Journal of Global Optimization **13:** 43–59, 1998.

Tracking Demands in Optimal Control of Managerial Systems with Continuously-divisible, Doubly Constrained Resources

KONSTANTIN KOGAN and EUGENE KHMELNITSKY
Department of Industrial Engineering, Tel-Aviv University, Tel-Aviv 69978, Israel
E-mail: kogan@eng.tau.ac.il

(Received 30 December 1996; accepted 13 November 1997)

Abstract. The paper addresses problems of allocating continuously divisible resources among multiple production activities. The resources are allowed to be doubly constrained, so that both usage at every point of time and cumulative consumption over a planning horizon are limited as it is often the case in project and production scheduling. The objective is to track changing in time demands for the activities as closely as possible. We propose a general continuous-time model that states the problem in a form of the optimal control problem with non-linear speed-resource usage functions. With the aid of the maximum principle, properties of the solutions are derived to characterize optimal resource usage policies. On the basis of this analytical investigation, numerical scheduling methods are suggested and computationally studied.

Key words: Resource constrained scheduling, Renewable and nonrenewable resources, Optimal control

1. Introduction

Significant work has been done to advance scheduling theory and its application to diverse industrial, management, service and other real-life systems and more is likely to come due to both great influence of scheduling solutions on the functioning of those systems and 'hardness' of the scheduling problems themselves. The state-of-the-art of scheduling theory and its applications can be found in recent literature (see, for example, Blazewicz et al. 1996, Gershwin 1994 and Pinedo 1995) which covers various scheduling approaches and environments. In particular, Blazewicz et al. (1996) stress importance and usefulness of considering continuously devisable resources in problems of scheduling constrained resources. They discuss some special cases of these problems which are solved optimally and suggest heuristic approaches for solving general problem formulations approximately. This is typically accomplished by reducing the original dynamic problem to a number of static convex programming problems.

In contrast to the described approach, this work addresses the continuous-time, optimal control model which allows the problem to be solved in its dynamic formulation. The formulation incorporates most important and common features of the modern production and project scheduling in a general scheduling model. The generalization concerns integration of the following major scheduling issues:

- two types of resource constraints;
- nonlinearity of resource utilization;
- nonlinear objective of tracking changing demands.

To detail the generalization, we start our discussion from the production scheduling related extensions.

Following the current literature on the continuous-time production scheduling or, more exactly, on the production flow control in a flexible manufacturing system, the inventory $X_i(t)$ flow of product-type $i(i = 1, 2, \ldots, I)$ is traditionally presented as the difference between the system production rate $u_i(t)$ for product i at moment t and demand rate $d_i(t)$ for this product at the moment:

$$\dot{X}_i(t) = u_i(t) - d_i(t), X_i(0) = X_i^0, \tag{1}$$

where X_i^0 is the given initial inventory level of product type i.

A generalization to be discussed in the present paper is modeling of the system production rate $u_i(t)$ as a given function of the resource required to produce product i (speed-resource function), i.e., we introduce the production flow as:

$$\dot{X}_i(t) = f_i(r_i(t)) - d_i(t), X_i(0) = X_i^0, \tag{2}$$

where the resource usage $r_i(t)$ is a bounded control variable,

$$0 \leqslant r_i(t) \leqslant b_i . \tag{3}$$

Clearly, that the equation (1) is a special case of the equation (2). The increasing continuous function $f_i(r_i(t))$, $f_i(0) = 0$ describes a real-life situation when the production rate of a system is not necessarily linearly dependent on the resource usage as it was in equation (1).

Another generalization related to production flow control concerns utilization of the resources. No matter the nature of the speed-resource functions, whether linear as stated in equation (1) or non-linear as allowed in equation (2), in industry, the resources are often doubly constrained as it is the case in project management. Typical examples of such resources are money and energy which are usually treated as non-renewable resources, while in practice they are doubly constrained due to restrictions on their current usage rate. At the same time, manpower cannot be always treated as a renewable resource, because the number of man-hours consumed is also often restricted. Therefore, we present the following two types of resource constraints: renewable and non-renewable. These affect differently the extremal behavior of the system in both linear and nonlinear cases of the resource usage and consumption as it will be discussed in the sequel.

The renewability of the resource at *every point* in time is constrained as follows:

$$\sum_i r_i(t) \leqslant N(t), \tag{4}$$

where $N(t)$ is the maximal level of the *resources usage* at time t. We will assume that $N(t)$ is greater than each specific bound b_i, i.e., $b_i < N(t)(\forall i$, and is less than the sum of these bounds, $\sum_i b_i > N(t)$. Since constraint (4) refers to each point in time separately and does not involve the preceding resource utilization, the resources it constrains, are commonly called renewable. However, the resources are allowed to be non-renewable in the sense that the maximum cumulative amount of the resource consumed up to time t is predetermined and fixed.

