A Positive Linear Discrete-Time Model Of Capacity Planning and Its Controllability Properties

L. Caccetta [1], L.R. Foulds [2] and V.G. Rumchev [1]

[1] School of Mathematics and Statistics
Curtin University of Technology
GPO Box U 1987, Perth, WA 6845, Australia
caccetta@cs.curtin.edu.au
rumchevv@cs.curtin.edu.au

[2] Department of Management Systems
University of Waikato
Private Bag 3105, Hamilton 2020, New Zealand
lfoulds@waikato.ac.nz

Abstract One of the most important concepts in production planning is that of the establishment of an overall, or aggregate production plan. In this paper we consider the problem of establishing an aggregate production plan for a manufacturing plant. A new dynamic model is developed. The basic issue is, given a set of production demands stated in some common unit, what levels of resources should be provided in each period? There has been a long history of academic research on aggregate planning, resulting in many mathematical programming models and in a variety of heuristics. However, as the firms attempt to implement manufacturing planning and control systems they find serious deficiencies in these models and heuristics. We attempt to overcome some of these drawbacks with a new approach, utilizing concepts arising in positive linear systems (PLS) theory. Applying recent results concerning PLS we are able to analyze controllability properties of the simplified model. Controllability is a property of the system that shows its ability to move in space. It is a fundamental property with direct implications not only in dynamic optimization problems (such as those arising in inventory and production control) but also in feedback control problems. We provide a number of interesting insights into capacity planning concerning controllability of the system and at the same time formulate some problems regarding controllability of stationary and non-stationary PLS with linear constraints.

1. THE MODEL

Our aim is to meet the pre-specified demand taking into account decisions concerning when to hire and fire, how much inventory to hold, when to use overtime and undertime, and set-up times. We adopt a common unit of production hours. We now introduce the model.

1.1 Dynamics Equations

For $t = 0, 1, 2, \ldots, T\text{-}1$,

$$I_{t+1} = \beta_t I_t + \gamma_t X_t + \delta_t O_t \qquad (1)$$

$$W_{t+1} = \alpha_t W_t + H_t, \qquad (2)$$

where

$$0 \le \alpha_t < 1, \; 0 \le \beta_t < 1, 0 \le \gamma_t \le 1, \qquad (3)$$

t is the time period (usually a week or a month) and T is the number of time periods in the *planning horizon*. In the difference equations (1)-(2) the *state variables* I_t and W_t, the *decision variables* X_t, O_t and H_t, and the *parameters* α_t, β_t, γ_t and δ_t of the production system have the following meaning:

W_t = the number of people employed in month t;

I_t = the hours stored in inventory at the end of month t.

X_t = the regular time production hours scheduled in month t;

O_t = the overtime production hours scheduled in month t;

H_t = the number of employees hired at the end of month t for work in month $(t+1)$;

α_t = the fraction of employees employed in month t that are retained in the month $(t+1)$, the *survival coefficient*;

β_t = the fraction of the total of the hours stored in inventory at the end of month t which is stored in inventory at the end of month $(t+1)$, the *storage coefficient*;

γ_t = the fraction of regular time production hours scheduled in month t which are stored in inventory in month $(t+1)$;

δ_t = the fraction of overtime production hours scheduled in month t which are stored in inventory in month $(t+1)$.

The coefficients α_t (survival), β_t (storage), δ_t and γ_t have an attractive economic interpretation and are quite helpful in the planning process. They are used in the model as exogenous parameters characterizing the production system but their role in the process of decision-making is, clearly, important since they (their values) determine the system evolution. Note also that in (2) $\alpha_t W_t$ is equal to the number of employees employed in month t that are retained in month $(t+1)$, and therefore $(1 - \alpha_t)W_t$ is equal to the number of employees fired in month $(t+1)$. Furthermore, it is not difficult to see from (1) that the hours of production sold in month t is equal to $(1-\beta_t)I_t + (1-\gamma_t)X_t + (1-\delta_t)O_t$.

