Scheduling in production, supply chain and Industry 4.0 systems by optimal control: fundamentals, state-of-the-art and applications

Alexandre Dolgui, Dmitry Ivanov, Suresh Sethi, Boris Sokolov

Scheduling in production, supply chain and Industry 4.0 systems by optimal control: fundamentals, state-of-the-art, and applications

Alexandre Dolgui[1] , Dmitry Ivanov[2*], Suresh Sethi[3], Boris Sokolov[4,5]

[1]Automation, Production and Computer Sciences Dept.
IMT Atlantique LS2N - CNRS UMR 6004 La Chantrerie
email: alexandre.dolgui@imt-atlantique.fr

[2*] Berlin School of Economics and Law
Department of Business Administration; Professor for Supply Chain Management
10825 Berlin, Germany
Phone: +49 30 85789155
E-Mail: divanov@hwr-berlin.de

[3] Eugene McDermott Chair
Center for Intelligent Supply Networks (C4iSN)
Naveen Jindal School of Management, Mail Station SM30
The University of Texas at Dallas
800 W. Campbell Rd., Richardson, Texas 75080-3021
E-mail: sethi@utdallas.edu

[4] Saint Petersburg Institute for Informatics and Automation of the RAS (SPIIRAS)
V.O. 14 line, 39 199178 St. Petersburg, Russia
E-Mail: sokol@iias.spb.su

[5]ITMO University, St. Petersburg, Russia
E-Mail: sokol@iias.spb.su

* Corresponding author
Dmitry Ivanov

Abstract

Specific scheduling problems with complex hybrid logical and terminal constraints, non-stationarity in process execution as well as complex interrelations between dynamics in process design, capacity utilization, and machine setups require further investigation and the application of a broad range of methodical approaches. One of these approaches is optimal control. The objectives of this survey are twofold. The first objective is to derive major contributions, application areas, limitations, as well as research and application recommendations for the future regarding optimal control applications to scheduling. The second objective is to explain control engineering models in terms of industrial engineering and production management. In this paper, we provided a survey on the applications of optimal control to scheduling in production, supply chain, and Industry 4.0 systems with a focus on the deterministic maximum principle. Optimal control approaches take a different perspective as mathematical programming methods which represent schedules as trajectories. We consider optimal control models, performance analysis qualitative methods, and computational methods for optimal control. We provide a brief historic overview and clarify major mathematical fundamentals whereby the control engineering terms are brought into correspondence with industrial engineering and management. The survey allows the grouping of models with only terminal constraints with application to master production scheduling, models with hybrid terminal-logical constraints with applications to short term job and flow shop scheduling, and hybrid structural-terminal-logical constraints with applications to customized assembly systems such as Industry 4.0. Computational algorithms in state, control, and adjoint variable spaces are discussed. Finally, we derive major contributions, application areas of different control methods, and their limitations. This paper also provides recommendations for future research and applications.

Keywords: optimal program control, deterministic control, maximum principle, scheduling, attainable sets, algorithms

1. Introduction

Short-term scheduling belongs to the fundamentals of scheduling theory. It considers jobs containing operation chains with equal (i.e., flow shop) or different (i.e., job shop) machine sequences and different processing times. The operations need to be scheduled for machines with different processing power subject to some criteria such as makespan, lead time, or due dates (Blazewicz et al. 2001, Pinedo 2008, Dolgui and Proth 2010, Werner and Sotskov 2014).

Over the last decades, mathematical optimization applications to scheduling problems have been studied from different perspectives whereby significant progress can be observed in the development of rigorous theoretical models and efficient solution techniques. Lauff and Werner (2004), Jungwattanakit et al. (2009), Sotskov et al. (2013), Choi et al. (2013), Harjunkoski et al. (2014), Bożek and Wysocki (2015), Ivanov et al. (2016a,c) have pointed out that specific scheduling problems with complex hybrid logical and terminal constraints, non-stationarity in process execution as well as complex interrelations between dynamics in process design, capacity utilization, and machine setups require further investigation and the application of a broad range of methodical approaches.