The non-renewability of the resources is imposed by the maximal cumulative *resource consumption* $M(t)$ from the beginning of the planning horizon through moment t:

$$\sum_i \int_0^t r_i(s)ds \leqslant M(t)\,. \tag{5}$$

These are the generalizations concerning the dynamic scheduling models suggested so far for production flow control in flexible manufacturing environment (Kimemia and Gershwin 1983, Khmelnitsky and Kogan 1994, Khmelnitsky, Kogan and Maimon 1995, Kogan and Khmelnitsky 1996).

On the other hand, the project scheduling continuous-time models are traditionally concerned with diverse resources including doubly constrained as well as speed-resource usage functions (for example, Weglarz 1981, Janiak and Stankiewicz 1983, Leachman et al. 1990). However, these continuous-time dynamic models do not account explicitly for due dates, especially when the due dates take the most general, dynamic form introduced in equation (2).

Thus, the generalization related to the project scheduling concerns the system dynamics and the objective function. Namely, in place of the equation of the cumulative state of project activity i (Weglarz 1981):

$$\dot{X}_i(t) = f_i(r_i(t))\,,$$

the current state equation (2) is utilized. Furthermore, instead of the typical reduction of the objective to minimizing the project durations, we utilize a general objective for both project and production scheduling. This objective is to track the dynamic demands as closely as possible:

$$\int_0^T \sum_i C(X_i(t))dt \to \min\,, \tag{6}$$

where T is the planning horizon and $C(X_i(t)) = 1/2\,c_i[X_i(t)]^2$ is a convex cost function traditionally chosen quadratic to clarify the presentation. When there are surpluses (overproduction) $X_i(t) > 0$ in the system, this function reflects inventory holding cost, while it is backlogging cost in the case of shortages (underproduction).

Thus, model (2)–(6) incorporates diverse features significant for both production and project scheduling and, hence, applicable to a wide range of scheduling problems. Moreover, when a demand for products is given as an amount of every product required per time unit along the planning horizon, this model presents a general case of the dynamic lot-sizing in manufacturing systems and preemptive project activities scheduling in project management.

However, if the demand is not dynamic, that is, they are given only at one point (i.e., at a due date) of the planning horizon and they require only a single job (or normalized activity with work content equal to one) to be carried out at that point, then the model presents classic preemptive job (activity) scheduling with the objective to minimize total dynamic lateness and tardiness.

2. Canonical statements of the problem

Optimal control theory has been successfully applied to various production control and planning problems when dealing with only renewable resources which usage is characterized by linear speed-resource functions (Sethi et al. 1992, Bergstrom and Smith 1970, Khmelnitsky and Kogan 1996).

In order to apply the optimal control to the problem (2)–(6) and develop a general solution approach, the problem is stated in a canonical form. The behavior of the state variables $X_i(t)$ has to be presented by differential equations with right-hand side functions being differentiable with respect to the control $r_i(t)$ and state $X_i(t)$ variables. The constraints on the state and control variables are to be described by algebraic equalities (inequalities) also differentiable in $r_i(t)$ and $X_i(t)$. Evidently, the constraint (5) is not canonical due to its integral form, and function $f_i(r_i(t))$ is not necessarily differentiable. Therefore, we first assume this function to be differentiable or approximated by a differentiable function with respect to $r_i(t)$.

To convert the non-canonical constraint (5), we introduce a new state variable $Y(t)$ which is a cumulative amount of the resource consumed by time t. Then, the resource consumption is described as follows:

$$\dot{Y}(t) = \sum_i r_i(t), \quad Y(0) = Y^0, \tag{7}$$

where, similar to the initial inventory level X_i^0 in the system (see (2)), Y^0 defines initial state of the resource consumed by the beginning of the planning horizon.

Consequently, constraint (5) now readily transforms into a canonical one:

$$Y(t) \leqslant M(t)\,. \tag{8}$$

Since the problem (2)–(4), (6)–(8) is canonical, the maximum principle can be applied to study the properties of the optimal solutions and develop a time-decomposition solution algorithm (Ilyutovich and Khmelnitsky 1991, Hartl, Sethi and Vickson 1995). However, the state constraints, for example, constraint (8), are well known to be the cause of a substantial computational burden for any time-decomposition procedure. Therefore, we relax the constraint (8) by replacing it with the same constraint, but valid only by the end of the planning horizon. This is a quite common manufacturing practice: to restrict the total nonrenewable resource consumption, while renewable resources are limited at each point of time,

$$Y(T) \leqslant M(T)\,. \tag{9}$$

3. Properties of optimal solutions for the linear speed-resource function

Given problem (2)–(4), (6), (7) and (9), the maximal principle applied to the problem, states (Bryson and Ho 1975, Dubovitsky and Milyutin 1981) that if a trajectory $(X_i(t), Y(t), r_i(t))$ is optimal, then there exist piecewise continuous functions (dual variables)

- the dual differential equations and transversality conditions hold

$$\dot{\psi}_i^X(t) = c_i X_i(t), \quad \psi_i^X(T) = 0 \tag{10}$$

$$\dot{\psi}^Y(t) = 0, \quad \psi^Y(T) \leqslant 0\,; \tag{11}$$

- the Hamiltonian H is maximized for each t by the controls $r_i(t)$:

$$H = -\frac{1}{2}\sum_i c_i[X_i(t)]^2 + \sum_i \psi_i^X(t)[f_i(r_i(t)) - d_i(t)] + \psi^Y(t)\sum_i r_i(t) \to \max\,. \tag{12}$$

  subject to constraints (3) and (4).
- the local maximum principle holds

$$\frac{\partial H}{\partial r_i(t)} = \psi_i^X(t)\,\frac{\partial f_i(r_i(t))}{\partial r_i(t)} + \psi^Y(t) = 0\,. \tag{13}$$

Since the dual variable $\psi^Y(t)$ is a non-positive constant (see (11)), the argument t can be further omitted.

