} \geq 0, \qquad n = 1, \ldots, N; t = 1, \ldots, T, \tag{4.6}$$

$$0 \leq X_t^n \leq \beta_t^n, \qquad n = 1, \ldots, N; t = 1, \ldots, T, \tag{4.7}$$

$$U_0^n \geq 0, \qquad n = 1, \ldots, N. \tag{4.8}$$

We refer to the above problem as (M3). The model is a linear programming model and its relation to (M0) is summarized in the next theorem.

THEOREM 4.5. *Assume that conditions* (1), (2) *and* (3) *hold.* (*M0*) *is feasible if and only if* (*M3*) *is feasible. If* $\langle U_0, X \rangle$ *is feasible in* (*M0*), *it is also feasible in* (*M3*). *If* $\langle U_0, X \rangle$ *is feasible in* (*M3*), *then* $\langle \lceil U_0^1 \rceil_{M^1}, \ldots, \lceil U_0^N \rceil_{M^N} \rangle$ *is a feasible partial solution to* (*M0*).

To evaluate how well an optimal partial solution $\langle U_0 \rangle$ to (M3) can approximate (M0), we let Z_0 be the optimal objective value of (M0) and Z_3 the objective value given by $\langle \lceil U_0^1 \rceil_{M^1}, \ldots, \lceil U_0^N \rceil_{M^N} \rangle$. The next theorem shows that the relative error $(Z_3 - Z_0)/Z_0$ caused by the LP approximation approaches zero asymptotically.

THEOREM 4.6. *Assume that conditions* (1), (2) *and* (3) *hold. Also assume that* (*M0*) *and* (*M3*) *are feasible. Then* $(Z_3 - Z_0)/Z_0 \to 0$ *as* $Q^0/T \to \infty$.

EXAMPLE 3. Let $N = 3$, $T = 3$, $s(3) = 2$, $s(2) = 1$, $s(1) = 0$, $C^3 = 1$, $C^2 = 6$, $C^1 = 15$, $\alpha^n = 1$ for $n \in \{1, 2, 3\}$, $e^{3,2} = 5$, $e^{2,1} = 2$, $e^{1,0} = 1$, $\beta_t^n = \infty$ for $n \in \{1, 2, 3\}$ and $t \in \{1, 2, 3\}$, $V_0^n = 0$ for $n \in \{1, 2, 3\}$, $W_0^n = 0$ for $n \in \{1, 2, 3\}$, $X_1^0 = 0$, $X_2^0 = 0$ and $X_3^0 = 2$. At optimality, $U_0^3 = 10$, $U_0^2 = 2$ and $U_0^1 = 2$ for (M0) while $U_0^3 = 20/3$, $U_0^2 = 4/3$ and $U_0^1 = 2$ for (M3). Since $\lceil 20/3 \rceil_{10} = 10$, $\lceil 4/3 \rceil_2 = 2$ and $\lceil 2 \rceil_1 = 2$, the approximation method actually finds the optimal $\langle U_0 \rangle$ for (M0) in this example. □

EXAMPLE 4. Assume the same parameters as used in Example 3, except that $\beta_3^3 = 0$. At optimality, $U_0^3 = 10$, $U_0^2 = 2$ and $U_0^1 = 2$ for (M0) while $U_0^3 = 40/3$, $U_0^2 = 4/3$ and $U_0^1 = 2$ for (M3). Since $\lceil 40/3 \rceil_{10} = 20 > 10$, the approximation method finds a nonoptimal feasible $\langle U_0 \rangle$ for (M0) in this case. With $Q^0/T = 2/3$, the relative error $(Z_3 - Z_0)/Z_0 = 10/52$.