Optimal control approaches take a different perspective as mathematical programming methods which represent schedules as trajectories. Optimal control applications to scheduling problems are encountered in production systems with single machines (Giglio 2015), job sequencing in two-stage production systems (Lou and Van Ryzin 1989, Sethi and Zhou 1996) and multi-stage machine structures with alternatives in job assignments and intensity-dependent processing rates such as flexible manufacturing systems (Sharifnia et al. 1991, Maimon et al. 1998, Yang et al. 1999, Ivanov and Sokolov 2013a, Pinha et al. 2015), supply chains as multi-stage networks (Ivanov and Sokolov 2012, Ivanov et al. 2013), and Industry 4.0 systems based on data interchange between the product and stations, flexible stations dedicated to various technological operations, and real-time capacity utilization control (Ivanov et al. 2016a).

This survey considers research on optimal control applications to production scheduling with analysis of model parameters and computational algorithms published in the last 55 years. The objectives of this survey are twofold. The first objective is to derive major contributions, application areas, limitations, as well as research and application recommendations for the future regarding optimal control applications to scheduling. The second objective is to explain control engineering models in terms of industrial engineering and production management. We provide a survey on the applications of optimal control to scheduling in production, supply chain, and Industry 4.0 systems whereby we restrict ourselves to deterministic maximum principle-based approaches and omit detailed analysis of stochastic optimal control approaches as well as dynamic programming algorithms. Regarding the related topics which are not covered in this paper, we refer to the surveys by Sethi (1978, 1984) for applications of the maximum principle to production-inventory problems and to works (Lou et al. 1994, Sethi and Zhang 1994, Presman et al. 1995, Samaratunga et al. 1997, Presman et al. 1997, 2000, Feng and Yan 2000, Sethi et al. 2002, Khmelnitsky et al. 2011) which extend this survey to stochastic scheduling problems.

The survey follows the structure "optimal control models – performance analysis qualitative methods – computational methods". In Section 2, a brief historic overview and clarification of major mathematical fundamentals are provided. The control engineering terms are brought into correspondence with industrial engineering and management. In Section 3, optimal control models for scheduling in production, supply chain, and Industry 4.0 systems are presented and classified in terms of their analytical contents

and application areas. Section 4 deals with attainable (reachable) sets as a method of qualitative analysis of optimal control performance. Section 5 is devoted to computational algorithms regarding state, control, and adjoint variable spaces. In Section 6, we derive major contributions, application areas of different methods, and their limitations. This section also provides recommendations for future research and applications. The paper concludes in Section 7 by summarizing the insights from this survey.

2. Fundamentals of optimal control models with applications to scheduling

2.1. Historical development

Optimal control approaches represent schedules as trajectories. The first studies in this area were devoted to inventory control. One of these studies (Eilon 1961) was published in the first volume of the *International Journal of Production Research* (*IJPR*). Later, Hwang *et al*. (1967, 1969) were among the first to apply optimal control and the maximum principle to multi-level and multi-period master production scheduling which determined the optimal control (i.e., production) with the corresponding state (i.e., inventory) trajectory. Albright and Collins (1977) developed a Bayesian approach to the optimal control of continuous industrial processes. Bedini and Toni (1980) developed a dynamic model for the planning of a manufacturing system. The maximum principle has been used to formulate the problem and obtain a solution. A large research area of flexible manufacturing systems and their dynamics has been examined in numerous studies (e.g., Stecke and Solberg 1981). The stream of production scheduling was continued by Kimemia and Gershwin (1985), Kogan and Khmelnitsky (1996), and Khmelnitsky *et al*. (1997), who applied the maximum principle in discrete form to the planning of continuous-time flows in flexible manufacturing systems