From the dual problem (10)–(12), it follows that there are two major options to search for an optimal solution. One is to optimize the Hamiltonian with respect to the control variables and control constraints analytically as the maximum principle (12) states. Based on the obtained optimal properties, the solution to the original problem can then be reduced to a procedure of global optimization search with respect to initial values of the dual variables. Specifically, for each set of the initial values of $\psi_i^X(0)$ and ψ^Y , the trajectory is built uniquely by integrating the primal (2), (7) and dual (10) equations from the left to the right. Consequently, the objective (6) becomes a function of these initial values. This methodological option is possible if a specific form of the speed-resource function is selected to be studied.

The other option is not directly related to a specific form of the speed-resource functions and, therefore, can be exploited in a general form. It is accomplished by optimizing the Hamiltonian numerically based only on the local maximum principle (13). Such an approach commonly leads to the projected-gradient based solution procedures.

Both described options are attended in the paper as time-decomposition methods because they are based on the Hamiltonian maximization at every point of time. The speed-resource functions, which enter the Hamiltonian - the dual problem objective, are to be further constrained to have only concave or linear forms. This ensures the unimodality of the dual problem so that its global optimum can be found by standard analytical and numerical tools.

We start first from the analytical option. Since the study of the optimal regimes requires a specific form of the speed-resource functions to be selected, the following linear $f_i(r_i(t)) = k_i r_i(t)$ and concave $f_i(r_i(t)) = k_i\sqrt{r_i(t)}$ functions are further chosen to illustrate the first solution methodology.

LEMMA 1 (necessary optimality conditions). *Given problem (2)–(4), (6), (7) and (9), $f_i(r_i(t)) = k_i r_i(t)$, $G_i(t) = k_i\psi_i^X(t)+\psi^Y$; optimal control $r_i(t)$ in the problem is necessarily defined as follows:*

- $r_i(t) = 0$, *if* $G_i(t) < 0$, $\forall i$ *(no-resource usage regime);*
- $r_i(t) = b_i$, *if* $G_i(t) > 0$ *and* $\sum_{i'|G_{i'}\geqslant G_i} b_{i'} \leqslant N(t)$; $r_i(t) = N(t) - \sum_{i'|G_{i'}>G_i} b_{i'}$ *if* $G_i(t) > 0$ *and* $\sum_{i'|G_{i'}>G_i} b_{i'} < N(t) \leqslant \sum_{i'|G_{i'}\geqslant G_i} b_{i'}$ *(greedy resource usage regime);*
- $r_i(t) \in [0, b_i]$, *if* $G_i(t) = 0$ *and* $\sum_{i'|G_{i'}>G_i} b_{i'} < N(t)$ *(first singular regime);*
- $r_i(t) \in [0, b_i]$, $r_{i'}(t) \in [0, b_{i'}]$, *if* $G_i(t) = G)_{i'}(t)$, $\sum_{i''|G_{i''}>G_i} b_{i''} < N(t)$ *and* $\sum_{i''|G_{i''}\geqslant G_i} b_{i''} > N(t)$ *(second singular regime);*

Proof. According to the maximum principle, the control-dependent term of the Hamiltonian

$$H = \sum_i \psi_i^X(t) k_i r_i(t) + \psi^Y \sum_i r_i(t) \tag{14}$$

must be maximized at each time with respect to the control constraints. Taking into account that this term is linear in $r_i(t)$ as well as the linearity of constraints (3) and (4), one can readily observe that the maximum is achieved in one of the cases stated in the lemma and it is illustrated in Figure 1 for the two-product system case.

Specifically, when coordinates of the Hamiltonian gradient

$$\frac{\partial H}{\partial r_i(t)} = G_i(t) = k_i\psi_i^X(t) + \psi^Y$$

are positive and not equal to one another, $G_1(t) > G_2(t)$ (greedy resource usage regime), the first resource is to be utilized as much as just possible, i.e., $r_1(t) = b_1$, while the second one contents with the remaining capacity, $r_2(t) = b_2$, if $b_1 + b_2 \leqslant N(t)$ and $r_2(t) = N(t) - b_1$ otherwise.