Now let $X_3^0 = 200$ while the other parameters remain unchanged. At optimality, $U_0^3 = 1330$, $U_0^2 = 134$ and $U_0^1 = 200$ for (M0) while $U_0^3 = 4000/3$, $U_0^2 = 400/3$ and $U_0^1 = 200$ for (M3). With $Q^0/T = 200/3$, the relative error $(Z_3 - Z_0)/Z_0 = 10/5134$. □

5. Conclusion

In this paper we have presented an optimization model for the Kanban system. The model is intended for a deterministic multi-stage capacitated assembly-type production setting. We have also provided a general solution procedure to the model and discussed three special cases of practical interest. Future research topics would include the development of Kanban models for the distribution-type and mixed-type production settings, the inclusion of independent (external) demands for upstream items, and a direct treatment of uncertainties in the model.[1]

[1] The authors are grateful to Professors Hirofumi Matuso and Devanath Tirupati, and two anonymous referees for their useful comments on an earlier version of this paper.

Appendix 1. Lemmas A.1 and A.2

Let (2.3)′ and (2.7)′ be respectively, for $n = 1, \ldots, N$ and $t = 1, \ldots, T$, the set of constraints derived by expressing (2.3) as inequalities and by extending the integrality condition to X_t^n.

LEMMA A.1. *If* $\langle U, V, W, X, Y\rangle$ *satisfies* (2.1)–(2.2), (2.3)′, (2.4)–(2.6), *and* (2.7)′, *then* $\langle U_0, X\rangle$ *satisfies* (3.1)–(3.4). *If* $\langle U_0, X\rangle$ *satisfies* (3.1)–(3.4), *then* $\langle U_0\rangle$ *is a feasible partial solution to* (2.1)–(2.2), (2.3)′, (2.4)–(2.6), *and* (2.7)′.

[PROOF] Let $\langle U, V, W, X, Y\rangle$ satisfy (2.1)–(2.2), (2.3)′, (2.4)–(2.6), and (2.7)′. It follows from (2.2), (2.3)′, (2.4), and (2.6) that

$$e^{n,s(n)}\alpha^{s(n)}X_t^{s(n)} \le \alpha^n V_{t-1}^n + W_{t-1}^n + \alpha^n X_t^n$$
$$= \alpha^n\Big(V_0^n + \sum_{\tau=1}^{t-1} X_\tau^n - \sum_{\tau=1}^{t-1} Y_\tau^n\Big) + \Big(W_0^n + \alpha^n \sum_{\tau=1}^{t-1} Y_\tau^n - e^{n,s(n)}\alpha^{s(n)} \sum_{\tau=1}^{t-1} X_\tau^{s(n)}\Big) + \alpha^n X_t^n.$$

Hence,

$$(W_0^n/\alpha^n) + V_0^n + \sum_{\tau=1}^{t} X_\tau^n - E^{n,s(n)} \sum_{\tau=1}^{t} X_\tau^{s(n)} \ge 0.$$

So, $\langle U_0, X\rangle$ satisfies (3.1). It follows from (2.5) and (2.6) that $\sum_{\tau=1}^{t} Y_\tau^n = \lceil (e^{n,s(n)}\alpha^{s(n)} \sum_{\tau=1}^{t} X_\tau^{s(n)} - W_0^n)/\alpha^n \rceil$; and it follows from (2.1) and (2.3)′ that $U_0^n - \sum_{\tau=1}^{t} X_\tau^n + \sum_{\tau=1}^{t-1} Y_\tau^n \ge 0$. Hence,

$$U_0^n - \sum_{\tau=1}^{t} X_\tau^n + E^{n,s(n)} \sum_{\tau=1}^{t-1} X_\tau^{s(n)} - (W_0^n/\alpha^n) + 1 - \epsilon \ge U_0^n - \sum_{\tau=1}^{t} X_\tau^n + \lceil (e^{n,s(n)}\alpha^{s(n)} \sum_{\tau=1}^{t-1} X_\tau^{s(n)} - W_0^n)/\alpha^n \rceil$$
$$= U_0^n - \sum_{\tau=1}^{t} X_\tau^n + \sum_{\tau=1}^{t-1} Y_\tau^n \ge 0.$$

So, $\langle U_0, X\rangle$ satisfies (3.2). The first half of the lemma is then proved.