The origins of control scheduling techniques can be found in network planning, dynamic programming and waiting line theory (Moiseev 1974, Sotskov et al. 2013). Consider the graph in Fig. 1.
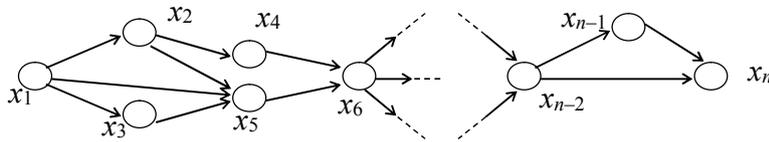

Fig. 1. Network planning graph

In the classical network planning theory, the state x_i of the *i*-job can be determined subject to Eq. (1):

$$x_i \Leftrightarrow T_i = \frac{Q_i}{\eta_i} \ , \qquad (1)$$

where T_i is the task time to process the *i*-job, Q_i is the processing volume of the *i*-job, and η_i is the intensity (i.e., processing rate) of processing the *i*-job, whereby the jobs *i= 1,...,n* are interconnected with each other in terms of precedence constraints "or"/"and". It is notable that even if elements of dynamics can be observed in the aforementioned system regarding process deployment in time, job execution dynamics itself have not been considered explicitly. In other words, operations execution has been treated in a static manner since task times nwere assumed to be fixed. In reality, task times may change in job execution dynamics. As such, dynamics of job execution requires an explicit description of job process execution, distribution (allocation) of resources required for job execution, and changes in the job states and the respective control inputs. The aforementioned dynamic interpretation of job execution was extensively developed in the 1970s (Zimin and Ivanilov 1971, Moiseev 1974) within the framework of optimal control theory.

Optimal control theory is devoted to determining some functions known as controls that lead to optimization(minimization or maximization) of an objective (Pontryagin et al. 1964, Athaus and Falb 1966, Lee and Markus 1967, Moiseev 1974, Bryson and Ho 1975, Hartl and Sethi 1984, Soner 1986, Gershwin 1994, Sethi and Thompson 2000). This theory evolved over the centuries based on calculus variation principles developed by Fermat, Lagrange, Bernulli, Newton, Hamilton and Jacobi. In the 20th century, two computational fundamentals of optimal control theory, the maximum principle (Pontryagin et al., 1964) and the dynamic programming method (Bellmann 1972), were developed. These methods extend the classical calculus variation theory which is based on control variations of a continuous trajectory and observing the performance impact of these variations at the trajectory end. Since control systems in the middle of the 20th century were increasingly characterized by piecewise continuous functions (such as 0-1 switch automats), the development of the maximum principle and the dynamic programming was needed for solving problems with complex constraints on state and control variables. This section aims to clarify the notions of state, control, and performance at the optimal control model level, bridging these notions to industrial management and engineering. Moreover, the computational level will be considered and the maximum principle, adjoint equation system, and transversality conditions will be clarified.

2.2. Major elements of an optimal control model for scheduling

Consider the evolution of a quantifiable object (e.g., inventory or production quantity) in time

$$\dot{\mathbf{x}}(t) = \mathbf{f}(t, \mathbf{x}(t), \mathbf{u}), \ \dot{\mathbf{x}}(t) = \frac{d\mathbf{x}}{dt}, t_0 \leq t \leq t_f$$

where $\mathbf{u} \in R^m$ is the control, $\mathbf{x} \in R^n$ is the terminal system state vector, $\mathbf{f} \in R^n$ is a given function, R^n is Euclid space of dimensionality n, t_0 is the initial point in time and t_f is final point in time. For example, in a job shop, $\mathbf{x}(t)$ can describe a *buffer* such as production quantity or inventory, and control variable $\mathbf{u}(t)$ can describe processed flow volume of a job at a machine. The system state vector is determined by the evolution of state variables $\mathbf{x}(t)$ that characterize the system at each point in time. State evolution in dynamics is determined by control variables $\mathbf{u}(t)$ that correspond to the decisions of a person or an algorithm governing the system dynamics.