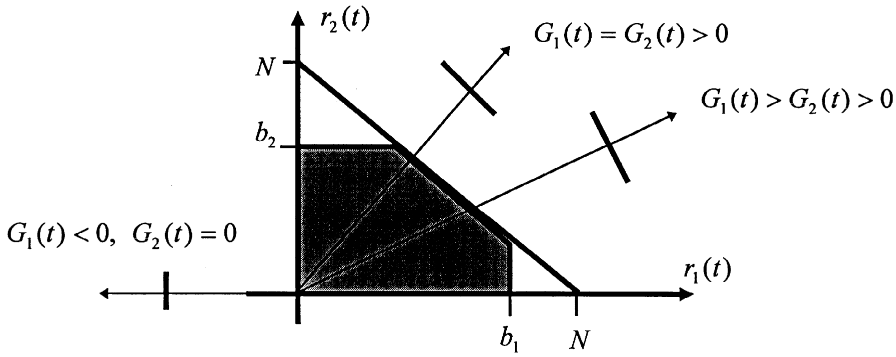

Figure 1. Optimal regimes: linear speed-resource function case.

In the same Figure, the two cases where the optimal controls are uncertain (singular regimes), are presented by the gradient vector perpendicular to either one of the axis or to the edge formed by the control constraint $r_1(t) + r_2(t) \leqslant N(t)$. □

Note, that unlike the regular no-resource usage and greedy resource usage regimes where the optimal controls are uniquely defined, the singular regimes possess undetermined controls. The following Lemmas resolve the ambiguity for the singular cases determined in Lemma 1.

LEMMA 2. *Given the singular regime condition for a product i:*

$$G_i(t) = 0 \tag{15}$$

satisfied on an interval of time; the optimal control value on the interval for this singular regime is defined as follows:

$$r_i(t) = \frac{d_i(t)}{k_i}\,. \tag{16}$$

Proof. By differentiating twice the condition (15) and taking into account the primal and dual differential equation (2) and (10) immediately result in equation (16). □

Note, to support the singular regime on an interval of time, the resource usage must follow the demand, as stated in the Lemma, and therefore it can only occur when the correspondent buffer is empty $X_i(t) = 0$ and demand is less than the operating capacity $d_i(t) \leqslant b_i k_i$.

LEMMA 3. *Given the singular regime conditions for a set of products I', $|I'| = n$:*

$$G_i(t) = G_{i'}(t), i, i' \in I', \sum_{i''|G_{i''}>G_i} b_{i''} < N(t) \text{ and } \sum_{i''|G_{i''}\geqslant G_i} b_{i''} > N(t) \tag{17}$$

satisfied on an interval of time; the optimal control values on the interval of this singular regime are defined from the system of the linear algebraic equations:

$$\begin{aligned} &k_i c_i (k_i r_i(t) - d_i(t)) = k_{i'} c_{i'} (k_{i'} r_{i'}(t) - d_{i'}(t)), \quad i, i' \in I'; \\ &\sum_{i \in I'} r_i(t) + \sum_{i'' | G_{i''} > G_i} b_{i''} = N(t) \end{aligned} \tag{18}$$

Proof. By differentiating twice the condition (17) and taking into account the primal and dual differential equation (2) and (10), we find $n-1$ algebraic equations:

$$k_i c_i (k_i r_i(t) - d_i(t)) = k_{i'} c_{i'} (k_{i'} r_{i'}(t) - d_{i'}(t)) .$$

The lemma's last equation is the resource usage equation, which complements the system of $n = |I'|$ relations for n unknown parameters of control. □

COROLLARY. *If, on an interval of time, the singular regime (17) is realized, either of the following conditions holds:*

$$\begin{aligned} &0 \leqslant k_i c_i d_i(t) - k_{i'} c_{i'} d_{i'}(t) \leqslant c_i k_i^2 b_i, \quad i, i' \in I' \\ &c_{i''} k_{i'}^2 b_{i'} \leqslant k_i c_i d_i(t) - k_{i'} c_{i'} d_{i'}(t) \leqslant 0, \quad i, i' \in I' \end{aligned} \tag{19}$$

Proof. The condition immediately follows from the condition (18) and control constraint (3). □

The studied optimization problem is unimodal, since its objective function is convex and the constraints are linear with respect to all state and control variables. The unimodal problems are known to have only one local optimum that brings the objective also to the globally minimum value. Therefore, the properties of the solution proved in Lemmas 1-3 are not only necessary but also sufficient for optimality.

The next section determines the necessary and sufficient optimality conditions for the case of a concave form of the speed-resource functions $f_i(r_i(t))$.

4. Properties of optimal solutions for the concave speed-resource function

The concave speed-resource functions approximated as $f_i(r_i) = k_i \sqrt{r_i}$ are considered in this section. The Hamiltonian in this case is:

$$H = -\frac{1}{2} \sum_i c_i X_i(t)^2 + \sum_i \psi_i^X(t)(k_i \sqrt{r_i(t)} - d_i(t)) + \psi^Y \sum_I i\, r_i(t) ,$$

and after the change of variables $z_i(t) = \sqrt{r_i(t)}$, the problem of the Hamiltonian maximization takes the following form:

$$\psi^Y \sum_i z_i(t)^2 + \sum_i \psi_i^X(t) k_i z_i(t) \rightarrow \max,$$
$$\text{s.t.} \quad 0 \leqslant z_i(t) \leqslant \sqrt{b_i}, \text{ and } \sum_i z_i(t)^2 \leqslant N(t)\,.$$