1.2 Constraints

$$X_t - A_{1t} W_t + U_t = 0 \qquad (4)$$

$$O_t - A_{2t} W_t + S_t = 0 \qquad (5)$$

$$I_t - B_t \geq 0, \; t = 0, 1, 2, \dots, m\text{-}1, \qquad (6)$$

where

U_t = the number of idle time regular production hours in month t;

S_t =the number of idle time overtime production hours in month t;

B_t = the minimum number of hours to be stored in inventory in month t;

A_{1t} = the maximum number of regular time hours to be worked per employee per month;

A_{2t} = the maximum number of overtime hours to be worked per employee per month.

Clearly from their meaning all the state and decision variables as well as the parameters introduced above are *non-negative* so that

$$I_t, X_t, O_t, W_t, H_t, U_t, S_t, B_t, A_{1t}, A_{2t} \geq 0, \quad t = 0, 1, 2, \dots, m\text{-}1. \qquad (7)$$

The restrictions (6) on the production system dynamics are mixed constraints imposed on the state and decision variables for every time period t. The number U_t of idle time regular production hours in month t, the number of idle time S_t overtime production hours in month t and the minimum number B_t of hours to be stored in inventory in month t are assumed to be exogenous parameters in the model.

1.3 Boundary Conditions

$$W_0 = A_3 \geq 0 \qquad (8)$$

$$I_0 = A_4 \geq 0 \qquad (9)$$

$$W_T = A_5 \geq 0 \qquad (10)$$

$$I_T = A_6 \geq 0, \qquad (11)$$

where

A_3 = the initial employment level;

A_4 = the initial inventory level;

A_5 = the desired number of employees in month T (the last month of the planning horizon);

A_6 = the desired inventory level at the end of month T;

The states W_0 and I_0 are called initial states, and the states W_T and I_T are final (terminal) states.

1.4 Assumptions

The dynamic model (1)-(11) described above is introduced under the following assumptions. In any month t:

- All regular time employees work overtime.
- Only existing regular time employees work overtime.
- All employees work the same number of regular time hours, up to the limit A_{1t}.
- All employees work the same number of overtime hours, up to the limit A_{2t}.

2. POSITIVE LINEAR SYSTEMS

The dynamic equations (1)-(2) can be rewritten in the matrix form

$$\begin{bmatrix} W_{t+1} \\ I_{t+1} \end{bmatrix} = \begin{bmatrix} \alpha_t & 0 \\ 0 & \beta_t \end{bmatrix} \begin{bmatrix} W_t \\ I_{t+1} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & \gamma_t & \delta_t \end{bmatrix} \begin{bmatrix} H_t \\ X_t \\ O_t \end{bmatrix}$$

$$t = 0, 1, 2, \dots, T\text{-}1, \qquad (12)$$

or, respectively,

$$x(t+1) = A(t)\,x(t) + B(t)\,u(t),$$
$$t = 0, 1, 2,\dots, T\text{-}1, \qquad (13)$$

where the vector of state variables $x(t)$, the decision (control) vector $u(t)$, the system matrix $A(t)$ and the control matrix $B(t)$ are given by the corresponding vectors and matrices in (12). Note that all of the entries of $u(t)$, $A(t)$ and $B(t)$ are greater than or equal to zero for any time period t. Vectors and matrices with nonnegative entries are called nonnegative vectors and matrices, see Berman and Plemmons [1994]. They are denoted as $u(t) \geq 0$ and $A(t) \geq 0$, respectively. Since the system matrix $A(t) \geq 0$, the control matrix $B(t) \geq 0$ and the decision vector $u(t) \geq 0$ are nonnegative for any t, it can be seen from (12) (and (13)) that the state vector $x(t)$ is a nonnegative vector whenever the initial state

$$x(0) = \begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} = \begin{bmatrix} W_o \\ I_o \end{bmatrix} = \begin{bmatrix} A_3 \\ A_4 \end{bmatrix} \geq 0 \qquad (14)$$