Let $\langle U_0, \bar X\rangle$ satisfy (3.1)–(3.4). Define $\langle X\rangle$ as follows:

$$X_1^n = \min\{\bar X_1^n, Q^n\} \quad \text{and} \quad X_t^n = \min\Big\{\sum_{\tau=1}^{t} \bar X_\tau^n, Q^n\Big\} - \min\Big\{\sum_{\tau=1}^{t-1} \bar X_\tau^n, Q^n\Big\} \quad \text{for} \quad t = 2, \ldots, T.$$

It follows from the definition of $\langle X\rangle$ that $\sum_{\tau=1}^{t} X_\tau^n = \min\{\sum_{\tau=1}^{t} \bar X_\tau^n, Q^n\}$ for $t = 1, \ldots, T$ and in particular, $\sum_{\tau=1}^{T} X_\tau^n \le Q^n$. We now show that $\langle U_0, X\rangle$ also satisfies (3.1)–(3.4). It follows from the definition of Q^n that $(W_0^n/\alpha^n) + V_0^n + Q^n - E^{n,s(n)}Q^{s(n)} \ge 0$. Since $\langle U_0, \bar X\rangle$ satisfies (3.1),

$$(W_0^n/\alpha^n) + V_0^n + \sum_{\tau=1}^{t} X_\tau^n - E^{n,s(n)} \sum_{\tau=1}^{t} X_\tau^{s(n)}$$
$$= (W_0^n/\alpha^n) + V_0^n + \min\Big\{\sum_{\tau=1}^{t} \bar X_\tau^n, Q^n\Big\} - E^{n,s(n)} * \min\Big\{\sum_{\tau=1}^{t} \bar X_\tau^{s(n)}, Q^{s(n)}\Big\} \ge 0$$

and hence $\langle U_0, X\rangle$ satisfies (3.1). It also follows from the definition of Q^n and ϵ that $-Q^n + E^{n,s(n)}Q^{s(n)} - (W_0^n/\alpha^n) + 1 - \epsilon > 0$. Since $\langle U_0, \bar X\rangle$ satisfies (3.2) and (3.4),

$$U_0^n - \sum_{\tau=1}^{t} X_\tau^n + E^{n,s(n)} \sum_{\tau=1}^{t-1} X_\tau^{s(n)} - (W_0^n/\alpha^n) + 1 - \epsilon$$

$$= U_0^n - \min\{\sum_{\tau=1}^{t} \bar{X}_\tau^n, Q^n\} + E^{n,s(n)} * \min\{\sum_{\tau=1}^{t-1} \bar{X}_\tau^{s(n)}, Q^{s(n)}\} - (W_0^n/\alpha^n) + 1 - \epsilon \geq 0$$

and hence $\langle U_0, X\rangle$ satisfies (3.2). By construction, $\langle X\rangle$ satisfies (3.3). We have hence proved that $\langle U_0, X\rangle$ satisfies (3.1)–(3.4). We next show that $\langle U_0, X\rangle$ is a feasible partial solution to (2.1)–(2.2), (2.3)′, (2.4)–(2.6), and (2.7)′. To construct a complement to $\langle U_0, X\rangle$, we let U_t^n, V_t^n, Y_t^n, and W_t^n ($n = 1, \ldots, N$; $t = 1, \ldots, T$) be defined according to (2.1)–(2.2) and (2.5)–(2.6). If $U_{t-1}^n + Y_{t-1}^n < X_t^n$, then $U_{t-1}^n - X_t^n + Y_{t-1}^n \leq -1$ since all the variables involved are integers. As a result

$$U_0^n - \sum_{\tau=1}^{t} X_\tau^n + E^{n,s(n)} \sum_{\tau=1}^{t-1} X_\tau^{s(n)} - (W_0^n/\alpha^n) + 1 - \epsilon \leq U_0^n - \sum_{\tau=1}^{t} X_\tau^n + \lceil E^{n,s(n)} \sum_{\tau=1}^{t-1} X_\tau^{s(n)} - (W_0^n/\alpha^n)\rceil + 1 - \epsilon$$

$$= U_{t-1}^n - X_t^n + Y_{t-1}^n + 1 - \epsilon \leq -1 + 1 - \epsilon = -\epsilon < 0.$$