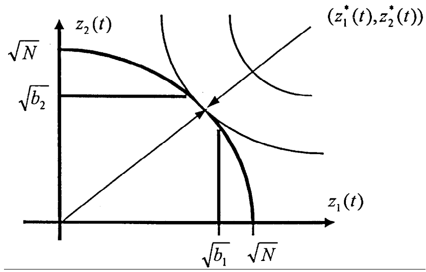

Figure 2. Optimal regimes: concave speed-resource function case

By classifying the optimal control regimes, we find the same no-resource and greedy resource usage regimes that are analogous to those proven in Lemma 1 for the linear statement. No singular regime can now occur, since the Hamiltonian is no longer linear. Figure 2 shows levels of equal values of the Hamiltonian against the admissible area of controls $z_i(t)$ for the two-product system case. The parabolic form of the Hamiltonian function with the vertex coordinate:

$$z_i^*(t) = -\psi_i^X(t) k_i / 2\psi^Y$$

causes the following additional regimes:

- $z_i(t) = z_i^*(t)$, if $0 \leqslant z_i^*(t) \leqslant \sqrt{b_i}$ and $|z_i^*(t)| \leqslant \sqrt{N(t)}$ (partial usage regime);
- $z_i(t) = z_i^*(t)\sqrt{N(t)}/|z_i^*)(t)|$, if $0 \leqslant z_i^*(t)\,\sqrt{N(t)}/|z_i^*(t)| \leqslant \sqrt{b_i}$ and $|z_i^*(t)| \geqslant \sqrt{N(t)}$ (full usage regime).

For the linear case of the speed-resource function, the necessary conditions of optimality have been also sufficient. For the concave case, the sufficiency issue should be given a special consideration.

LEMMA 4 (sufficient optimality conditions). *Given problem (2)–(4), (6), (7) and (9), in order for the solution $r_i(t)$ of the problem which satisfies the necessary optimality conditions to be globally optimal, it is sufficient if:*

- *function $f_i(r_i(t))$ is concave in $r_i(t)$, and*

· *for every t, such that $X_i(t) > 0$ the following inequality is satisfied*

$$\left[\int_0^t \frac{\partial f_i(r_i(\tau))}{\partial r_i}\, d\tau\right]^2 \geqslant -X_i(t)\int_0^t \frac{\partial^2 f_i(r_i(\tau))}{\partial r_i^2}\, d\tau\,. \qquad (20)$$

.

Proof. Note, that all constraints of the problem are linear except for the primal differential equation (2), while all the cost functions are convex. Thus, to verify the conditions of the lemma, we have to consider the influence of the nonlinear equation (2) on the unimodality of the problem. It is accomplished by replacing the equation (2) with its integral expression:

$$X_i(t) = X_i(0) + \int_0^t (f_i(r_i(\tau)) - d_i(\tau))d\tau$$

into the objective:

$$\frac{1}{2}\int_0^T \sum_i c_i \left[X_i^0 + \int_0^t (f_i(r_i(\tau)) - d_i(\tau))d\tau\right]^2 dt.$$

From the obtained objective, one can readily see that there is no even need to construct the Hessian matrix in order to verify the convexity of the objective, since it is simply the sum of independent functions of one variable. Thus, the problem is unimodal if at every point of time:

$$\frac{1}{2}\frac{\partial\left[X_i^0 + \int_0^t (f_i(r_i(\tau)) - d_i(t))d\tau\right]^2}{\partial r_i^2} \geqslant 0, \quad i = 1, 2, \ldots, I\,,$$

Taking the derivatives, we finally obtain:

$$X(t)\int_0^t \frac{\partial^2 f_i(r_i(\tau))}{\partial r_i^2} d\tau + \left[\int_0^t \frac{\partial f_i(r_i(\tau))}{\partial r_i}\, d\tau\right]^2 \geqslant 0. \qquad (21)$$

There are two conditions when the last inequality is trivially satisfied. The first condition concerns the linearity of the speed-resource function, that has been considered in the previous section. The second one is derived from the concavity of function $f_i(r_i(t))$, i.e., $\partial^2 f_i(r_i(t))/\partial r_i^2 \leqslant 0$. If demands in the system are greater than or equal to the available production rates in cumulative sense, so that $X_i(t) \leqslant 0$, the concavity of $f_i(r_i(t))$ is the sufficient condition. However, if this is not the case, then inequality (21) must be satisfied as stated in the Lemma. □

REMARK. *If demands in the system are pressing in terms of the available resources (i.e., $X_i(t) = X_i^0 + \int_0^t (f_i(r_i(\tau)) - d_i(\tau))d\tau \leqslant 0$), which is the most important and complex scheduling case in real-life manufacturing, the sufficient condition is met by any concave speed-resource function.*

5. Solution methods

Basics of the two time-decomposition methods have been discussed in the previous section. One method optimizes the Hamiltonian numerically by projection of the gradient on the set of control constraints. The other utilizes properties of the optimal solutions for solving the initial-point boundary-value problem (2), (7) and (10) (Burden and Faires 1989).

