Journal of Global Optimization **8**: 349–378, 1996.

Transient and Steady-State Analysis of a Manufacturing System with Setup Changes

SHERMAN X. BAI and MOHSEN ELHAFSI
Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL 32611, U.S.A.

(Received: 15 February 1995; accepted: 5 August 1995)

Abstract. This paper deals with the optimal scheduling of a one-machine two-product manufacturing system with setup, operating in a continuous time dynamic environment. The machine is reliable. A known constant setup time is incurred when switching over from a part to the other. Each part has specified constant processing time and constant demand rate, as well as an infinite supply of raw material. The problem is formulated as a production flow control problem. The objective is to minimize the sum of the backlog and inventory costs incurred over a finite planning horizon. The global optimal solution, expressed as an optimal feedback control law, provides the optimal production rate and setup switching epochs as a function of the state of the system (backlog and inventory levels). For the steady-state, the optimal cyclic schedule (Limit Cycle) is determined. This is equivalent to solving a one-machine two-product Lot Scheduling Problem. To solve the transient case, the system's state space is partitioned into mutually exclusive regions such that with each region is associated an optimal control policy. A novel algorithm (Direction Sweeping Algorithm) is developed to obtain the optimal state trajectory (optimal policy that minimizes the sum of inventory and backlog costs) for this last case.

Key words: Dynamic setups, setup and production flow control, optimal control.

1. Introduction

The setup scheduling problem arises in manufacturing systems, where the switch over from one product to another consumes a certain amount of time or incurs a cost. Setup times imply a down time during which there is no production. Cost may be the result of scrap losses due to testing and tolerance adjustment of the machine for the next part.

The setup scheduling problem has received the attention of many researchers due to its importance, since almost no manufacturing system is perfectly flexible. A special version of the setup scheduling problem is known as the Economic Lot Scheduling Problem (ELSP). Elmaghraby (1978) gives a thorough review of the models used for the ELSP. These models have focused on combinatorial optimization including mixed integer programming formulations. The typical objective reflected in those models, is to schedule a number of jobs with fixed processing times on a set of machines, so as to minimize some performance measure. The latter includes flow time, make span, tardiness, lateness etc. This class of problems is known to be very difficult to solve (i.e., their computation time grows exponentially with the size of the problem). Recent research in the field includes the work of Dobson (1987), Goyal (1984), and Carreno (1990). The purpose of their effort is the scheduling of several products that must be produced on a fixed number of identical reliable machines. There is a constant demand for each product, that must be fulfilled immediately (i.e., no backlog is allowed). Their objective is to minimize the average inventory holding and setup costs per unit time. Gallego (1989) extended the ELSP problem to allow backlog.

The general version of the setup scheduling problem can be formulated as a feedback control problem. The latter must respond to random events so as to minimize a certain criterion. This kind of formulation is usually classified under *production flow control* models. The first production flow control model was introduced by Kimemia and Gershwin (1983) in the early 1980s, where they modeled the movement of parts as a continuous flow and suggested a feedback control of the flow rates of parts through a flexible manufacturing system, in response to machine failures so as to closely track the demand of all parts. Gershwin (1994) gives a more general framework, where he suggests a hierarchical approach for scheduling manufacturing systems. He groups the events from the least frequent (at the top level of hierarchy) to the most frequent ones (at the bottom level). If the hierarchy levels are well separated (frequency wise), each level can be formulated as a continuous flow control problem.

Using the formalism of Kimemia and Gershwin (1983) and Gershwin (1994), Sharifnia *et al.* (1991) investigated a single machine setup scheduling problem. They proposed a feedback setup scheduling policy which uses corridors in the surplus (inventory/backlog) space to determine the epochs of setup changes. The corridors are chosen so as to guide the surplus trajectory to a target cycle which they referred to as the *Limit Cycle.* Srivatsan and Gershwin (1990) extended the ideas of Sharifnia *et al.* and developed methods for choosing the parameters of the corridors when the setup frequencies are not all the same. Caramanis *et al.* (1991) derived the optimality conditions for set-up changes and solved them numerically for a two-part type system using a quadratic cost criterion. Hu and Caramanis (1992, 1995) solved the three-part type setup problem numerically and deduced structural properties of the optimal policies. Based on the numerical results, they proposed near-optimal policies. Perkins and Kumar (1989) and Kumar and Seidman (1990) studied the performance of distributed real-time setup scheduling policies and investigated the conditions under which the system remains stable. Connolly (1992) proposes a heuristic for the two-part-type one-machine setup system, based on known results from the non-setup system. Her approach is based on a local optimization that maximizes the progress toward a target Limit Cycle. Bai and Elhafsi (1993) studied the real-time scheduling of an unreliable one-machine two-part-type non-resumable setup system. They provide a continuous dynamic programming formulation of the problem which they discretize and solve numerically. Based on the numerical solution they provide two heuristics to solve the stochastic problem. They also provide necessary and sufficient conditions for demand feasibility and stability of the control policies. Gallego (1989) studied the ELSP problem in the case of a machine subject to disruptions of small magnitude. He shows that the optimal policy selects the production lot sizes as a linear function of the current inventory levels.

In this paper, we study the setup scheduling of a deterministic one-machine two-part-type system within a feedback control framework (Kimemia and Gershwin's framework). The remainder of the paper is organized as follows: In Section 2, we present an optimal control formulation of the setup problem. In Section 3, we provide the optimal solution of the problem in steady-state. In Section 4, the transient solution is obtained by partitioning the surplus/backlog space into two mutually exclusive major regions. In one region the optimal solution is obtained by inspection. In the second region an algorithm that gives the optimal state trajectory is developed. We conclude our study with Section 5.

2. Problem Formulation

We consider a manufacturing system which has a single machine and produces two distinct parts (or products). The system should satisfy a constant demand rate d_i $(i = 1,2)$ for each part. The machine incurs a non-zero setup time when switching from one product to the other. The setup times δ_i $(i = 1,2)$ are given constants. Let $x_i(t)$ be the production surplus (positive or negative) of Part Type i $(i = 1,2)$ at time t; a positive value of $x_i(t)$ represents inventory while a negative value represents backlog. Here, we follow the general framework introduced by Kimemia and Gershwin (1983) and model the production flow as continuous rather than discrete. Let $u_i(t)$ be the controlled production rate of the machine producing Type i parts at time t. Denote by $\sigma_i(t)$ the setup state at time t. It is a binary variable which is 1 when the machine is ready to produce Type i parts, and 0 otherwise. We assume that initially, the machine is not setup to either part type.

System Dynamics and Constraints: The dynamics of the system can now be described by

$$\frac{dx_i(t)}{dt} = u_i(t) - d_i, \quad i = 1,2; \tag{1}$$

$$0 \leq u_i(t) \leq U_i\sigma_i(t), \quad i = 1,2. \tag{2}$$

where U_i is the maximum machine capacity of producing Type i parts. Denote by $\sigma_{i,j}(t)$ the transition index which is 1 if the machine is undergoing a setup change from Type i to Type j parts at time t, and 0 otherwise.

The setup states and the transition indices obey the following set of equations:

$$\sigma_1(t) + \sigma_2(t) + \sigma_{1,2}(t) + \sigma_{2,1}(t) = 1 \tag{3}$$

$$\text{if } \sigma_{i,j}(t^-) = 1 \text{ and } t < \delta_i, \text{ then } \sigma_{i,j}(t) = 1; \tag{4}$$

$$\text{if } \sigma_i(t^-) = 1 \text{ and } \sigma_i(t) = 1, \text{ then } \sigma_{i,j}(t) = 1; \tag{5}$$

$$\text{if } \sigma_{i,j}(t^-) = 1, t > \delta_i \text{ and } \int_{t-\delta_i}^{t-} \sigma_{i,j}(\tau)\mathrm{d}\tau = \delta_i, \text{ then } \sigma_j(t) = 1; \tag{6}$$

$$\text{if } \sigma_{i,j}(t^-) = 1, t > \delta_i \text{ and } \int_{t-\delta_i}^{t-} \sigma_{i,j}(\tau)\mathrm{d}\tau < \delta_i, \text{ then } \sigma_{i,j}(t) = 1. \tag{7}$$

i=1,2, j=1,2, $i \neq j$.

The above equations, (3)–(7), merely state that if $\sigma_i(t) = 1$, we can either continue producing Part Type i, or decide to switch production to Part Type j. In the latter case we must spend exactly δ_j amount of time setting up the machine for Part Type j. That is, $\sigma_{i,j}(t) = 1$ for exactly δ_j amount of time. After the setup change, $\sigma_j(t) = 1$, and the machine is ready to produce Part Type j.

Penalty Function: The instantaneous cost of the system at time t, is the sum of instantaneous costs of all part types at time t. The costs for individual part types are assumed to be piecewise linear and are given by:

$$g_i(x_i(t)) = c_i^+ x_i^+(t) + c_i^- x_i^-(t), \quad \text{for } i = 1, 2 \tag{8}$$

where c_i^+ and c_i^- are the per unit instantaneous inventory and backlog costs respectively, and $x_i^+(t) = \max\{x_i(t), 0\}$ and $x_i^-(t) = \max\{-x_i(t), 0\}$. The total instantaneous cost is given by

$$g(x(t)) = \Sigma_{i=1}^2 g_i(x_i(t)). \tag{9}$$

This cost function penalizes the system for over producing ($x_1 > 0$, $x_2 > 0$) or under producing ($x_1 < 0$, $x_2 < 0$). Notice that, we do not consider setup costs, but we do include setup times explicitly in the formulation.

State Variables and Control Variables: The state variable of the system is given by $x(t) = (x_1(t), x_2(t))$. The variables $u(t) = (u_1(t), u_2(t))$, and $\sigma(t) = (\sigma_1(t), \sigma_2(t))$ are the control variables. We denote by (σ, u) the complete control vector.

Capacity Set: The capacity set represents the set of feasible production rates, when the setup state is $\sigma(t)$, at time t. It is given by:

$$\Omega(\sigma(t)) = \{u(t) \mid 0 \leq u_i(t) \leq U_i, \sigma_i(t), i = 1, 2\}. \tag{10}$$

Hence, for each setup state we have a different capacity set. These are given below:

$$\begin{aligned} \Omega(0,0) &= \{u(t) \mid u_i(t) = 0, i = 1, 2\} \\ \Omega(1,0) &= \{u(t) \mid 0 \leq u_1(t) \leq U_1, u_2(t) = 0\}, \\ \Omega(0,1) &= \{u(t) \mid u_1(t) = 0, 0 \leq u_2(t) \leq U_2\}. \end{aligned} \tag{11}$$

Setup Constraint Set: The setup constraints set is the set of all possible setup vectors $\sigma(t) = (\sigma_1(t), \sigma_2(t))$ satisfying constraints (3)–(7). Let Θ be this set.