A contradiction arises and hence $U_{t-1}^n + Y_{t-1}^n \geq X_t^n$. By straightforward substitution, we can show that $\langle U, V, W, X, Y\rangle$ satisfies (2.3)′ and (2.4). The second half of the lemma is then proved. □

LEMMA A.2. *If $\langle U, V, W, X, Y\rangle$ satisfies* (2.1), (2.2), (2.3)′, (2.4), (2.5), (2.6) *and* (2.7)′, *then $\langle U_0\rangle$ is a feasible partial solution to* (2.1)–(2.7).

Appendix 2. Proof of Theorem 4.1

[PROOF] Let $\langle U_0, X\rangle$ satisfy (3.1)–(3.4). Since all the variables involved are integral, $0 \leq (W_0^n/\alpha^n) < 1$, $0 < -(W_0^n/\alpha^n) + 1 - \epsilon < 1$ and $E^{n,s(n)} = 1$, we conclude that $\langle U_0, X\rangle$ also satisfies (4.1)–(4.4).

Let $\langle U_0, \bar{X}\rangle$ satisfy (4.1)–(4.4). We want to show $\langle U_0\rangle$ is a feasible partial solution to (3.1)–(3.4). To construct a complement to $\langle U_0\rangle$, we let $X_1^n = \lceil \bar{X}_1^n\rceil$ and $X_t^n = \lceil \sum_{\tau=1}^{t} \bar{X}_\tau^n\rceil - \lceil \sum_{\tau=1}^{t-1} \bar{X}_\tau^n\rceil$ for $t = 2, \ldots, T$. It follows that $\sum_{\tau=1}^{t} X_\tau^n = \lceil \sum_{\tau=1}^{t} \bar{X}_\tau^n\rceil$ and X_t^n is a nonnegative integer. If $V_0^n + \sum_{\tau=1}^{t} X_\tau^n - \sum_{\tau=1}^{t} X_\tau^{s(n)} < 0$, then $V_0^n + \sum_{\tau=1}^{t} X_\tau^n - \sum_{\tau=1}^{t} X_\tau^{s(n)} \leq -1$ since all the variables involved are integral. Then,

$$V_0^n + \sum_{\tau=1}^{t} \bar{X}_\tau^n - \sum_{\tau=1}^{t} \bar{X}_\tau^{s(n)} \leq (V_0^n + \sum_{\tau=1}^{t} X_\tau^n - \sum_{\tau=1}^{t} X_\tau^{s(n)}) + (\sum_{\tau=1}^{t} X_\tau^{s(n)} - \sum_{\tau=1}^{t} \bar{X}_\tau^{s(n)}) \leq -1 + (\lceil \sum_{\tau=1}^{t} \bar{X}_\tau^{s(n)}\rceil - \sum_{\tau=1}^{t} \bar{X}_\tau^{s(n)}) < 0,$$

which contradicts (4.1) for $\langle U_0, \bar{X}\rangle$. Hence, we conclude that $V_0^n + \sum_{\tau=1}^{t} X_\tau^n - \sum_{\tau=1}^{t} X_\tau^{s(n)} \geq 0$ and $\langle U_0, X\rangle$ satisfies (4.1). Similarly, we can show that $\langle U_0, X\rangle$ satisfies (4.2). It follows from the construction of $\langle X\rangle$ that