Both methods are quite well computationally studied in the literature for diverse scheduling problems with different number of differential equations, constraints and variables (Kogan and Khmelnitsky 1995, Khmelnitsky, Kogan and Maimon 1995), and therefore are only briefly discussed here. The advantage of the projected-gradient method is in its applicability to large-scale systems, while the global search based on initial-point boundary-value problems is well solvable (usually by shooting methods) only for relatively small-scale cases. However, large problems are tackled effectively with the former if a moderate accuracy is required, while the performance of the latter improves significantly if the missing boundary values (in our case initial values for the dual variables) can be estimated in advance. Therefore a combined method can be suggested where the projected-gradient approach is used to compute not only a fast approximate solution, but also initial boundary value approximation for the dual variables. Consequently, if the accuracy of the obtained gradient-based solution is not sufficient, these dual variable values can be beneficially employed by a global search procedure to elaborate the result on the basis of analytical Hamiltonian optimization.

PROJECTED-GRADIENT-BASED ALGORITHM

The main idea behind this time-decomposition method and underlying algorithm suggested here is to locate iteratively all types of the optimal regimes (see Lemmas 1-3) by the gradients (14) projected on the control constraints (3) and (4).

The method involves integration of the primal and dual systems under initial and terminal boundary conditions respectively in order to find gradients and, thus, control variations improving the objective on every iteration. Since the terminal boundary condition $\psi^Y(T)$ in the presented model is not completely defined for the dual equation (11), the following lemma is formulated to determine it.

LEMMA 5. *Given primal (2)-(4), (6), (7), (9) and dual (3), (4), (10), (12) problems and a trajectory $(r_i(t), X_i(t), Y(t))$ satisfying all the constraints of the problems, if a gradient-based method is applied to improve this trajectory on which $Y(T) - M(T) = 0$, then the new terminal boundary condition for the dual differential equations is determined by the following equation:*

$$\psi^Y = \frac{1}{T|I|} \sum_i \int_0^T \frac{\partial f_i(r_i(t))}{\partial r_i} \int_i^T c_i X_i(\tau) d\tau \, dt \tag{22}$$

· *otherwise if* $Y(T) - M(T) < 0$ *then* $\psi^Y = 0$. (23)

Proof. First, condition (23) is readily obtained from the constraint (9) related complementary slackness condition of the maximum principle:

$$(Y(T) - M(T))\psi^Y = 0 .$$

Let us now consider the case when constraint (9) is active, i.e. transforms into equality. Then any small control variation on the subsequent iteration may violate it, unless the variation of the state variable at the terminal time point is zero:

$$\delta Y(T) = 0 . \tag{24}$$

Integrating differential equation (7) and replacing it with $Y(T)$ in equation (24) we express the variation in this equation through the corresponding control variations:

$$\delta Y(T) = \int_0^T \sum_i \delta r_i(t)\, dt = 0$$

Consequently choosing a small ε for providing the control variation in the direction of the gradient (14) we obtain:

$$\delta Y(T) = \int_0^T \sum_i \varepsilon \left[\psi_i^X(t) \frac{\partial f_i(r_i(t))}{\partial r_i} + \psi^Y \right] dt = 0 ,$$

Finally, the stated in this lemma terminal dual variable equation (22) is readily obtained by replacing the dual variables with their integral expressions:

$$\psi_i^X(t) = -\int_t^T c_i X_i(\tau) d\tau$$

resulting from equation (10). □

The projected gradient-based algorithm is described as follows:

Step 1. Choose a feasible solution of the problem. For example, controls $r_i(t) = 0$, $\forall i$ meet all the constraints of the problem (2)–(4), (6), (7) and (9).

Step 2. Integrate from the left to the right the primal differential equations (2) and (7) to determine inventory $X_i(t)$ and resource consumption $Y(t)$ levels for the given controls.

Step 3. For the obtained inventory levels and controls calculate the objective according to equation (6).

Step 4. Calculate the missing terminal condition for the dual equation (11) as determined in Lemma 5. Integrate from the right to the left the dual differential equations (10).

Step 5. Calculate direction of the descent $p_i(t)$ by projecting the Hamiltonian gradients (13) on the constraints (3) and (4) to determine small control variations as a step(along the calculated direction.

Step 6. Make a step in the direction of the descent: $\delta r_i(t) = \varepsilon p_i(t)$.

Step 7. Integrate from the left to the right the primal differential equations (2) and (7) to determine inventory $X_i(t)$and resource consumption $Y(t)$ levels for the newly obtained controls.

Step 8. For the obtained inventory levels and controls calculate the objective (6). If all of the constraint are not violated and the objective is improved then go to the next step. Otherwise decrease (and go back to Step 6.

Step 9. Check the standard stop criterion value $\int_0^T \sum_i |p_i(t)|dt$; if it is less than a given tolerance then stop, the solution has been found, otherwise proceed to step 4.