Admissible Control Policies: Let $\Xi(\Theta,\Omega)$ be the set of feasible controls, which depends on Θ and Ω. The set of admissible control policies, A, is the set of all mappings μ, such that:

$$\begin{aligned} \mu &: \Re^2 \rightarrow \Xi(\Omega, \Phi) \\ &\mu(x) \mapsto (\sigma, u) \end{aligned} \tag{12}$$

are piecewise continuously differentiable (e.g., sufficiently smooth). These admissible control policies are feedback controls that specify the control actions (setup and production level of the machine) to be taken, given the state of the system.

Objective Function: The objective is to find a control policy $\mu^* \in$ A, corresponding to a setup control $\sigma^* = ({\sigma^*}_1, {\sigma^*}_2)$ and a production flow rate control $u^* = ({u^*}_1, {u^*}_2)$, that minimizes, for each initial state $x(t)$, the following cost function:

$$J_\mu(x(t), t) = \int_t^{t_f} g(x(s))\mathrm{d}s \tag{13}$$

where the minimization is over all functions $\mu(x(\tau)) = (\sigma(\tau), u(\tau))$, such that $x(\tau)$, $\sigma(\tau)$ and $u(\tau)$ satisfy constraint (1) and $(\sigma(\tau), u(\tau)) \in \Xi\ (\Theta,\Omega)$ for $t \leq \tau \leq t_f$.

The cost function can be divided into two components. One is due to a transient period and the other is due to a steady-state period. The definitions of the transient and steady-state periods will be given in the next section. We will also show in the next section, that once at the steady-state, the system follows an optimal cyclic schedule (called the Limit Cycle) in the state space. That is, starting with the machine setup for Part Type i (i=1,2), the machine commences to produce this Part Type at maximum machine capacity. After t_i amount of time, the machine stops producing, and a setup for Part Type j (j = 1,2, $j \neq i$) is initiated. Once the setup is completed, production of Part Type j begins. After t_2 amount of time, we stop the machine, and start a setup change for Part Type i. At the end of the setup, the system is in the state it started from at the beginning of the cycle. This procedure will repeat itself, until the end of the planning horizon.

Let t_s be the time instant, the system reaches the steady-state. The total cost can then be written as

$$\begin{aligned} J_\mu(x(t), t) &= \int_t^{t_s} g(x(s))\mathrm{d}s + \int_{t_s}^{t_f} g(x(s))\mathrm{d}s \\ &= J_\mu^T(x(t), t) + (t_f - t_s) J_\mu^s(x(t_s)t_s). \end{aligned} \tag{14}$$

We refer to $J_\mu^T(x(t),t)$ as the transient cost,

$$J_\mu^T(x(t), t) = \int_t^{t_s} g(x(s))\mathrm{d}s \tag{15}$$

and $J^S_\mu\ (x(t_s),\ t_s)$ as the average steady-state cost,

$$J^s_\mu(x(t_s),t_s) = \frac{1}{t_f - t_s}\int_{t_s}^{t_f} g(x(s))\mathrm{d}s \tag{16}$$

Here, we assume that $(t_f\ -\ t)$, the planning horizon, is long enough, so that the system reaches the steady-state and stays there for a long period. Also, we assume that the demand set $\{d_1,d_2\}$ is feasible. That is, $U_i > d_1 + d_2$ for $i = 1,2$ (see Bai and Elhafsi (1993), for necessary and sufficient conditions for demand feasibility).

3. Steady-State Solution

If the machine were perfectly flexible (i.e., with zero setup change times), the production surplus can be kept at the zero level. That is, it would be optimal to produce both parts simultaneously at the demand rates. Therefore, if the setup times are negligible, the steady-state solution is very simple and easy to get. Unfortunately, the steady-state solution is not that easy, when the machine is not flexible (i.e. incurs a nonzero setup time before switching to the other part type). The main difficulty is that it is not possible to produce both part types at the same time. Thus, the surplus vector cannot be kept at a constant level and therefore must follow a cyclic schedule. Sharifnia *et al.* (1991) suggest corridor policies to schedule the timing of the setup changes and their frequencies. They show that if the horizon is long enough, and by choosing the appropriate corridors, the surplus levels converge to a cycle, in which parts are produced according to a round robin sequence. They refer to this cycle as the Limit Cycle, since the latter is reached asymptotically. In the remainder of this paper, we will refer to the steady-state solution as the Limit Cycle and vice versa. A formal definition of the steady-state for our problem is given as follows:

DEFINITION 3.1. We say that the system has reached a steady-state, when the surplus levels touch the Limit Cycle at a point where it would be possible for the surplus of both parts to stay on the Limit Cycle. The steady-state solution is completely characterized when the optimal location of the Limit Cycle is known. As can be seen in the next subsection, determining the Limit Cycle is equivalent to solving a one-machine two-part-type LSP problem, since we ignore setup costs. Lot sizes are induced in the model by the explicit presence of setup times.

3.1. OPTIMAL LOCATION OF THE LIMIT CYCLE

In the steady-state, the average cost can be written as follows:

$$J^s_\mu(x(t_s),t_s) \;=\; \frac{1}{t_f - t_s}\int_{t_s}^{t_f} g(x(s))\mathrm{d}s$$

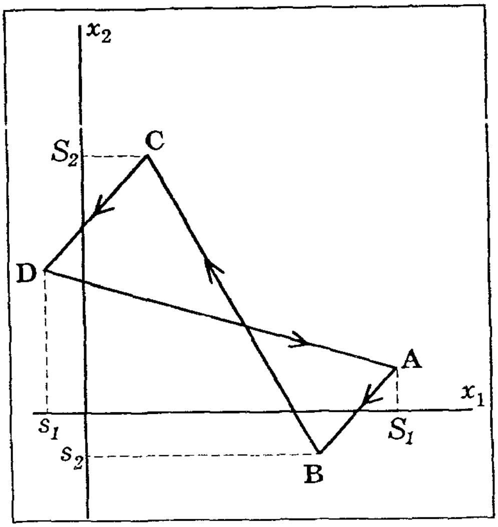

Fig. 1. Location of the limit cycle in x-space.

$$\cong \frac{1}{nT}\left(\Sigma_{i=1}^{n}\int_{(i-1)T}^{iT} g(x(s)ds + \int_{nT}^{t_f} g(x(s))ds\right)$$

$$\cong \frac{1}{T}\int_0^T g(x(s))\mathrm{d}s + \frac{1}{nT}\int_{nT}^{t_f} g(x(s))ds \tag{17}$$

where T is the time duration of the Limit Cycle and $nT \leq t_f - t_s < (n+1)T$ (n a positive integer). The second term in (17) can be neglected, since we assumed that $t_f - t_s$ is large enough (thus n is large) so that the system stays on the Limit Cycle for a long time. It is clear from (17) that, minimizing the total average steady-state cost is equivalent to minimizing the average cost over the Limit Cycle. This, we use as our criterion for the steady-state solution.

If the system is heavily loaded, that is, the demand level is close to the system capacity (see Elhafsi and Bai (1995) for more details), it can be shown that the Limit Cycle has the shape shown in Figure 1. The location of the Limit Cycle is defined by four points A, B, C, and D, in x-space. Notice that in Figure 1, starting at any point on the Limit Cycle, we always come back to the same point, after exactly one cycle. For example, if we started at point A, the cycle would be to set up the machine for Part Type 2, produce Part Type 2, set up the machine for Part Type 1, and then produce Part Type 1 until the surplus vector reaches A. This observation leads to the following fact:

FACT 3.1. starting at a point P on the Limit Cycle, to be able to come back to the same point, the amount of surplus generated during the production of either part type should be equal to that lost during times when we do not produce that part type. In other words the surplus generated during one Limit Cycle should be equal to zero for either part type. The steady-state solution is based on Fact 3.1. Before proceeding with the solution we need the following additional notation:

For Part Type i (i = 1,2):

t_i time spent in Setup i over the Limit Cycle (i.e., producing Part Type i).

T length of the Limit Cycle ($T = t_1 + t_2 + \delta_1 + \delta_2$).

S_i maximum surplus level.

s_i minimum surplus level.

q_i replenishment quantity ($q_i = S_i - s_i$).

Based on Fact 3.1 and assuming that the production rates are constant over the Limit Cycle (we will show that this assumption holds indeed), we can write the following balance equations: where t_i (i = 1,2) is the time spent in setup i and T is the length of the Limit Cycle. Note that the equations hold only if $u_i \geq d_1 + d_2$, (i = 1,2). After rearranging we have

$$\begin{cases} (u_1 - d_1)t_1 - d_1 t_2 = (\delta_1 + \delta_2)d_1 \\ -d_2 t_1 + (u_2 - d_2)t_2 = (\delta_1 + \delta_2)d_2. \end{cases}$$

Solving the above system of linear equations for t_1 and t_2, we get

$$t_1 = \frac{u_2 d_1}{u_1 u_2 - u_1 d_2 - u_2 d_1}(\delta_1 + \delta_2), \tag{18}$$

$$t_2 = \frac{u_1 d_2}{u_1 u_2 - u_1 d_2 - u_2 d_1}(\delta_1 + \delta_2). \tag{19}$$

and T is given by:

$$T = \frac{u_1 u_2}{u_1 u_2 - u_1 d_2 - u_2 d_1}(\delta_1 + \delta_2) \tag{20}$$

The above expressions were also obtained in Bai and Elhafsi (1993) (when $u_i = U_i$, for i = 1,2), as limits of the transient cycles. Also, It has been shown in Bai and Elhafsi (1993) that if the demand set $\{d_1, d_2\}$ is feasible, then $(U_1 - d_1)(U_2 - d_2) > d_1 d_2$ which guarantees that the denominator in the expressions of t_1, t_2, and T, is strictly positive. We will show later that u_i has to be in $\{0, U_i\}$, i = 1,2, to minimize the cost of operating according to the Limit Cycle.