is nonnegative. Thus, nonnegativity (positivity) is an intrinsic property of the system (12), that is (13). Such systems are called *positive systems*, see, for example, Luenberger [1979]. It can be proved that the conditions $u(t) \geq 0$, $A(t) \geq 0$ and $B(t) \geq 0$ are necessary and sufficient for the *state trajectory* $\{x(t)\}$ to be nonnegative for any t. Note also that the nonnegativity of the decision variables $u_1(t) = H_t$, $u_2(t) = X_t$ and $u_3(t) = O_t$ guarantees the nonnegativity of the state variables in the mixed functional constraints (4)-(5). Note that the final (terminal) state

$$x(T) = \begin{bmatrix} x_1(T) \\ x_2(T) \end{bmatrix} = \begin{bmatrix} W_T \\ I_T \end{bmatrix} = \begin{bmatrix} A_5 \\ A6 \end{bmatrix} \geq 0$$

is non-negative too.

The (dynamic) system theory for positive systems has been rapidly developing during the last decade although one of the cornerstones of this theory is the famous Frobenius-Perron theorem for nonnegative matrices known for over 80 years, see Berman and Plemmons [1994] or Luenberger [1979]. The Frobenius-Perron theorem plays a fundamental role in mathematical economics, input-output analysis, economic dynamics, probability theory and mathematical statistics, and any linear theory involving positivity.

The dynamic model for capacity planning (1)-(11) can be built-in in decision support systems. It is somewhat easier for simulation and decision-making than the static models. On the other hand, introducing a relevant objective (cost) function we can consider the related optimal control problem and determine the optimal decision sequences and the corresponding optimal state trajectory over the planning horizon T. But because of the restrictions (8)-(11) the related problem will be a two-point boundary-value problem and as it is well known a solution to such a problems might not exist.

The optimal control approach to the theory of the firm is motivated by three issues : (i) the need for policies, (ii) the contribution of deductive analysis, and (iii) the need to incorporate time. Van Hilton *at al* [1993] have well exposed the state-of-the art of this area but they discuss only continuous-time systems and exploit Pontryagin Maximum Principle developed for such systems. They do not consider positive systems as well as discrete-time models. Discrete-time models are somewhat more suitable to describe the firm's dynamics. Moreover, the model (1)-(11) not only represents a discrete-time positive system but it contains a number of important parameters not included in the dynamic models described in the literature. At the same time, the first question that arises when solving any two-boundary optimal control problem is whether a solutions exists. This question is closely related to the controllability properties of the system. Unfortunately not much attention to date is paid to controllability of the dynamic models of the firm as it is evident from Luenberger [1979] and van Hilton *at al* [1993]. In the next section we study controllability properties of the discrete-time positive model (12), i.e. (13), with non-negative decision sequence $u(t) \geq \mathbf{0}$.

3. REACHABILITY AND CONTROLLABILITY OF POSITIVE SYSTEMS

3.1 Some Definitions

The definitions of reachability, null-controllability and controllability introduced by Rumchev and James [1989] for stationary (time-invariant) positive systems are extended below for non-stationary (time-variant) positive systems.

The system (13) (and the non-negative pair $(A(t),B(t)) \geq 0$) is said to be

(a) *reachable* (or controllable-from-the origin) if for any non-negative state $x \in \Re_+^n$, $x \neq \mathbf{0}$, and some finite t there exists a non-negative control sequence $\{u(s), s = 0,1,, \ldots \ldots, t-1\}$ that transfers the system from the origin into the state $x = x(t)$;

(b) *null-controllable* (or controllable-to-the origin) if for any non-negative state $x \in \Re_+^n$ and some finite t there exists a non-negative control sequence $\{u(s), s = 0,1,, \ldots \ldots, t-1\}$ that transfers the system from the state $x = x(0)$ into the origin;

(c) *controllable* if for any non-negative pair $\{x_o, x\}$ $x \in \Re_+^n$ and some finite t there exists a non-negative control sequence $\{u(s), s = 0,1,, \ldots \ldots, t-1\}$ that transfers the system from the state $x_o = x(0)$ into the state $x = x(t)$.