$$X_t^n = \lceil \sum_{\tau=1}^{t} \bar{X}_\tau^n - \lceil \sum_{\tau=1}^{t-1} \bar{X}_\tau^n\rceil\rceil = \lceil \bar{X}_t^n - (\lceil \sum_{\tau=1}^{t-1} \bar{X}_\tau^n\rceil - \sum_{\tau=1}^{t-1} \bar{X}_\tau^n)\rceil \leq \lceil \bar{X}_t^n\rceil \leq \beta_t^n;$$

hence, $\langle U_0, X\rangle$ satisfies (4.3). Noting that all the variables in $\langle U_0, X\rangle$ are integral, $0 \leq (W_0^n/\alpha^n) < 1$, $0 < -(W_0^n/\alpha^n) + 1 - \epsilon < 1$ and $E^{n,s(n)} = 1$, we conclude that $\langle U_0, X\rangle$ also satisfies (3.1)–(3.4). □

Appendix 3. Proof of Theorem 4.2

[PROOF] Note that $0 \leq \bar{W}_0^n - \bar{\bar{W}}_0^n \leq e^{n,s(n)}\alpha^{s(n)} - 1$, or equivalently, $0 \leq (\bar{W}_0^n/\alpha^n) - (\bar{\bar{W}}_0^n/\alpha^n) \leq E^{n,s(n)} - (1/\alpha^n)$. Also note that $0 < \epsilon < (1/\alpha^n) \leq E^{n,s(n)}$.

Let $\langle U_0, X\rangle$ satisfy (3.1)–(3.4) with $\langle W_0\rangle = \langle \bar{W}_0\rangle$. It follows from (3.1) that

$$(\bar{\bar{W}}_0^n/\alpha^n) + V_0^n + \sum_{\tau=1}^{t} X_\tau^n - E^{n,s(n)} \sum_{\tau=1}^{t} X_\tau^{s(n)} = (\bar{W}_0^n/\alpha^n) + V_0^n + \sum_{\tau=1}^{t} X_\tau^n - E^{n,s(n)} \sum_{\tau=1}^{t} X_\tau^{s(n)}$$

$$- ((\bar{W}_0^n/\alpha^n) - (\bar{\bar{W}}_0^n/\alpha^n)) \geq -((\bar{W}_0^n/\alpha^n) - (\bar{\bar{W}}_0^n/\alpha^n)) \geq -(E^{n,s(n)} - (1/\alpha^n)).$$

Since $(\bar{\bar{W}}_0^n/\alpha^n)$, V_0^n, X_τ^n (for $\tau = 1, \ldots, t$) and $E^{n,s(n)}X_\tau^{s(n)}$ (for $\tau = 1, \ldots, t$) are all integral multiples of $E^{n,s(n)}$, we conclude that $(\bar{\bar{W}}_0^n/\alpha^n) + V_0^n + \sum_{\tau=1}^{t} X_\tau^n - E^{n,s(n)} \sum_{\tau=1}^{t} X_\tau^{s(n)} \geq 0$. It is straightforward to show that

$$U_0^n - \sum_{\tau=1}^{t} X_\tau^n + E^{n,s(n)} \sum_{\tau=1}^{t-1} X_\tau^{s(n)} - (\bar{\bar{W}}_0^n/\alpha^n) + 1 - \epsilon \geq 0.$$

Let $\langle U_0, X\rangle$ satisfy (3.1)–(3.4) with $\langle W_0\rangle = \langle \bar{\bar{W}}_0\rangle$. It follows from (3.2) that $U_0^n - \sum_{\tau=1}^{t} X_\tau^n + E^{n,s(n)} \sum_{\tau=1}^{t-1} X_\tau^{s(n)} - (\bar{\bar{W}}_0^n/\alpha^n) + 1 \geq E^{n,s(n)}$ since U_0^n, X_τ^n (for $\tau = 1, \ldots, t$), $E^{n,s(n)}X_\tau^{s(n)}$ (for $\tau = 1, \ldots, t-1$), $(\bar{\bar{W}}_0^n/\alpha^n)$ and 1 are all integral multiples of $E^{n,s(n)}$ while $0 < \epsilon < (1/\alpha^n) \leq E^{n,s(n)}$. Consequently,