To be able to compute the average cost incurred over the Limit Cycle, we use the techniques of Sivazlian and Stanfel (1975). That is, we convert the inventory system with finite production rates and finite setup times to a system with instantaneous replenishment quantities. We first calculate the inventory and backlog costs incurred by each Part Type separately. The total cost incurred over the Limit Cycle is the sum of the costs incurred individually by each Part Type. Notice here that we start the Limit Cycle at point D (see Figure 1). It is clear that the unknowns are S_1, S_2 (the maximum surplus levels), s_1 and s_2 (the minimum surplus levels), and that $q_i = S_i -$

s_i, is the instantaneous replenishment quantity for Part Type i. Before calculating the cost of operating the system over the period of length T, we introduce the quantity d_i' which is computed as follows:

$$q_i = S_i - s_i = Td_i' = t_i(u_i - d_i), \tag{21}$$

substituting t_i and T by their expressions above and solving for d_i', we get

$$d_i' = d_i(1 - d_i/u_i). \tag{22}$$

Let τ_i denote the total amount of time inventory costs are incurred, and let τ_i' be the total amount of time backlog or shortage costs are incurred. The average inventory cost of Part Type i is then given by

$$\frac{c_i^+}{T}\int_0^{\tau_i}(S_i - d_i't)dt = \frac{c_i^+}{T}\left(S_i\tau_i - \frac{1}{2}d_i'\tau_i^2\right) \tag{23}$$

The average backlog cost of Part Type i is given by

$$-\frac{c_i^-}{T}\int_{\tau_i}^{T}(S_i - d_i't)dt = -\frac{c_i^-}{T}\left(S_i(T - \tau_i) - \frac{1}{2}d_i'(T^2 - \tau_i^2)\right) \tag{24}$$

The total cost of inventory and shortage (over a Limit Cycle) of Part Type i, $F_i(S_i,\tau_i)$, is then given by:

$$F_i(S_i,\tau_i) = \frac{c_i^+}{T}\left(S_i\tau_i - \frac{1}{2}d_i'\tau_i^2\right) - \frac{c_i^-}{T}\left(S_i(T - \tau_i) - \frac{1}{2}d_i'(T^2 - \tau_i^2)\right) \tag{25}$$

Noticing that $\tau_i = S_i/d_i'$, and substituting for τ_i, we get

$$F_i(S_i) = \frac{1}{2}\frac{S_i^2}{Td_i'}(c_i^+ + c_i^-) + \frac{1}{2}c_i^- d_i T - S_i c_i^-. \tag{26}$$

The total inventory and shortage cost over the Limit Cycle, $F(S_1,S_2)$, is given by

$$F(S_1, S_2) = \Sigma_{i=1}^2 F_i(S_i),$$

S_1 and S_2 are the values that minimize the total Limit Cycle average cost $F(S_1,S_2)$. This is an unconstrained optimization problem, and the function $F(S_1,S_2)$ is a separable function of S_1 and S_2. Thus, minimizing $F(S_1,S_2)$ with respect to S_1 and S_2, is equivalent to minimizing $F_1(S_1)$ with respect to S_1, and minimizing $F_2(S_2)$ with respect to S_2 separately. The minimum of $F_i(S_i)$ (i = 1,2) is obtained for

$$S_i = \frac{c_i^-}{c_i^- + c_i^+}Td_i',$$

using (22), notice that

$$q_i = (S_i - s_i) = Td_i' = \frac{u_1 u_2}{u_1 u_2 - u_1 d_2 - u_2 d_1} d_i (1 - d_i/u_i)(\delta_1 + \delta_2)$$
$$\Leftrightarrow q_i = \frac{(1 - d_i/u_i)}{1 - d_1/u_1 - d_2/u_2} d_i (\delta_1 + \delta_2). \tag{27}$$

Hence,

$$S_i = \frac{c_i^-}{c_i^- + c_i^+} q_i, \quad \text{for } i = 1, 2. \tag{28}$$

To verify that S_i, given above, is the minimum we compute the second derivative of the cost function $F_i(S_i)$, which is given below

$$\frac{\mathrm{d}^2 F_i(S_i)}{\mathrm{d}S_i^2} = \frac{c_i^+ + c_i^-}{q_i} \tag{29}$$

which is clearly positive, and therefore S_i is a global minimum.

The minimum surplus level s_i can be easily obtained from the expression $q_i = S_i - s_i$, and is given as follows:

$$s_i = \frac{-c_i^+}{c_i^- + c_i^+} q_i, \quad \text{for } i = 1, 2. \tag{30}$$

The total average cost over the Limit Cycle is then

$$F(S_1, S_2) = \frac{1}{2} \frac{c_1^- c_1^+}{c_1^- + c_1^+} q_1 + \frac{1}{2} \frac{c_2^- c_2^+}{c_2^- + c_2^+} q_2. \tag{31}$$

Note that the total average cost is implicitly a function of the production rates u_1 and u_2. Now, we have to choose u_1 and u_2, so as to minimize the total average cost. For this, we differentiate the cost function $F(S_1,S_2) = F(u_1,u_2)$ with respect to u_1 and u_2. Doing so, we obtain the following:

$$\begin{aligned}
\frac{\partial F}{\partial u_1} &= \frac{1}{2} \frac{c_1^- c_1^+}{(c_1^- + c_1^+)} \frac{\partial q_1}{\partial u_1} + \frac{1}{2} \frac{c_2^- c_2^+}{(c_2^- + c_2^+)} \frac{\partial q_2}{\partial u_1}, \\
\frac{\partial F}{\partial u_2} &= \frac{1}{2} \frac{c_1^- c_1^+}{(c_1^- + c_1^+)} \frac{\partial q_1}{\partial u_2} + \frac{1}{2} \frac{c_2^- c_2^+}{(c_2^- + c_2^+)} \frac{\partial q_2}{\partial u_2}, \\
\frac{\partial q_i}{\partial u_i} &= \frac{-(\delta_1 + \delta_2)}{(1 - d_1/u_1 - d_2/u_2)^2} \frac{d_j}{u_j} \frac{d_i^2}{u_i^2} < 0, \text{ for } i = 1, 2, j = 1, 2 \text{ and } i \neq j; \\
\frac{\partial q_j}{\partial u_i} &= \frac{-(\delta_1 + \delta_2)}{(1 - d_1/u_1 - d_2/u_2)^2} d_j (1 - d_j/u_j) \frac{d_i}{u_i^2} < 0, \text{ for } i = 1, 2, j = 1, 2 \\
&\quad \text{and } i \neq j.
\end{aligned}$$

The gradient of $F(u_1,u_2)$ is strictly negative. Hence, the minimum cost is obtained by setting u_1 and u_2 to their maximum values, U_1 and U_2 respectively.

The optimal location of the Limit Cycle is then given by:

$$A = \begin{pmatrix} S_1 \\ s_2 + \delta_2 d_2 \end{pmatrix}, B = \begin{pmatrix} s_1 - \delta_2 d_1 \\ s_2 \end{pmatrix}, C = \begin{pmatrix} s_1 + \delta_1 d_1 \\ S_2 \end{pmatrix} \text{ and}$$
$$D = \begin{pmatrix} s_1 \\ S_2 - \delta_1 d_2 \end{pmatrix}. \tag{32}$$

Where

$$S_i = \frac{c_i^-}{c_i^- + c_i^+} q_i, \quad \text{for } i = 1,2; \tag{33}$$

$$s_i = \frac{-c_i^+}{c_i^- + c_i^+} q_i, \quad \text{for } i = 1,2; \tag{34}$$

$$q_i = d_i(1 - d_i/U_i)(\delta_1 + \delta_2)/(1 - d_1/U_1 - d_2/U_2), \quad \text{for } i = 1,2. \tag{35}$$

and

$$t_1 = (\delta_1 + \delta_2)d_1/U_1/(1 - d_1/U_1 - d_2/U_2); \tag{36}$$

$$t_2 = (\delta_1 + \delta_2)d_2/U_2/(1 - d_1/U_1 - d_2/U_2); \tag{37}$$

$$T = (\delta_1 + \delta_2)/(1 - d_1/U_1 - d_2/U_2). \tag{38}$$

In this section, we determined the optimal location of the Limit Cycle by converting the problem to a LSP problem. The optimal location of the Limit Cycle corresponds to an optimal cyclic schedule of the two parts in the system's state space. In the next section, we give the transient solution of the problem.

4. Transient Solution

In this section, we will determine the optimal transient solution of our problem. This corresponds to finding optimal trajectories that lead to the Limit Cycle (or steady-state). First, we define what we call a transient and an optimal transient solution.

DEFINITION 4.1. A transient solution is defined as a trajectory in the x-space (or surplus space) emanating from an initial point and reaching the steady-state (i.e., the Limit Cycle) in a finite amount of time.

DEFINITION 4.2. An optimal transient solution is a transient solution that minimizes the cost of reaching the steady-state. That is, among all transient trajectories that lead to the Limit Cycle, it is the one that gives the minimum cost.

REMARK. In previous research work of this type (see Sharifnia *et al.* (1991)), the Limit Cycle has been usually assumed to be reachable asymptotically (see Figure 1). However, based on the evidence from the numerical solution to this problem (finite horizon discrete dynamic programming version of this problem), we believe that the Limit Cycle should be finite-time reachable for the system under consideration. In the definitions above, we assume that the Limit Cycle has to be reached in finite time.

Our problem is to find such optimal transient solution for every initial point in the surplus space. The transient solution to our problem is based on the following theorem and the subsequent facts.

THEOREM 4.1. *The optimal production rate vector* $u^* = ({u^*}_1, {u^*}_2)$ *belongs to the finite set of vectors* $\Omega^* = \{(0,0),(U_1,0),(d_1,0),(0,U_2),(0,d_2)\}$.