Rumchev and James [1989] have proved that the discrete-time positive linear system is controllable if and only if it is reachable and null-controllable. Thus controllability implies both reachability and null-controllability and, vice versa, reachability and null-controllability *together* imply controllability. Reachability, null-controllability and controllability are general properties of the system (but not of its environment). They express the *ability* of the system to move in space, this being the non-negative orthant for the class of positive systems under consideration.

The t-step *reachability matrix* of the pair $A(t),B(t)$ (and the system (13)) is defined, see Sontag [1998], as

$$\Re_t(A(t), B(t)) = [B(t\text{-}1)\ \ A(t\text{-}1)B(t\text{-}2)\ A(t\text{-}1)\,A(t\text{-}2)B(t\text{-}3)\ \ldots\ A(t\text{-}1)\,A(t\text{-}2)\,A(t\text{-}3)\ldots A(2)A(1)\,B(0)] \qquad (15)$$

The reachability matrix of the system (13) is clearly a non-negative matrix since $A(t) \geq 0$ and $B(t)) \geq 0$ for any integer t.

For stationary system, i.e. $A(t) = A$ and $B(t) = B(t)$ are constant real matrices, the t-step reachability matrix becomes

$$\Re_t(A, B) = [B\ \ AB\ \ldots\ A^{t\text{-}1}B] \qquad (16)$$

A vector (column, row) with exactly one non-zero entry is called *monomial*, see Berman and Plemmons [1994]. The product of a non-singular diagonal matrix and a permutation matrix is call a *monomial matrix*. Any monomial matrix consists of linearly independent monomial columns. A monomial vector is called i-monomial if the non-zero entry is in the ith position.

3.2 Criteria for Stationary Systems

The following criteria for identifying the reachability and controllability properties of stationary discrete-time positive linear systems can be found in Caccetta and Rumchev [1999].

The non-negative pair $(A, B) \geq \mathbf{0}$ is

(i) reachable if and only if the n-step reachability matrix $\Re_n(A, B)$ contains an $n{\times}n$ monomial submatrix;

(ii) null-controllable if and only if A is a nil-potent matrix;

(iii) controllable (in finite time) if and only if it is reachable and null-controllable.

Reachability and controllability are generic properties of the pair $(A, B) \geq \mathbf{0}$ (and the positive system). These properties depend on the structure of the pair (A, B) and are not affected by the values of the entries of A and B, that is of the parameters of the positive system.

The analysis of reachability and controllability properties of positive linear systems (even time-invariant) with additional linear constraints on state and decision variables is more complicated. It is still an open problem.

3.3 Non-Stationary Systems: the Capacity Planning Model

There are no results in the literature on reachability and controllability of non-stationary positive linear systems even without additional linear constraints on state and decision variables. Such a class of systems is more difficult to study. Therefore, to get some insight into the problem we consider in this subsection the *simplified* (without the restrictions (4)-(6)) capacity planning *model* presented in Section 1, that is the model (12) with the non-negativity restrictions. The system and control matrices in (12) are, respectively,

$$A(t) = \begin{bmatrix} \alpha_t & 0 \\ 0 & \beta_t \end{bmatrix}, \text{ and } B(t) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \gamma_t & \delta_t \end{bmatrix}. \qquad (17)$$