$$U_0^n - \sum_{\tau=1}^{t} X_\tau^n + E^{n,s(n)} \sum_{\tau=1}^{t-1} X_\tau^{s(n)} - (\bar{W}_0^n/\alpha^n) + 1 - \epsilon = U_0^n - \sum_{\tau=1}^{t} X_\tau^n + E^{n,s(n)} \sum_{\tau=1}^{t-1} X_\tau^{s(n)} - (\bar{\bar{W}}_0^n/\alpha^n) + 1 - \epsilon$$

$$+ (\bar{\bar{W}}_0^n/\alpha^n) - (\bar{W}_0^n/\alpha^n) \geq E^{n,s(n)} - \epsilon - E^{n,s(n)} + (1/\alpha^n) > 0.$$

It is straightforward to show that

$$(\bar{W}_0^n/\alpha^n) + V_0^n + \sum_{\tau=1}^{t} X_\tau^n - E^{n,s(n)} \sum_{\tau=1}^{t} X_\tau^{s(n)} \geq 0. \quad \square$$

Appendix 4. Proof of Theorem 4.3

[PROOF] We shall prove only the third part of the theorem. Let $\langle \bar{U}_0, \bar{X} \rangle$ satisfy (3.1)–(3.2), (3.3)′ and (3.4)′. Define $U_0^n = \lceil \bar{U}_0^n \rceil + 1$, $X_1^n = \lceil \bar{X}_1^n \rceil$ and $X_t^n = \lceil \sum_{\tau=1}^{t} \bar{X}_\tau^n \rceil - \lceil \sum_{\tau=1}^{t-1} \bar{X}_\tau^n \rceil$ for $t = 2, \ldots, T$. We want to show that $\langle U_0, X \rangle$ satisfies (3.1)–(3.4). If $(W_0^n/\alpha^n) + V_0^n + \sum_{\tau=1}^{t} X_\tau^n - E^{n,s(n)} \sum_{\tau=1}^{t} X_\tau^{s(n)} < 0$, then

$$(W_0^n/\alpha^n) + V_0^n + \sum_{\tau=1}^{t} X_\tau^n - E^{n,s(n)} \sum_{\tau=1}^{t} X_\tau^{s(n)} \leq -E^{n,s(n)}.$$

Hence,

$$(W_0^n/\alpha^n) + V_0^n + \sum_{\tau=1}^{t} \bar{X}_\tau^n - E^{n,s(n)} \sum_{\tau=1}^{t} \bar{X}_\tau^{s(n)} \leq -E^{n,s(n)} + E^{n,s(n)} (\lceil \sum_{\tau=1}^{t} \bar{X}_\tau^{s(n)} \rceil - \sum_{\tau=1}^{t} \bar{X}_\tau^{s(n)}) < 0,$$

which contradicts (3.1) for $\langle \bar{U}_0, \bar{X} \rangle$. Thus, we conclude that $\langle U_0, X \rangle$ satisfies (3.1). Since $\langle \bar{U}_0, \bar{X} \rangle$ satisfies (3.2), it follows that

$$U_0^n - \sum_{\tau=1}^{t} X_\tau^n + E^{n,s(n)} \sum_{\tau=1}^{t-1} X_\tau^{s(n)} - (W_0^n/\alpha^n) + 1 - \epsilon \geq 1 - (\lceil \sum_{\tau=1}^{t} \bar{X}_\tau^n \rceil - \sum_{\tau=1}^{t} \bar{X}_\tau^n) > 0$$