Proof. The proof is based on the Bellman equation. The latter has been derived in Gershwin (1994) for a flexible multiple part system. In our case, the Bellman equation, assuming that J^T (the transient cost component) is differentiable in x and t, is given by:

$$-\frac{\partial}{\partial t}J^T(x,t) = \min_{\sigma,u\in\Xi(\Phi,\Omega)}\left\{g(x) + \frac{\partial}{\partial x_i}J^T(x,t)(u_1 - d_1) + \frac{\partial}{\partial x_2}J^T(x,t)(u_2 - d_2)\right\}.$$

It is clear that when the machine is undergoing a setup change for a Part Type, there is no decision to make and $({u^*}_1, {u^*}_2)$ is forced to be equal to (0,0). Now assume that we know the optimal setup state of the machine. Let $\sigma = (1,0)$ be this setup state. That is, the machine can produce Part Type 1. In this case the Bellman equation can be rewritten as follows:

$$-\frac{\partial}{\partial t}J^T(x,t) = \min_{u\in\Omega(1,0)}\left\{g(x) + \frac{\partial}{\partial x_1}J^T(x,t)(u_1 - d_1) + \frac{\partial}{\partial x_2}J^T(x,t)(u_2 - d_2)\right\}$$

Now, notice that at each time instant t, if we knew $J^T(x,t)$, we would solve a linear programming problem, for which u_1 and u_2 are the decision variables, $\partial J^T/\partial x_1$ and $\partial J^T/\partial x_2$ are the cost coefficients and $\Omega(1,0)$ is the constraint set. $\Omega(1,0) =$

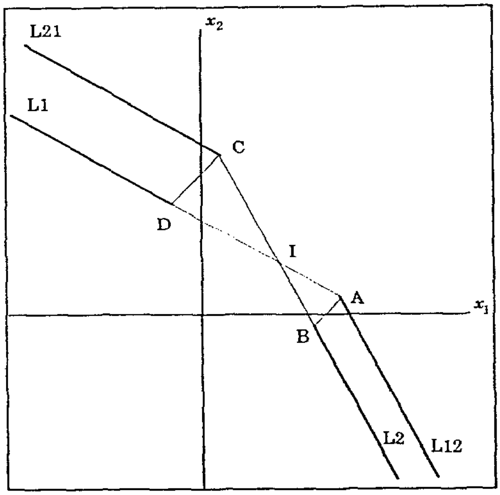

Fig. 2. Illustration of lines L1, L2, L12 and L21.

$\{(u_1,u_2) \mid 0 \leq u_1 \leq U_1, u_2 = 0\}$ is a convex set. We know that the solution of the above linear programming problem is always at an extreme point of the constraint set $\Omega(1,0)$, that is (u^*_1,u^*_2) is either equal to (0,0) (if $\partial J^T/\partial x_1 > 0$), or equal to $(U_1,0)$ (if $\partial J^T/\partial x_1 < 0$). Furthermore the solution is unique, if the cost coefficient $\partial J^T/\partial x_1$ is non-zero. In the case $\partial J^T/\partial x_1 = 0$, the solution is not unique anymore, since any solution (u^*_1,u^*_2) will not affect the objective function of the linear programming problem at time instant t. However, to keep the cost coefficient $\partial J^T/\partial x_1$ equal to zero at time instant $t + \delta t$, we should produce Part Type 1 at the demand rate d_1 so as to minimize the rate of increase of the cost function J^T. In this case (u^*_1,u^*_2) is equal to $(d_1,0)$. A similar argument is used when the optimal setup state is $\sigma = (0,1)$. ■

The linear programming problem above, has also another very important consequence, which we state in the following fact.

FACT 4.1. Since the cost coefficients in the Bellman equation are functions of x, given an optimal choice $\sigma^*(t \leq \tau \leq t_s)$, the linear programming problem above suggests a partition of the x-space into mutually exclusive regions (see Gershwin 1994). Each region corresponds to an optimal setup state $\sigma^* = (\sigma^*_1,\sigma^*_2)$, and an optimal production rate $u^* = (u^*_1,u^*_2)$ which belongs to the set Ω^*, defined above.

Another important fact drawn from the Bellman equation is given as follows:

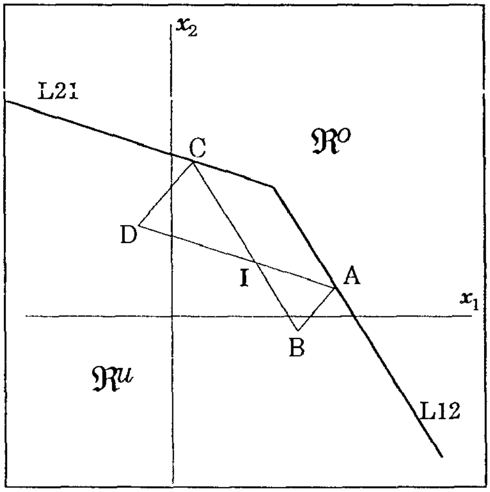

Fig. 3. Regions $\Re^u$ and $\Re_o$.

FACT 4.2. It has been shown in Connolly (1992) that J^T, the cost function, does not depend explicitly on the time, for the flexible two-part case. She also showed that J^T is piecewise quadratic and therefore the cost coefficients of the Bellman equation are linear in x. The latter implies that the boundaries of the regions in x-space must be linear (see Gershwin 1994). These observations are still valid in our case for an optimal choice of σ. In other words, suppose that we know how to choose σ optimally, then for each such a choice, the cost function is not an explicit function of time and is piecewise quadratic in x. Therefore, the boundaries of the regions in the surplus space must be linear. Of course in our case (as opposed to the non-setup case), this x-space partition is more complicated by the choice of σ.

For the subsequent analysis, the following important quantities need to be defined. First we need to find analytic expressions for the equations of line (A,D), line (B,C), line parallel to (A,D) containing Point C and the line parallel to (B,C) containing Point A (see Figure 2). These line equations are easy to find, since we know the coordinates of the points A, B, C, and D (from the steady-state optimal solution). These are given below.

$$\text{Line (A,D)} : d_2t_1(x_1 - S_1) + q_1(x_2 - s_2 - \delta_2 d_2) = 0. \tag{39}$$

$$\text{Line (B,C)} : q_2(x_1 - s_1 - \delta_1 d_1) + d_1t_2(x_2 - S_2) = 0. \tag{40}$$

$$\text{Line parallel to (A,D) containing C}: d_2t_1(x_1 - s_1 - \delta_1 d_1) + q_1(x_2 - S_2) = 0. \tag{41}$$

$$\text{Line parallel to (B,C) containing A}: q_2(x_1 - S_1) + d_1t_2(x_2 - s_2 - \delta_2 d_2) = 0. \tag{42}$$

We will refer to line (A,D) as Line L1, line (B,C) as Line L2, line parallel to (A,D) containing C as Line L21, and line parallel to (B,C) containing A as Line L12 (see Figure 2).

Having the equations of Lines L1 and L2, we can easily find the coordinates of the intersection point I, which is also of importance in the subsequent analysis. Intersection Point I is given by (after some manipulations):

$$I = \begin{pmatrix} i_1 \\ i_2 \end{pmatrix} = \begin{pmatrix} c_1^- q_1/(c_1^- + c_1^+) - \delta_2 q_1/(\delta_1 + \delta_2) \\ c_2^- q_2/(c_2^- + c_2^+) - \delta_1 q_2/(\delta_1 + \delta_2) \end{pmatrix} \tag{43}$$

Notice that intersection Point I can be located in any quadrant of the surplus space, depending on the values of the penalty function coefficients and the setup time duration.

Without loss of generality we index the parts such that Part Type 1 is the part type with the larger setup time. To solve the transient problem we divide the surplus space into two major mutually exclusive Regions $\Re^u$ and $\Re^o$ defined as follows: $\Re^u$ is the region in x-space located bellow the Lines L12 and L21. $\Re^u$ is the region in x-space located above the Lines L12 and L21 (see Figure 3). Algebraically,

$$\begin{aligned} \Re^u &= \{(x_1, x_2) \mid q_2(x_1 - S_1) + d_1 t_2(x_2 - s_2 - \delta_2 d_2) \\ &\quad < 0; d_2 t_1(x_1 - s_1 - \delta_1 d_1) + q_1(x_2 - S_2) < 0\}; \\ \Re^o &= \{(x_1, x_2) \mid q_2(x_1 - S_1) + d_1 t_2(x_2 - s_2 - \delta_2 d_2) \\ &\quad \geq 0; d_2 t_1(x_1 - s_1 - \delta_1 d_1) + q_1(x_2 - S_2) \geq 0\}. \end{aligned}$$

We divide the analysis into two parts. In the first part, we find the optimal transient solution for all initial surplus levels in region $\Re^o$. In the second part, we do the same for all initial surplus levels in Region $\Re^u$.

4.1. Optimal Transient Solution in Region $\Re^o$

As mentioned above, our task is to bring the initial surplus levels in x-space to a point on the Limit Cycle. Because, once on the Limit Cycle, the surplus will behave according to it thereafter (since the machine is reliable), and we know that operating according to the Limit Cycle is optimal (from the steady-state solution). The solution approach for initial surplus levels in Region $\Re^o$ is based on inspection and intuition. However, it is easy to see that the solution is indeed optimal. For instance, from Theorem 4.1 and Fact 4.1, we know that the optimal production rates are piecewise constant and are selected from a very small set. Hence, from the system dynamics (differential equations (1)), it is easy to verify that the optimal trajectories in x-space are piecewise linear and therefore it would not be difficult to see that they are indeed optimal.

We will provide a detailed solution for the case where Intersection Point I is located in the first quadrant (see Figure 2). The positive location of Intersection Point I may be the most common in practice. It has been reported in the literature that the backlog and inventory costs, $c^-{}_i$ and $c^+{}_i$ (i = 1,2), are usually chosen at

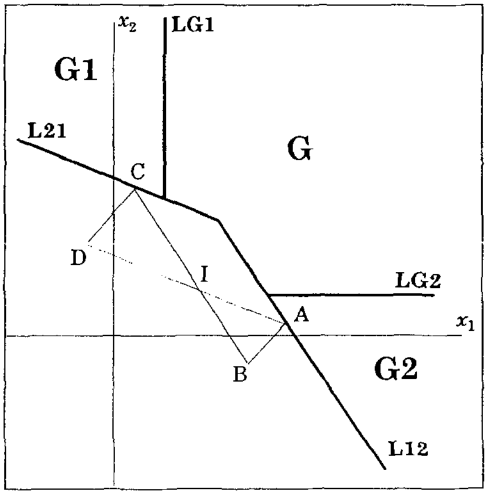

Fig. 4. Partition of Region $\Re^o$.

least in the ratio of 5 to 1 (because backlog usually results in sales and customers good will losses and other undesirable effects). The setup times δ_i (i = 1,2) are of the same order of magnitude. Hence from the expression of the coordinates of Intersection Point I in (42), we have

$$c_i^-/(c_i^- + c_i^+) > \delta_j/(\delta_1 + \delta_2) \quad \text{for } i = 1,2; j = 1,2 \text{ and } j \neq i.$$

Hence, i_1 and i_2 are positive numbers. In any case, the solution procedure is exactly the same for Intersection Point I located in any other quadrant.