The state of the model (13) (and (12)) at time t can be represented as

$$x(t) = A(t\text{-}1)\,A(t\text{-}2)\,A(t\text{-}3)\ldots A(2)A(1)A(0)\,x(0) + \Re_t(A(t), B(t))\,u_t, \qquad (18)$$

where $u_t = [u'(t-1), \ldots, u'(1), u'(0)]'$ is the expanded decision vector and " ' " denotes the transposed vector. For *reachability* $x(0) = \mathbf{0}$, and the expression (18) becomes

$$x(t) = \Re_t(A(t), B(t))\, u_t . \qquad (19)$$

Since the expanded decision vector $u_t \geq \mathbf{0}$ and the reachability matrix $\Re_t(A(t), B(t)) \geq \mathbf{0}$ are non-negative the non-negative quadrant (in the case under consideration) can be spanned if and only if $\Re_t(A(t), B(t))$ contains two linearly independent monomial columns. Then, it readily follows from the structure of (15) and the form of $B(t)$ given by (17) that matrix $\Re_t(A(t), B(t))$ contains two linearly independent monomial columns (1, 0) and (0, γ_{t-1}) (or (0, δ_{t-1})), and any non-negative state can represented as

$$x(t) = c_1 e_1 + c_2 e_2 \ , \quad \text{with} \quad c_1, c_2 \geq 0 \qquad (20)$$

where e_1 and e_2 are the basis unit vectors, and c_1 and c_2 are some (non-negative) constants. In other words, any non-negative state can be reached from the origin by a suitably chosen non-negative decision sequence in finite time - the *positive system* (12) *is reachable*. As a matter of fact any state of the model (12) can be reached from the origin in at most two steps so that the *reachability index* (see, for example, Sontag [1998]) of the pair $A(t), B(t) \geq \mathbf{0}$ is equal to two.

For *null-controllability* $x(t) = \mathbf{0}$ and $x_o = x(0) \neq \mathbf{0}$ so that the equation (18) becomes

$$\mathbf{0} = A(t-1)\,A(t-2)\,A(t-3) \ldots A(2)A(1)A(0)\,x_o + \Re_t(A(t), B(t))\, u_t . \qquad (22)$$

Rumchev and James [1989] have shown that for positive systems the decision sequence does not contribute to speed up the system to the origin so that $u_t = \mathbf{0}$, and hence the equation (22) can be reduced to

$$\mathbf{0} = A(t-1)\,A(t-2)\,A(t-3) \ldots A(2)A(1)A(0)\,x_o \qquad (23)$$

or

$$\mathbf{0} = \Phi(t)\,x_o, \qquad t = 0, 1, \ldots, T-1 \qquad (24)$$

where

$$\Phi(t) = \prod_{k=0}^{t-1} A(k) = \begin{bmatrix} \alpha_t^c & 0 \\ o & \beta_t^c \end{bmatrix} \text{ with}$$

$$\alpha_t^c = \prod_{k=0}^{t-1} \alpha_k < 1 \text{ and } \beta_t^c = \prod_{k=0}^{t-1} \beta_k < 1 \qquad (25)$$

is the fundamental matrix of the system (12). Let now $\alpha_k = 0$ for some $k = s_1$ and $\beta_r = 0$ for some $r = s_2$, and let $s = \max\{s_1, s_2\}$. Then the matrix $\Phi(s+1) = \mathbf{0}$ and the equation (24) is satisfied for any non-negative state x_o and any $t > s$ which implies the null-controllability (in finite time) of the non-stationary positive system (12) is, and since the system is reachable it is also (finite time) *controllable*. But having $\alpha_t = 0$ or $\beta_t = 0$ for some t means that all the employees are fired or, respectively, no production is stored in inventory at the end of the time-period t. Such a situation, clearly, does not seem quite realistic.