In real practice of control engineering, control variables are typically considered bounded piecewise continuous functions. Examples of controls in operational systems include processing rates ofmachines in manufacturing or shipment rates in transportation. In production scheduling, binary control variables are used to describe the assignment of a job to a machine. Consider an example (Fig. 2).
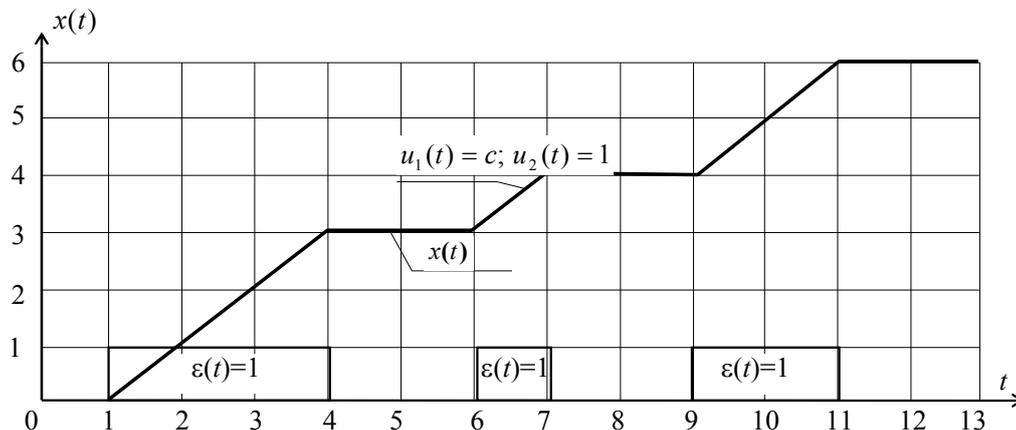
Figure 2. Example of job execution dynamics

Fig. 2 represents job execution dynamics in which the non-stationarity of job execution is reflected. Machine availability times (subject to maintenance or other restrictions) are given as a preset matrix time function ε(t). The $u_1(t)$ is a control decision variable characterizing the processed flow volume at the machine subject to some processing intensity and to another control decision $u_2(t) = 1$, namely an assignment of the job to this machine. It can be observed in Fig. 2 that job processing at the machine starts at $t = t_1$ and ends at $t = t_{11}$ whereby a flow (volume) of six product units is produced.

A standard dynamic network process control model has the form of Eqs. (2)-(8).

Mathematical model of job execution control

$$\frac{dx_i^{(o)}}{dt} = \dot{x}_i^{(o)} = u_i^{(o)}\gamma_+(a_i^{(o)} - x_i^{(o)})\left[\prod_{\bar{\alpha} \in \Gamma_{i1}^-}\gamma_-(x_{\bar{\alpha}}^{(o)} - a_{\bar{\alpha}}^{(o)}) + \sum_{\bar{\beta} \in \Gamma_{i2}^-}\gamma_-(x_{\bar{\beta}}^{(o)} - a_{\bar{\beta}}^{(o)})\right] \qquad (2)$$

where $x_i^{(o)}$ is the current state of the execution of the i-job, $\gamma_+(a_i^{(o)} - x_i^{(o)})$ are asymmetric step functions that set logical constraints to avoid overproduction in regard to the planned volume $a_i^{(o)}$ (see further in Eq. 4), $\prod_{\bar{\alpha} \in \Gamma_{i1}^-}\gamma_-(x_{\bar{\alpha}}^{(o)} - a_{\bar{\alpha}}^{(o)}) + \sum_{\bar{\beta} \in \Gamma_{i2}^-}\gamma_-(x_{\bar{\beta}}^{(o)} - a_{\bar{\beta}}^{(o)})$ are asymmetric step functions that set logical precedence constraints "and"/"or" regarding the i-job (cf. Eq. 5). The job numbers $\bar{\alpha}$ and $\bar{\beta}$ from the sets Γ_{i1}^- Γ_{i2}^- determine preceding jobs for the i-job. $u_i^{(o)}$ is a continuous control variable subject to the maximum processing intensity of operations execution (Eq. 3).