Throughout this paper, we assume that initially the machine is not setup to either part type. For initial surplus levels in Region $\Re^o$, we can state our problem as follows: *Subject to constraints (1)–(7), find an optimal trajectory (i.e., with minimum cost) emanating from an initial point $x(t) \in \Re^o$ in x-space and reaching the Limit Cycle at a point $x(t_s)$, where it would be possible to move according to it thereafter.*

To determine the optimal solution for initial surplus levels in Region $\Re^o$, we first partition Region $\Re^o$ into three mutually exclusive regions G, G1, and G2 by introducing two linear boundaries LG1 and LG2. Line LG1 is the line for which the surplus level of Part Type 1 is exactly equal to $d_1\delta_1$ and Line LG2 is the line for which the surplus level of Part Type 2 is exactly equal to $d_2\delta_2$. Then, we further partition Region G into eight mutually exclusive regions G11, G12, G21, G22, H11, H12, H21, and H22. Figures 4 and 5 show the partition of Region $\Re^o$ and G respectively. These regions are defined as follows:

$$G11 = \{(x_1, x_2) \mid x_1 - \delta_1 d_1 \geq 0; -d_2(x_1 - s_1 - \delta_1 d_1) + d_1(x_2 - S_2) \geq 0\};$$

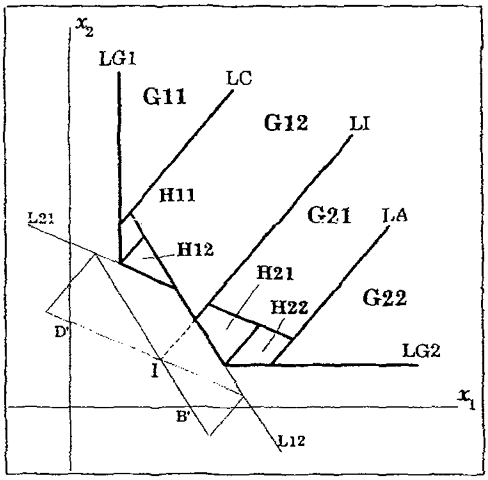

Fig. 5. Partition of Region G.

$$
\begin{aligned}
G12 = \{(x_1, x_2) \mid\ & d_2(x_1 - s_1 - \delta_1 d_1) - d_1(x_2 - S_2) > 0; -d_2(x_1 - i_1) \\
& + d_1(x_2 - i_2) \geq 0; q_2(x_1 - S_1) + d_1 t_2(x_2 - s_2 - \delta_2 d_2) \geq 0\}; \\
G21 = \{(x_1, x_2) \mid\ & -d_2(x_1 - S_1) + d_1(x_2 - s_2 - \delta_2 d_2) \geq 0; d_2(x_1 - i_1) \\
& - d_1(x_2 - i_2) > 0; d_2 t_1(x_1 - s_1 - \delta_1 d_1) + q_1(x_2 - S_2) \geq 0\}; \\
G22 = \{(x_1, x_2) \mid\ & x_2 - \delta_2 d_2 \geq 0; d_2(x_1 - S_1) - d_1(x_2 - s_2 - \delta_2 d_2) > 0\}; \\
H11 = \{(x_1, x_2) \mid\ & d_2(x_1 - s_1 - \delta_1 d_1) - d_1(x_2 - S_2) > 0; x_1 - \delta_1 d_1 \geq 0; \\
& -q_2(x_1 - S_1) - d_1 t_2(x_2 - s_2 - \delta_2 d_2) \geq 0; -d_2(x_1 - g_{11}) \\
& + d_1(x_2 - g_{12}) \geq 0\}; \\
H12 = \{(x_1, x_2) \mid\ & d_2(x_1 - g_{11}) - d_1(x_2 - g_{12}) > 0; d_2 t_1(x_1 - s_1 - \delta_1 d_1) \\
& + q_1(x_2 - S_2) \geq 0; -q_2(x_1 - S_1) - d_1 t_2(x_2 - s_2 - \delta_2 d_2) \geq 0\}; \\
H21 = \{(x_1, x_2) \mid\ & q_2(x_1 - S_1) + d_1 t_2(x_2 - s_2 - \delta_2 d_2) \geq 0; d_2(x_1 - i_1) \\
& - d_1(x_2 - i_2) > 0; -d_2 t_1(x_1 - s_1 - \delta_1 d_1) - q_1(x_2 - S_2) > 0; \\
& - d_2(x_1 - g_{21}) + d_1(x_2 - g_{22}) \geq 0\}; \\
H22 = \{(x_1, x_2) \mid\ & -d_2(x_1 - S_1) + d_1(x_2 - s_2 - \delta_2 d_2) > 0; x_2 - \delta_2 d_2 \geq 0; \\
& -d_2 t_1(x_1 - s_1 - \delta_1 d_1) - q_1(x_2 - S_2) > 0; -d_2(x_1 - g_{21}) \\
& + d_1(x_2 - g_{22}) > 0\}; \\
G1 = \{(x_1, x_2) \mid\ & -x_1 + \delta_1 d_1 > 0; d_2 t_1(x_1 - s_1 - \delta_1 d_1) + q_1(x_2 - S_2) \geq 0\}; \\
G2 = \{(x_1, x_2) \mid\ & -x_2 + \delta_2 d_2 > 0; q_2(x_1 - S_1) + d_1 t_2(x_2 - s_2 - \delta_2 d_2) \geq 0\}.
\end{aligned}
$$

Where, $g_1 = (g_{11}, g_{12})^T$ is the point in x-space given by:

$$
\begin{cases}
g_{11} = \delta_1 d_1; \\
g_{12} = S_2 + s_1 d_2 t_1 / q_1.
\end{cases}
$$

$g_2 = (g_{21}, g_{22})^T$ is the point in x-space given by:

$$\begin{cases} g_{21} = S_1 + s_2 d_1 t_2 / q_2; \\ g_{22} = \delta_2 d_2. \end{cases}$$

And $I = (i_1, i_2)^T$ is the point in x-space given by (42).

In the following, we provide the optimal transient solution for each of the aforementioned regions.

4.1.1. *Initial Surplus Levels in Region G*

In this case, both surplus levels are positive. Therefore, the production is ahead of the demand for both parts and it is optimal to let the surplus levels deplete at the maximum possible rate. That is, we set the production rates to zero. However we also want to reach the Limit Cycle as quickly as possible. Hence, we need to decide which setup we must change the machine to, and when is the best time to do it. We answer this question by considering each sub-region of Region G separately.

Initial surplus in Region G11: The optimal trajectory is obtained by first setting the production rates to zero. This corresponds to a trajectory moving downward in the southwest direction with speed $(-d_1, -d_2)$. When the surplus of Part Type 1 reaches the level $\delta_1 d_2$ (that is the trajectory hits Line LG1), we immediately start a setup change for Part Type 1. When the setup change is finished, the surplus level of Part Type 1 is exactly 0. That of Part Type 2 is still positive and we still have not reached the Limit Cycle. Therefore, we need to decrease the surplus level of Part Type 2 and keep the surplus level of Part Type 1 at the zero level until we touch the Limit Cycle. The optimal way to do this, is to produce Part Type 1 at the demand rate d_1 (see Theorem 4.1). This corresponds to a trajectory moving downward along the x_2 axis with a speed $(0, -d_2)$ until the Limit Cycle is touched at point D′.

Initial surplus in Region G12: The optimal trajectory is obtained by setting the production rates to zero, until the surplus trajectory touches the boundary of Region G12 on Line L12. At this point, we immediately start a setup change to Part Type 2. At the end of the setup change the trajectory touches the Limit Cycle at a point on the Segment [I,C].

Initial surplus in Region G21: The optimal trajectory is obtained by setting the production rates to zero, until the surplus trajectory touches the boundary of Region G21 on Line L21. At this point, we immediately start a setup change to Part Type 1. At the end of the setup change the trajectory touches the Limit Cycle at a point on the Segment [I,A].

Initial surplus in Region G22: The optimal trajectory is obtained by setting the production rates to zero, until the surplus of Part Type 2 reaches the level $\delta_2 d_2$ (that is the trajectory hits Line LG2). At this point, we immediately start a setup change for Part Type 2. At the end of the setup change, the surplus level of Part Type 2 is exactly zero, that of Part Type 1 is still positive and we still have not reached the Limit Cycle. Therefore, we need to decrease the surplus level of Part Type 1 and keep the surplus level of Part Type 2 at the zero level until we touch the Limit Cycle. To do this, we produce Part Type 2 at the demand rate d_2. This corresponds to a trajectory moving along the x_1 axis with a speed $(-d_1,0)$, until the Limit Cycle is touched at point B′.

Initial surplus levels in Region H11: Trajectories emanating in Region H11 are similar to those emanating in Region G11.

Initial surplus levels in Region H12: Trajectories emanating in Region H12 are similar to those emanating in Region G21.

Initial surplus levels in Region H21: Trajectories emanating in Region H21 are similar to those emanating in Region G12.

Initial surplus levels in Region H22: Trajectories emanating in Region H22 are similar to those emanating in Region G22.

4.1.2. *Initial Surplus Levels in Region G1*

Notice that for initial points in this region, the surplus level of Part Type 2 is always positive and the surplus level of Part Type 1 is less than $\delta_1 d_1$. Therefore, we need to produce Part Type 1 so that we can reach the Limit Cycle. Now, suppose that the surplus level of Part Type 1 is positive. In this case, even if we started a setup change to Part Type 1, we would end up with a backlog of Part Type 1, since during the setup change to Part Type 1, $\delta_1 d_1$ amount of Part Type 1 is depleted. The optimal trajectories emanating from Region G1 and leading to the Limit Cycle are obtained as follows: First we start a setup change to Part Type 1. When the setup change is completed, x_1 is negative and x_2 is positive. Hence we produce Part Type 1 at the maximum production rate to eliminate the backlog of Part Type 1. This corresponds to a trajectory moving southeast with a speed of $(U_1 - d_1, -d_2)$. When the surplus level of Part Type 1 becomes zero, that of Part Type 2 is still positive (that is the point where the trajectory hits the x_2 axis), and we still have not reached the Limit Cycle. Therefore, we need to decrease the surplus level of Part Type 2 and keep the surplus of Part Type 1 at the zero level until we touch the Limit Cycle. The optimal way to do this is to produce Part Type 1 at the demand rate d_1 (see Theorem 4.1). This corresponds to a trajectory moving downward along the x_2 axis with a speed $(0,-d_2)$, until the Limit Cycle is touched at Point D′.