INITIAL-VALUE-BASED GLOBAL SEARCH ALGORITHM

The algorithm is intended for defining time points of entering and exiting the optimal regimes analytically formulated in Lemmas 1-3 while simultaneously integrating the primal (2), (7) and dual (10), (11) differential equations in concurrent directions.

The algorithm is described as follows:

Step 1. Select arbitrary initial values for the dual variables, for example, $\psi_i^X(0) = \psi^Y = 0, \forall i$.

Step 2. Set current point of time to be critical and equal to zero. Define all intervals of time which satisfy Lemmas 2 and 3 and their Corollaries. Assign left endpoints as the potentially critical moments for entering the singular regimes. If there are no such moments go to Step 7.

Step 3. Integrate concurrently from the left to the right the primal differential equations (2), (7) and dual differential equations (10), (11) from the current critical point up to the nearest moment of entering a singular regime as defined by Lemmas 2-3 and their Corollaries. Set optimal controls during the integration up to that moment as prescribed in Lemma 1.

Step 4. Find a moment of exiting the current singular regime which satisfies the conditions of the subsequent potentially critical moment, i.e., the moment of entering the subsequent singular regime. If there are no more subsequent singular regimes on the planning horizon, then this moment is set at t, and the execution proceeds to Step 7.

Step 5. If there is no corresponding (legitimate) moment of exiting of this singular regime, i.e. the moment of entering it is greater than the found moment of exiting, then go to the next step, otherwise the singular regime is fixed and the moment of exiting of it is declared to be the current critical moment. Go to Step 3.

Step 6. Merge the current singular regime with the subsequent one. If entire planning horizon is already covered by the singular regimes, go to Step 7, otherwise go to Step 4.

Step 7. If a stop criterion of the chosen global optimization method is satisfied with a given accuracy, the optimal solution has been found. Otherwise, correct the values of $\psi_i^X(0)$, ψ^Y with respect to the method so that to minimize the objective (6). Go to Step 2.

6. Computational example

A solution of a four-product-type scheduling problem computed with the combined (projected-gradient and global random search) time-decomposition method is presented here. The solution illustrates the optimal behavior of the resource usage rates and the corresponding inventory (activity) levels when the square root case of the speed-resource usage functions is employed. The demands given and the production parameters selected for the problem are summarized in Table 1, where c_i^+ is the holding cost of product type I, i.e. when $X_i(t) \geqslant 0$, c_i^- is the backlogging cost ($X_i(t) < 0$).

The problem is solved on the planning horizon T of 25 time units, where the maximal resource usage $N(t)$ is equal to 8.37 product units per time unit and the total resource consumption $M(T)$ is allowed at 198.6 product units. The initial inventory levels X_i^0 as well as the initial resource consumption Y_i^0 are set at zero.

Figure 3 shows the obtained allocation of the doubly constrained resource to the four products along the planning horizon. The found peaks on the optimal trajectory illustrate how the constrained resource is scheduled in response to the changing in time demand conditions. Due to the pressing demand, the resource is found to be fully consumed, while full usage of it is observed only on a part of the planning horizon.

Table 1. Demand and system parameters

i	Demand					System parameters			
	1-5	6-10	11-15	16-20	21-25	k	b	c^+	c^-
1	1	1	2	4	2	1.3	4.2	0.1	10
2	2	0	1	2	1	1.5	1.1	0.2	13
3	2	1	3	0	3	1.2	2.3	0.08	11
4	2	1	3	4	2	1.5	4.9	0.1	8

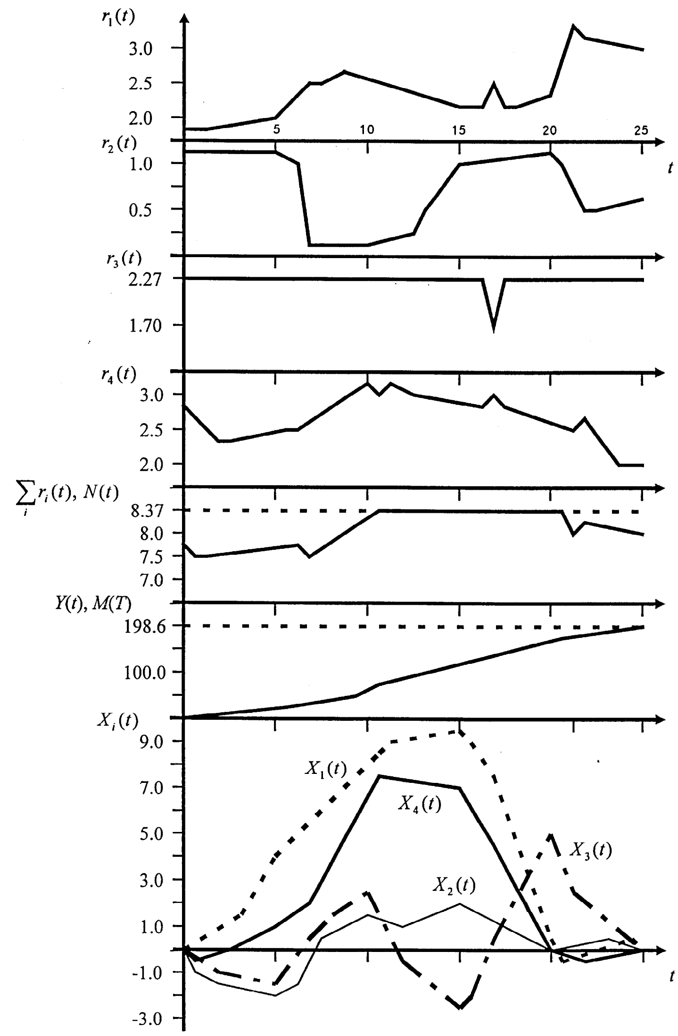

Figure 3. Evolution of the resource usage, consumption and inventory (activity) levels.