Let at least one of the fractions α_t and β_t be strictly positive for the whole planning horizon. Then the fundamental matrix $\Phi(t) \neq \mathbf{0}$ for $t = 1, \ldots, T$ and the positive system (12) is not null-controllable in finite time. If no control is exersized, the trajctory (free motion) of the system from any initial state $x_o \geq \mathbf{0}$ but $x_o \neq \mathbf{0}$ is given by

$$x(t) = \Phi(t)\,x_o. \qquad (26)$$

Take $\rho = \max\{\alpha_t, \beta_t; t = 0, 1, \ldots, T-1\}$. It readily follows from (3) that $\rho < 1$. Consider now the free motion

$$y(t) = P^t x_o \geq \mathbf{0} \qquad (27)$$

of the time-invariant system

$$y(t+1) = Py(t) \text{ with } P = \text{diag}\{\rho, \rho\} \qquad (28)$$

from the same initial state x_o. Clearly,

$$y(t) \rightarrow \infty \text{ as } t \rightarrow \infty \qquad (29)$$

since $\rho < 1$. On the other hand, since $A(t) \leq P$ for any $t \geq 0$ the trajectory (26) of the model (12) is dominated by the time-invariant positive system trajectory (27),

$$\mathbf{0} \leq x(t) \leq y(t) \quad \text{for any } t. \qquad (30)$$

Then it readily follows from (29) and (30) that

$$x(t) \rightarrow \infty \text{ as } t \rightarrow \infty, \qquad (31)$$

so the free motion of the positive systems (12) converges to the origin asymptotically. This means that the positive system (12) is *weakly* (asymptotically) *null-controllable*, see Caccetta and Rumchev [1999]. Reachability and weak null-controllability together imply weak controllability of the positive system, see again Caccetta and Rumchev [1999]. Thus if at least one of the fractions α_t and β_t be strictly positive for the whole planning horizon the positive system (12) is *weakly controllable*. This result is derived for large T and without taking into account the restriction (4)-(6) imposed on the state and decision variables. A study of reachability and controllability properties of the non-stationary model (12) with the linear constraints (4)-(6) is a subject of a related paper.

4. CONCLUSIONS

In this paper a new discrete-time dynamic model of capacity planning is developed. The model is motivated not only by the need for policies but also by the need to incorporate time and open the way for deductive analysis. Some interesting new characterizations of production systems important for aims of planning and control appear in the model. The model can be built-in in decision support systems. It is somewhat easier for simulation and decision making than the static models.

The model presented in this paper belongs to the class of non-stationary discrete-time positive linear systems. The theory of positive systems has been rapidly developing during the last decade. Criteria for identifying reachability and controllability properties of time-invariant positive linear systems have been obtained quite recently. Reachability and controllability criteria for non-stationary (time-variant) positive systems even without side linear constraints are not known to date. We have studied the controllability properties of the capacity planning model (without the side linear constraints) and obtain interesting results. It turns out that the model is controllable in short term in some not quite realistic situations but it is *weakly controllable* in long term.

Introducing relevant objective functions we can consider the related optimal control problems and determine the optimal decision sequences and the corresponding state trajectories. We study such optimal control problems as well as the reachability properties of the model with the side linear constraints in related papers.

5. REFERENCES

Berman, A., and R. J. Plemmons, *Nonnegative matrices in mathematical sciences*, Philadelphia, SIAM, 1994.

Berry, W.L., Vollman, T.E. and D.C.Whybark (), *Manufacturing planning and control systems*, 3rd ed., Homewood, IL, Business One/Irwin, 1992.

Caccetta, L., and V.G. Rumchev, A survey of reachability and controllability of positive linear systems, *Annals of Operations Research*, (in press), 1999.

van Hilten, O., P.M. Kort, and P.JJ.M. van Loon, *Dynamic policies of the firm: An optimal control approach*, Berlin, Springer-Verlag, 1993.

Luenberger, D.G., *Introduction to Dynamic Systems: Theory, Models & Applications*, New York, Wiley & Sons, 1979.

Ozdamar, L., Bozyel, M.A. and S.L. Birbil, A Hierarchical Decision Support System for Production Planning, *European Journal of Operations Research*, **104**, 403-422, 1998.

Rumchev, V.G and D.J.G. James, Controllability of positive linear systems, *International Journal of Control*, **50** (3), 845-857, 1989.

Sontag, E.D., *Mathematical Control Theory: Deterministic Finite Dimensional Systems*, New York, Springer, 1998.