4.1.3. *Initial Surplus Levels in Region G2*

For all initial points in this region, the surplus level of Part Type 1 is always positive and the surplus level of Part Type 2 is less than $\delta_2 d_2$. This is the symmetric case of points in region G1. That is, the optimal way to get to the Limit Cycle is to set up the machine for Part Type 2 first, produce Part Type 2 at the maximum production rate until its surplus level becomes zero. At this point, change the level of production of Part Type 2 to d_2 parts per time unit and continue producing this part until the trajectory touches the Limit Cycle at the point B′.

To summarize the control actions in Region $\Re^{o}$, let $x = (a,b)$ be the vector of initial surplus levels in Region $\Re^{o}$. Then,

If $x \in$ G1:

1: Set up the machine for Part Type 1;
2: After the setup change, produce Part Type 1 at the rate U_1;
3: When the surplus level of Part Type 1 becomes 0, change the production rate to d_1;
4: When the surplus level of Part Type 2 becomes $x_{2D'} = d_2 t_1 S_1/q_1 + s_2 + \delta_2 d_2$ (point D′ in Figure 5), switch to the control actions of the Limit Cycle.

If $x \in$ G11 ∪ H11:

1: Do not produce either part type;
2: When the surplus level of Part Type 1 becomes $\delta_1 d_1$, start a setup change for Part Type 1;
3: After the setup change, produce Part Type 1 at the demand rate d_1;
4: When the surplus level of Part Type 2 becomes $x_{2D'} = d_2 t_1 S_1/q_1 + s_2 + \delta_2 d_2$, switch to the control actions of the Limit Cycle.

If $x \in$ G12 ∪ H21:

1: Do not produce either part type;
2: When the surplus level of Part Type 2 reaches level l_2, immediately start a setup change for Part Type 2. l_2 is the surplus level of Part Type 2, when the trajectory hits Line L12 and is given by

$$l_2 = \frac{(bd_1 + (S_1 - a)d_2)q_2 + d_1 d_2 t_2 (s_2 + \delta_2 d_2)}{d_1(q_2 + t_2 d_2)}$$

3: at the end of the setup change switch to the Limit Cycle control actions.

If $x \in$ G21 ∪ H12:

1: Do not produce either part type;
2: When the surplus level of Part Type 1 reaches the level l_1, immediately start a setup change for Part Type 1. l_1 is the surplus level of Part Type 1, when the trajectory hits Line L21 and is given by

$$l_1 = \frac{(ad_2 + (S_2 - b)d_1)q_1 + d_1 d_2 t_1 (s_1 + \delta_1 d_1)}{d_2(q_1 + t_1 d_1)}$$

3: at the end of the setup change switch to the Limit Cycle control actions.

If $x \in$ G22 ∪ H22:

1: Do not produce either part type;
2: When the surplus level of Part Type 2 becomes $\delta_2 d_2$, start a setup change for Part Type 2;
3: After the setup change, produce Part Type 2 at the demand rate d_2;

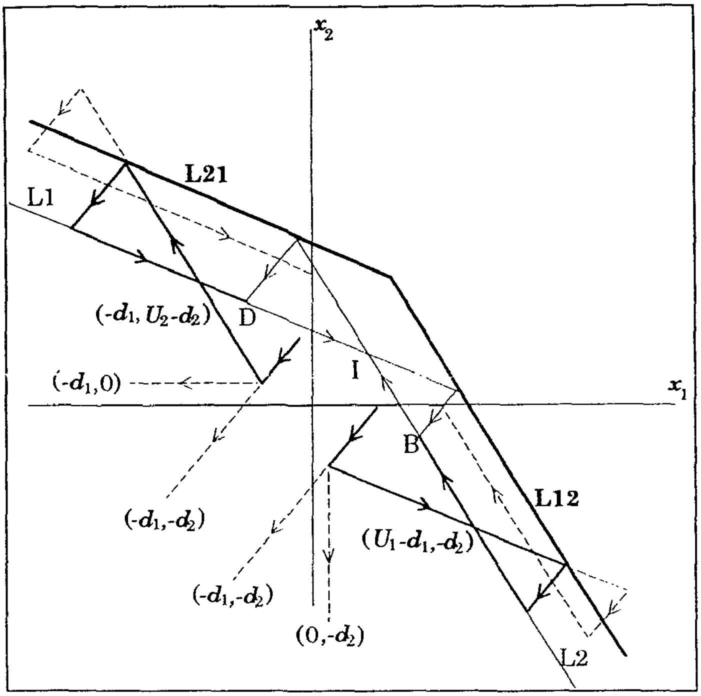

Fig. 6. Illustration of Fact 4.3 and Fact 4.4.

4: When the surplus level of Part Type 1 becomes $x_{1B'} = d_1S_2/q_2 + s_1 + \delta_1 d_1$ (point B′ in Figure 5), switch to the control actions of the Limit Cycle.

If $x \in$ G2:

1: Set up the machine for Part Type 2;
2: After the setup change, produce Part Type 2 at the rate U_2;
3: When the surplus level of Part Type 2 becomes 0, change the production rate to d_2;
4: When the surplus level of Part Type 1 becomes $x_{1B'} = d_1t_2S_2/q_2 + s_1 + \delta_1 d_1$, switch to the control actions of the Limit Cycle.

Notice that, all optimal trajectories emanating in Region $\Re^{o}$ reach the Limit Cycle by moving downward and that only one setup change to either part type is performed before the Limit Cycle is reached. In other words, for initial points in Region $\Re^{o}$, the Limit Cycle is always reached from above and with only one setup change.

4.2. Optimal transient solution in region $\Re^{u}$

In this subsection, we will develop an algorithm to obtain the optimal trajectories emanating in Region $\Re^{u}$ and reaching the Limit Cycle in finite time. But before we proceed with the algorithm, we establish the following facts and definitions.

In the previous subsection, we showed that for initial surplus levels in Region $\Re^{o}$, the Limit Cycle is always reached in finite time, from above, with only one setup change. For initial surplus levels in Region $\Re^{u}$, if we immediately started a setup change to a part, at the end of the setup change, we would miss either Segment [A,D] if we setup for Part Type 1, or Segment [B,C] if we setup for Part Type 2 (which are the target segments for the surplus to stay on the Limit Cycle). To bring the surplus levels to either Segment [A,D] or [B,C] in finite time, we need to generate a surplus excess for the part type the machine is set up for, so that when we switch to the production of the other part type, we end up on or above the appropriate segment of the Limit Cycle. Now, after the setup change, we need to produce the part type we set up the machine for. Based on Theorem 4.1, we can set the production rate to zero, the demand rate, or maximum machine capacity. It is clear that, if we set the production rate to zero, both surplus levels deplete and the generated trajectory moves in the direction $(-d_1, -d_2)$ and hence further drifts away from the Limit Cycle (see Figure 6). If we produced at the demand rate the part type the machine is setup for, we would keep the surplus level of this part type constant, while the surplus of the other part type depletes. In this case, the generated trajectory will move parallel to one of the axis in x-space in the direction that further drifts away from the Limit Cycle (see Figure 7). If we produced at maximum machine capacity the part type the machine is set up for, the surplus level of this part type increases, that of the other part type decreases. This way it is possible to generate an excess surplus for the part type being produced. The following fact formally states this result.

FACT 4.3. For initial surplus levels in Region $\Re^{u}$, the only way to progress toward the Limit Cycle is by producing at maximum machine capacity whenever it is possible.

Based on Fact 4.3, the trajectories emanating in Region $\Re^{u}$, either move along direction $(-d_1, -d_2)$ during a setup change, along direction $(U_1 - d_1, -d_2)$ during the production of Part Type 1 or along direction $(-d_1, U_2 - d_2)$ during the production of Part Type 2. Therefore, the Limit Cycle cannot be reached from below (i.e., from points in Region $\Re^{u}$) in finite time. To be able to touch the Limit Cycle in a finite time, starting in Region $\Re^{u}$, we should bring the surplus levels in Region $\Re^{u}$, then apply the optimal control actions of that Region (since we know that those controls lead to the Limit Cycle in finite time). Doing so will generate a surplus excess of either part type. To reach the Limit Cycle with the least amount of excess, we must bring the surplus levels to the boundary of Region $\Re^{o}$ (Line Lij = 1,2; j = 1,2, $i \neq j$) with minimum cost. Once on the boundary Lij, we switch to Part Type j and produce this part type (this corresponds to a trajectory moving along Line Lj) until the Limit Cycle is reached at Point D if $j = 1$ or B if $j = 2$ (see Figure 6). The following fact formally states this result.

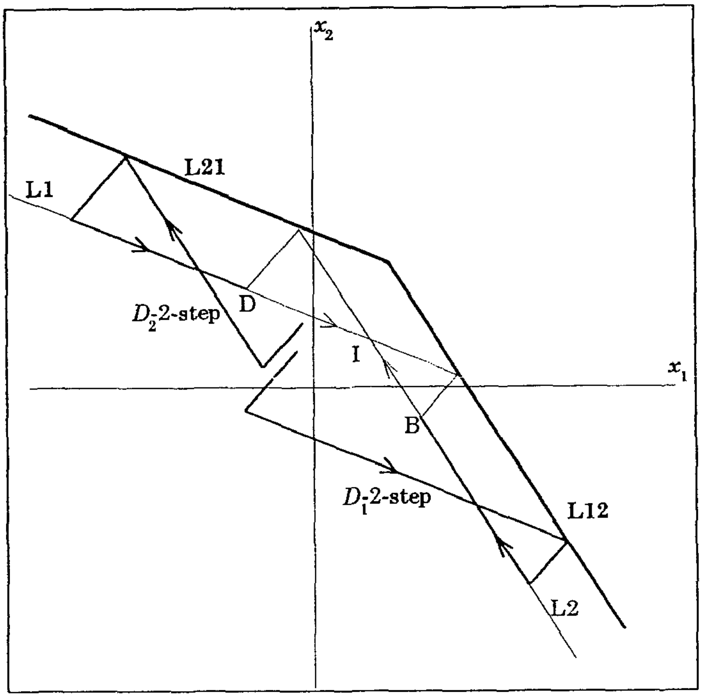

Fig. 7. D_1 – 2-*step* and D_2 – 2-*step* trajectories.