and we conclude that $\langle U_0, X \rangle$ satisfies (3.2). $\square$

Appendix 5. Proof of Theorem 4.4

[PROOF] Let $\langle U_0, X \rangle$ and $\langle \bar{U}_0, \bar{X}_0 \rangle$ be the optimal solutions to (M0) and (M2), respectively. Since $\sum_{n=1}^{N} C^n \bar{U}_0^n \leq \sum_{n=1}^{N} C^n U_0^n \leq \sum_{n=1}^{N} C^n (\lceil \bar{U}_0^n \rceil + 1) < \sum_{n=1}^{N} C^n (\bar{U}_0^n + 2)$, we conclude that $Z_2 - Z_0 < 2 \sum_{n=1}^{N} C^n$. Since $\langle U_0, X \rangle$ satisfies (3.1) and (3.2),

$$(W_0^n/\alpha^n) + V_0^n + \sum_{\tau=1}^{t-1} X_\tau^n - E^{n,s(n)} \sum_{\tau=1}^{t-1} X_\tau^{s(n)} \geq 0 \quad \text{and} \quad U_0^n - \sum_{\tau=1}^{t} X_\tau^n + E^{n,s(n)} \sum_{\tau=1}^{t-1} X_\tau^{s(n)} - (W_0^n/\alpha^n) + 1 - \epsilon \geq 0.$$

By adding the previous two inequalities together, we have $U_0^n + V_0^n - X_t^n + 1 - \epsilon \geq 0$. Since U_0^n, V_0^n and X_t^n are all integral and $0 < 1 - \epsilon < 1$, we have $U_0^n + V_0^n \geq X_t^n$. Since $U_0^n + V_0^n \geq X_t^n$ holds for all $t = 1, \ldots, T$, we have $U_0^n + V_0^n \geq (\sum_{\tau=1}^{T} X_\tau^n)/T \geq Q^n/T$. It follows that $Z_0 \geq \sum_{n=1}^{N} C^n Q^n/T$ and $(Z_2 - Z_0)/Z_0 < 2 \sum_{n=1}^{N} C^n / (\sum_{n=1}^{N} C^n Q^n/T)$. By the definition of Q^n, $Q^n/T \to \infty$ as $Q^0/T \to \infty$. Hence, $(Z_2 - Z_0)/Z_0 \to 0$ as $Q^0/T \to \infty$. $\square$

References

GABBAY, HENRY, "Multi-Stage Production Planning," *Management Sci.*, 25, 11 (November 1979), 1138–1148.

KIMURA, OSAMU AND HIROSUKE TERADA, "Design and Analysis of Pull System, A Method of Multi-Stage Production Control," *Internat. J. Production Res.*, 19, 3 (1981), 241–253.

KRAJEWSKI, LEE J., BARRY E. KING, LARRY P. RITZMAN AND DANNY S. WONG, "Kanban, MRP and Shaping the Production Environment," Working Paper Series 83-19, College of Administrative Science, The Ohio State University, April 1983.

MONDEN, YASUHIRO, "What Makes the Toyota Production System Really Tick?", *Industrial Engineering*, 13, 1 (January 1981a), 36–46.

———, "Smoothed Production Lets Toyota Adapt to Demand Changes and Reduce Inventory," *Industrial Engineering*, 13, 6 (August 1981b), 42–51.

———, "How Toyota Shortened Supply Lot Production Time, Waiting Time, and Conveyance Time," *Industrial Engineering*, 13, 9 (September 1981c), 22–30.

RICE, JAMES W. AND TAKEO YOSHIKAWA, "A Comparison of Kanban and MRP Concepts for the Control of Repetitive Manufacturing Systems," *Production and Inventory Management*, 23, 1 (First Quarter 1982), 1–13.

SCHONBERGER, RICHARD J., *Japanese Manufacturing Techniques: Nine Hidden Lessons in Simplicity*, The Free Press, New York, 1982.

SUGIMORI, Y., K. KASUNOKI, F. CHO AND S. UCHIKAWA, "Toyota Production System and Kanban System, Materialization of Just-in-Time and Respect-for-Human System," *Internat. J. Production Res.*, 15, 6 (1977), 553–564.