7. Concluding remarks

Consideration of doubly constrained resources is important from both theoretical and practical standpoints in aggregate production and project management. It is also of special interest due to many real-life applications where considering only either resource constraint is inaccurate and may result in improper managerial decision to be made. The continuous time approach to scheduling resources is introduced in the paper in a form of the optimal control problem. Based on the maximum principle, optimal resource usage regimes are derived for linear and concave speed-resource functions. As a result, two time-decomposition methods are developed to solve the problem with one of them maximizing the Hamiltonian

Table 2. Deviation of objective function values relative to the 20% accuracy GAMS solutions, in percents

Number of product types	Projected-gradient		Global search		Combined method	
	Computation time (min)					
	4	8	4	8	4	8
5	98	110	80	87	105	117
10	95	105	68	74	101	111
15	90	98	58	63	96	106
20	82	89	50	55	91	100

analytically and the other doing this numerically. Sufficient conditions for the methods to provide a global optimal solution are also discussed. As shown in Table 2, the best computational results are obtained when combining the methods. Moreover, in almost all experiments, the combined method computed in minutes, improved the objective computed within hours by the Non-Linear Programming optimization GAMS solver set up with 20% accuracy. Table 2 gives the average deviation of the objective obtained on a PC-486 by the three methods relative to that of the GAMS' solution.

Acknowledgment

The authors gratefully acknowledge Prof. Jan Weglarz whose ideas initiated this research, and whose valuable suggestions helped to complete this manuscript.

References

Bergstrom, G. and B. Smith (1970), Multi-item production planning - an extension of the HMMS rules, *Management Science* 16(10): B614–B629.

Blazewicz, J., K.Ecker, E. Pesch, G.Schmidt and J.Weglarz (1996), *Scheduling Computer and Manufacturing Processes*. Springer Verlag, Berlin.

Bryson, A. E., and J. Y-C. Ho (1975), *Applied Optimal Control*. Hemisphere Publishing Corporation, Washington, DC.

Burden, R.L., and J.D. Faires (1989), *Numerical Analysis*. PWS-KENT, Boston.

Dubovitsky, A.Y., and A.A. Milyutin (1981), Theory of the Maximum Principle, *in Methods of Theory of Extremum Problems in Economy,* pp 6–47. Nauka, Moscow.

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

Hartl, R.F., Sethi, S.P., and R.G. Vickson (1995), A survey of the maximum principles for optimal control problems with state constraints, *SIAM Review* 37(2): 181–218.

Ilyutovich, A., and E. Khmelnitsky (1991), *A numerical method for an optimal control problem with constraints on state variables, based on the maximum principle*. VNIISI Publishing, Moscow.

Janiak, A., and A. Stankiewicz (1983), The equivalence of local and global time-optimal control of a complex of operations, *International Journal of Control* 38: 1149–1165.

Kimemia, J.G., and S.B. Gershwin (1983), An algorithm for the computer control of a flexible manufacturing system, *IIE Transactions* 15: 353–362.

Khmelnitsky, E., and K. Kogan (1994), Necessary optimality conditions for a generalized problem of production scheduling, *Optimal Control Applications and Methods* 15: 215–222.

Khmelnitsky, E., K. Kogan, and O. Maimon (1995), A maximum principle based combined method for scheduling in a flexible manufacturing system, *Discrete Event Dynamic Systems* 5: 343–355.

Khmelnitsky E. and K. Kogan (1996), Optimal policies for aggregate production and capacity planning under rapidly changing demand conditions, *International Journal of Production Research* 34(7): 1929–1941.

Kogan, K., and E. Khmelnitsky (1996), An optimal control model for continuous-time production and setup scheduling, *International Journal of Production Research* 34(3): 715–725.

Leachman, R.C., A., Dincerler, and S., Kim (1990), Resource-constrained scheduling of projects with variable-intensity activities, *IIE Transactions* 22: 31–39.

Pinedo M. (1995), *Scheduling: Theory, Algorithms and Systems*. Prentice Hall, Englewood Cliffs.

Sethi, S.P., M.I. Taksar and Q. Zhang (1992), Capacity and production decisions in stochastic manufacturing systems: an asymptotic optimal hierarchical approach, *Production and Operations Management* 1(4): 367–392.

Weglarz, J. (1981), Project scheduling with continuously-divisible, doubly constrained resources, *Management Science* 27: 1040–1053.