FACT 4.4. To reach the Limit Cycle in finite time, starting with initial surplus levels in Region $\Re^u$, the trajectory leading to the Limit Cycle must touch the boundary L*ij* of Region $\Re^o$, just before switching to Part Type *j* and reaching the Limit Cycle at one of the points B or D on the Limit Cycle.

DEFINITION 4.3.We say that a trajectory is following Direction D_i, if it moves parallel to Line L*i* (see Figure 2) in the direction of increasing x_i. That is, the machine is producing Part Type *i* ($i = 1,2$).

REMARK. Since for surplus levels in Region $\Re^u$ the machine always produces at its maximum rate, the trajectories will move either along Direction D_1 or along Direction D_2.

DEFINITION 4.4. We call a direction D_i – *n-step* trajectory ($i = 1,2$; $n > 1$), a trajectory that performs alternately m_i setup change-production runs of Part Type *i* and m_j setup change-production runs of Part Type *j* ($j \neq i$)); with the initial setup change to Part Type *i* and the last segment touching the Limit Cycle at point B or D. If *n* is even then $m_i = m_j = n/2$. If *n* is odd then $m_i = (n+1)/2$ and $m_j = (n-1)/2$. Figure 7 shows a D_1 – 2-*step*, and a D_2 – 2-*step* trajectory.

We can now state the problem as follows: *Given surplus levels in Region* $\Re^u$, *find a* $D_{i*} - n^*$*-step trajectory, that minimizes the cost of reaching the Limit Cycle and satisfying constraints (1)–(7). Where* n^* *is the optimal number of steps (setup change-production runs) before reaching the Limit Cycle and* i^* *is the first direction to follow (i.e., the initial setup change).*

As a consequence of Facts 4.3 and 4.4, the optimal trajectory leading to the Limit Cycle in finite time will be at least D_i -2*-step*. Since, we need a setup change at the initial point and one more, once the boundary of Region $\Re^o$ is reached. This leads to the following proposition.

PROPOSITION 4.1. *Let* $C^n{}_i(x)$ $(i = 1,2)$ *be the cost of moving along a* $D_i - n$*-step trajectory with initial surplus point x in surplus space. An upper bound on the optimal cost of reaching the Limit Cycle is given by* $min_{i=1,2}\{C_i^2(x)\}$.

Proof. The proof is straight forward and is based on Facts 4.3 and 4.4. Assume that the optimal trajectory with initial point x is $D_{i*} - n^*$*-step* ($n^* \geq 2$). We know that an optimal trajectory originating at the point x in surplus space is at least D_i − 2*-step*. It follows that,

$$C_{i*}^{n*}(x) \leq C_1^2(x) \text{ and } c_{i*}^{n*}(x) \leq C_2^2(x). \text{ Therefore, } C_{i*}^{n*}(x) \leq \min_{i=1,2}\{C_i^2(x)\}.$$

■

Based on Facts 4.3 and 4.4, The optimal trajectory emanating at a point in Region $\Re^u$, and leading to the Limit Cycle in finite time can be obtained as follows: Given an initial surplus point in Region $\Re^u$, we choose the first setup and calculate the cost of the trajectory leading to the Limit Cycle with two setup changes only. At this point, we have a D_i − 2*-step* trajectory, where i is the initial setup for Part Type i. The next step is to try to lower the cost of the current trajectory by introducing a setup change to the other part type so that the Limit Cycle is reached at the opposite side. If the cost can be reduced, then the obtained new trajectory is a D_i − 3*-step* trajectory. We keep trying to reduce the cost of the current trajectory by introducing, each time, a setup change before the Limit Cycle is reached, until we cannot lower the cost anymore. At this point we have an optimal $D_i - n$*-step* trajectory (provided we start with a setup change to Part Type i) emanating in Region $\Re^u$ and reaching the Limit Cycle in finite time. In the same manner, we obtain the optimal $D_j - n$*-step* ($j \neq i$) trajectory starting with a setup to Part Type j first. The optimal trajectory would be the one with the lowest cost. The above procedure leads to the following algorithm which is called Direction Sweeping Algorithm (DSA). The reason for this will be given shortly.

4.2.0.1. *Direction Sweeping Algorithm*

Notation:

$x(0)$: *Initial surplus in x-space (Input)*;

$\{x(0), x(1), \ldots, x(k)\}$: Optimal trajectory (Output);

J_x: *Optimal trajectory cost (Output);*

$C(x(k))$: *Cost of the trajectory up to the point* $x(k)$;

$Cgm(Y, Z)$: *Cost along Segment [Y,Z].*

Dr: Direction of search. Dr=D_1 or Dr=D_2 as defined in Definition 4.3.

$\bar{D}r$*=1 if Dr=2 and* $\bar{D}r$*=2 if Dr=1.*

$C^n_{Dr}(X)$: *Cost of Dr-n-step trajectory starting at the point X in x-space.*

$I(X, Dr)$: *Intersection point of the line containing X and with direction Dr, with Line* $L_{Dr\bar{D}r}$, *where* $L_{Dr\bar{D}r} = L_{12}$ *or* $L_{Dr\bar{D}r} = L_{21}$, *depending on the direction of search Dr.*

$P_{Dr} : P_1 = B$ *and* $P_2 = D$. *B and D are given by equation (32) in Section 3.*

Main Algorithm:

// Initialization;

$k := 1;$

$m := 1;$

STOP:=FALSE;

// Find an upper bound on the cost of the optimal trajectory emanating from x(0):

$(J_x, Dr) : \min_{Dr=1,2}\{C^2_{Dr}(x(0))\};$

// Update trajectory:

$k := k + 1;$

$x(k) := x(k-1) + \delta^*_{Dr}(-d_1, -d_2);$

$\bar{D}r := notDr;$

// Update trajectory cost up to the point x(k):

$C(x(k)) := Cgm(x(k-1), x(k));$

WHILE STOP $\neq$ *TRUE*

DO

Swap(Dr, $\bar{D}r$*);*

// Add a new step to the trajectory:

$(c^2_{Dr}(Y(m)), Y(m)) := \min_{Y\in[x(k),I(x(k),Dr)]}\{C^2_{Dr}(y) + Cgm(x(k), Y, \bar{D}r)\};$

// This is a Line Search Procedure.

IF $C(x(k)) + Cgm(x(k), Y(m)) + C^2_{Dr}(Y(m)) < J_X$ *THEN*

// Update trajectory:

$k := k + 1;$

$x(k) := Y(m);$

$k := k + 1;$

$x(k) := x(k-1) + \delta^*_{Dr}(-d_1, -d_2);$

// Update upper bound on the optimal cost; $J_X := C(x(k)) + Cgm(x(k), Y(m))$ $C^2_{Dr}(Y(m));$

// Update trajectory cost up to the point x(k);

$C(x(k)) := C(x(k)) + Cgm(x(k-2), x(k-1)) + Cgm(x(k-1), x(k));$

ELSE

// Cannot add another step to the current trajectory.
// Then the trajectory x(k) is optimal.
STOP:= TRUE;
END
END DO
$Swap(Dr, \bar{Dr});$
$k := k + 1;$
$x(k) := I(x(k), Dr);$
$k := k + 1;$
$x(k) := x(k-1) + \delta^*_{\bar{D}r}(-d_1, -d_2);$
$k := k + 1;$
$x(k) := P_{Dr}$
// Optimal trajectory cost:
$C(x(k)) := J_X;$

Notice, each time we try to add a new step to the current trajectory, we sweep all possible trajectories in the direction of search and pick the one that minimizes the cost of reaching the Limit Cycle (this is done by means of the line search procedure). This is why the algorithm is called Direction Sweeping Algorithm.

The numerical solution of various examples suggests that the above algorithm be further simplified based on the following Conjecture.

CONJECTURE 4.1. *The initial setup of the optimal trajectory is given by* $i^* = argmin_{i=1,2}\{C^2{}_i\,(x)\}$.

Conjecture 4.1 implies that, we only need to compare two $D^i - 2$-*step* trajectories instead of comparing two $D_i - n$-*step* trajectories. Which is simpler and faster to compute.

In the following, we prove the validity of the line search and the uniqueness of its solution. For this, we use the algorithm notation.

PROPOSITION 4.2. *The cost function* $C^2{}_{Dr}(x)$ *is strictly convex in* $x \in [x(k), I(x(k), \bar{Dr})]$.

Proof. Let

$$\begin{aligned}C^2_{Dr}(x) = {} & Cgm(x, x', 0) + Cgm(x', I(x', Dr), Dr)\\ & + Cgm(I(x', Dr), x'', 0) + Cgm(x'', P_{\bar{D}r}, \bar{Dr}),\end{aligned}$$

where $Cgm(X,Y,l)$ is the cost of Segment $[X,Y]$ when direction l is followed. Here, direction 0 is the South-West direction. In other words the machine is undergoing a setup change. In general, we can write the cost along Segment $[X,Y]$ in x-space as follows:

$$Cgm(X, Y, l) = \sum_{i=1}^{2} \frac{1}{2} \frac{h_i}{(\sigma_i U_i - d_i)} \{y_i^2 - x_i^2\};$$

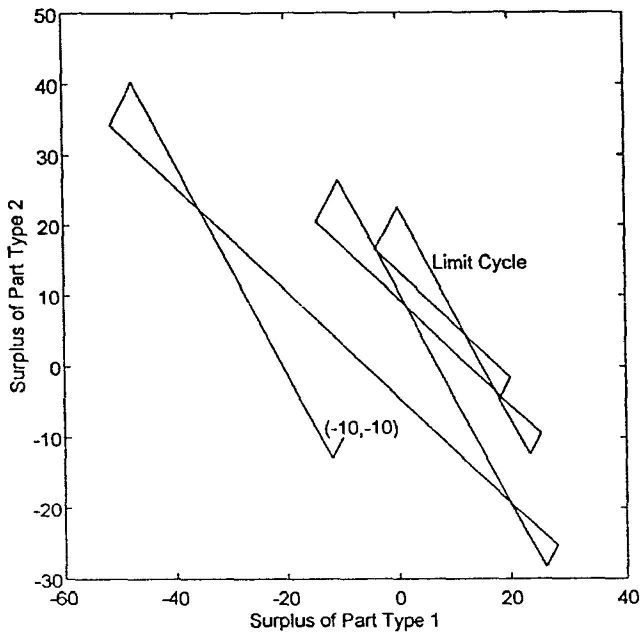

Fig. 8. Optimal trajectory obtained by the DSA algorithm.

Where, $X = (x_1,x_2)$, $Y = (y_1,y_2)$ and $h_i = c^{+}{}_i$ if x_i and y_i are positive and $h_i = -c^{-}{}_i$ if x_i and y_i are negative for $i = 1,2$.

Now, if Y is fixed, then it is easy to show $Cgm(X,Y,l)$ is a strictly convex function of X.

In particular, $Cgm(x'', P_{\bar{D}r}, \bar{D}r)$ is a strictly convex function of x'' for fixed $P_{\bar{D}r}$ (recall that $P_{\bar{D}r}$ is either point B or point D given by (31)). Since x'' can be expressed as a linear function of $P_{\bar{D}r}$, $Cgm(I(x', Dr), x'', 0) + Cgm(x'', P_{\bar{D}r}, \bar{D}r)$ is a strictly convex function of $I(x',Dr)$ for fixed $P_{\bar{D}r}$. Continuing the reasoning in the same manner, and since the domain $[x(k), I(x(k), \bar{D}r)]$ is convex, it follows that $C^2{}_{Dr}(x)$ is a strictly convex function of $x \in [x(k), I(x(k), \bar{D}r)]$. ∎

Using Proposition 4.2, it follows that the line search in the algorithm always provides a unique optimal solution.

THEOREM 4.2. *The DSA algorithm gives a unique optimal trajectory with the optimal cost* $C^{m*}{}_{i*}(x(0))$*, where* $x(0)$ *is the initial point of the trajectory in x-space,* m^* *is the optimal number of steps in the trajectory and* i^* *is the initial direction to follow.*

Proof. To see that the trajectory obtained by the DSA is indeed optimal, consider an initial point $x(0)$. Then, the DSA algorithm can be described by the following

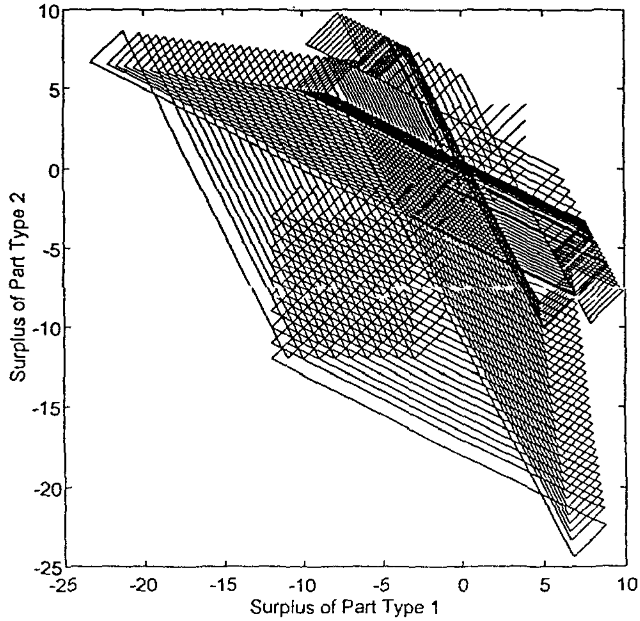

Fig. 9. Linearity of the Boundaries in Region $\Re^u$.

recursive expression:

$$C^2_{i*}(x(0)) = \min_{i=1,2}\{C^2_i(x(0))\};$$

$$C_i *^{m+1}(x(0)) =$$

$$\min\left\{ C^m_{i*}(x(0)); \min_{Y\in[x(2m-3),I(x(2m-3),im)]}\{C(Y) + C^2_{\bar{i}_m}(Y)\}\right\};$$

where m=2,3,...,m^*; $i_1 = \bar{i}^*, i_2 = i^*, i_3 = \bar{i}^*, \ldots$; i^* is the starting direction of search, $C^m{}_{i*}(x(0))$ is the cost of the D_i*-m-step trajectory and $C(Y)$ is the cost of the current trajectory up to point Y. The remaining quantities are the same as defined by the algorithm. The recurrence relation above and Proposition 4.2 guarantee that each step added to the trajectory is optimal. Therefore, by the principle of optimality, the trajectory obtained is optimal. Furthermore by Proposition 4.2 the optimal trajectory is unique. ■

Figure 8 shows an optimal trajectory starting at the point $(-10,-10)$ for the following problem data: $d_1 = 2$, $d_2 = 3$, $U_1 = 6$, $U_2 = 6$, $\delta_1 = 2$, $\delta_2 = 1$, $c^+{}_2 = 10$, $c^+{}_2 = 10$, $c^-{}_1 = 50$, and $c^-{}_2 = 50$. Notice that, the optimal trajectory is D_1-5-*step*. Figure 9 shows different optimal trajectories starting at different initial points. Notice that the setup switching policy is a special corridor policy. The latter has linear walls, which means that the boundaries in Region $\Re^u$ are linear as mentioned in Fact 4.2.

5. Conclusion

In this paper, we studied a deterministic one-machine two-product manufacturing system with setup changes. We formulated the problem as an optimal control. We divided the planning horizon into a transient period and a steady-state period. For the steady-state period, the optimal setup switching schedule and production flow rates were derived. For the transient period, the surplus/backlog space was divided into two major Regions $\Re^o$ and $\Re^u$. For initial surplus levels in Region $\Re^o$, the optimal solution was obtained by inspection. For initial surplus levels in Region $\Re^u$, an algorithm to obtain the optimal state trajectory was developed. The algorithm involves only a line search procedure, which makes it very fast. The complete optimal solution is a feedback control policy.

In this paper, we dealt with a deterministic problem. A stochastic version of this problem, with random production capacities, can be considered as an extension of this work. Moreover, fixed setup costs can be introduced in addition to setup times. In this case, we suspect that the Limit Cycle will have a completely different structure.

Acknowledgements

The authors would like to thank the editor and two anonymous reviewers for their helpful comments. This work is partially supported by a grant from the IEEE/Engineering Foundation under contract RI-A-94-07.

References

S.X. Bai and M. Elhafsi (1993), Real-Time Scheduling of a Manufacturing System with Non-Resumable Setup Changes, Research Report 93–28. Department of Industrial and Systems Engineering. University of Florida.

M. Caramanis, A. Sharifnia, J. Hu, and S.B. Gershwin (1991), Development of a Science Base for Planning and Scheduling Manufacturing Systems, *Proceedings of the 1991 NSF Design and Manufacturing Systems Conference*, The University of Texas Austin, Austin, Texas, January 9–11, 1991.

J.J. Carreno (1990), Economic Lot Scheduling for Multiple Products on Parallel Identical Processors, *Management Science* **36**, 348–358.

S. Connolly (1992), A Real-Time Policy for Performing Setup Changes in a Manufacturing System, Master's Thesis, MIT Operations Research Center, May; MIT Laboratory for Manufacturing and Productivity. Report LMP-92-0005.

G. Dobson (1987), The Economic Lot Scheduling Problem: Achieving Feasibility Using Time-Varying Lot Sizes, *Operations Research* **35**, 764–771.

M. Elhafsi and S.X. Bai (1995), Optimal Production Control of a Dynamic Two-Product Manufacturing System with Setup Costs and Setup Times, Research Report 95–1. Department of Industrial and Systems Engineering. University of Florida.

S.E. Elmaghraby (1978), The Economic Lot Scheduling Problem (ELSP): Review and Extensions, *Management Science* **24**, 587–598.

G. Gallego (1989), Real-Time Scheduling of Several Products on a Single Facility with Setups and Backorders, Ph.D. Thesis, School of OR&IE, Cornell University, Ithaca, NY.

S.B. Gershwin (1994), *Manufacturing Systems Engineering*. Prentice Hall, Inc.

S.K. Goyal (1984), Determination of Economic Production Quantities for a Two-Product Single Machine System, *International Journal of Production Research* **22**, 121–126.

J. Hu and M. Caramanis (1992), Near Optimal Set-up Scheduling for Flexible Manufacturing Systems, Technical Report, Department of Manufacturing Engineering, Boston University, Boston, MA, 1992; *Proceedings of the Third Renselaer Polytechnic Institute International Conference on Computer Integrated Manufacturing*, Troy, New York, May 20–22, 1992.

J. Hu and M. Caramanis (1995), Dynamic Set-up Scheduling of Flexible Manufacturing Systems: Design and Stability of Near Optimal General Round Robin Policies, to appear in *Discrete Event Systems*, IMA Volumes in Mathematics and Its Applications Series.

J. Kimemia and S.B. Gershwin (1983), An Algorithm for the Computer Control of a Flexible Manufacturing System, *IIE Transactions* **15**(4), 353–362.

P.R. Kumar and T.I. Seidman (1990), Dynamic Instabilities and Stabilization Methods in Distributed Real-Time Scheduling of Manufacturing Systems, *IEEE Transactions on Automatic Control* **35**, 289–338.

J. Perkins and P.R. Kumar (1989), Stable, Distributed, Real-Time Scheduling of Flexible Manufacturing/Assembly/Disassembly Systems, *IEEE Transactions on Automatic Control* **34**, 139–148.

A. Sharifnia, M. Caramanis, and S.B. Gershwin (1991), Dynamic Setup Scheduling and Flow Control in Manufacturing Systems, *Discrete Event Dynamic Systems: Theory and Applications* **1**, 149–175.

B.D. Sivazlian and L.E. Stanfel (1975), *Analysis of Systems in Operations Research*. Prentice Hall, Inc.

N. Srivatsan and S.B. Gershwin (1990), Selection of Setup Times in a Hierarchically Controlled Manufacturing System, *Proceedings of the 29th IEEE Conference on Decision and Control*, Hawaii (December).