$$0 \le u_i^{(o)} \le c_i^{(o)} \qquad (3)$$

The peculiarity of optimal control models are constraints. In practice, the system trajectory cannot belong to some areas in the space R^n. The above-mentioned boundary conditions belong to the constraints on the trajectory that can differ regarding fixed or free ends. In scheduling models, we typically have fixed end boundary conditions (Eq. 7). Next, the constraints on control need to be defined. Since the search for optimal control is performed within the class of functions $u(t)$ that depend only on t, this problem class is called optimal program control. Typically, in production scheduling we have optimal program control problems with two types of constraints on control, i.e., terminal and logical constraints. In future systems of Industry 4.0 production, the third type of hybrid constraints, i.e., structural-logical-temporal constraints will be used (see further in this paper in Sect. 3.3). Terminal constraints (see Sect. 3.1 in this paper) describe limited control resources (e.g., Eqs. 4 and 5). An example of a terminal constraint is a capacity restriction at a machine. Logical constraints (see further in this paper in Sect. 3.2) describe α-precedence relations of the type "or" and β-precedence relations of the type "and" regarding the job sequences in the jobs (Eq. 5).

$$\gamma_+(a_i^{(o)} - x_i^{(o)}) \qquad \in \{0,1\} \qquad (4)$$

$$\prod_{\bar{\alpha} \in \Gamma_{i1}^-}\gamma_-(x_{\bar{\alpha}}^{(o)} - a_{\bar{\alpha}}^{(o)}) + \sum_{\bar{\beta} \in \Gamma_{i2}^-}\gamma_-(x_{\bar{\beta}}^{(o)} - a_{\bar{\beta}}^{(o)}) \qquad \in \{0,1\} \qquad (5)$$

Terminal constraints (4) describe restrictions for limiting job execution subject to fixed processing volume $a_i^{(o)}$. Constraints (5) determine precedence relations by blocking job D until the previous operations $D_{\bar{\alpha}}, D_{\bar{\beta}}$ have been completed. Optimal control scheduling models with only terminal constraints typically address the master production scheduling level (Hwang et al. 1967, 1969, Kimemia and

Gershwin 1983, Jiang and Sethi 1991, Khmelnitsky et al. 1997, Kogan and Khmelnitsky 2000). Scheduling models with both terminal and logical constraints can also be applied to flow shop and job shop scheduling (Kalinin and Sokolov 1985, 1987, Ivanov and Sokolov 2013a, Ivanov et al. 2016a,c) as well as to supply chain scheduling (Ivanov and Sokolov 2012, Ivanov et al. 2013). Along with constraints (3), integral constraints on resources are typically written in the form of Eq. (6):

$$\int_{t_0}^{t_f} F(t, x(t), u) dt \leq 0 \qquad (6)$$

In order to assess the scheduling results, optimal control models define boundary conditions, i.e., start and end conditions such as Eq. (7)

$$\mathbf{h}_0(\mathbf{x}(t_0)) \leq \mathbf{0}; \ \mathbf{h}_1(\mathbf{x}(t_f)) \leq \mathbf{0}, \qquad (7)$$

where $\mathbf{h}_0^{(o)}$, $\mathbf{h}_1^{(o)}$ are known differentiable functions that determine the start and end conditions of the vector $\mathbf{x}(t)$. For example, an initial condition might specify that the executed volume of jobs at the beginning of the scheduling horizon is equal to zero. End conditions could reflect the desired end state, i.e., the completion of the jobs by the time t_f.

The optimal program control vector $\mathbf{u}(t)$ and the state trajectory $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}, t)$ should be determined so that the boundary conditions are met; in other words, the desired values of the scheduling performance indicators should be achieved as an analogy to goal programming.

The performance assessment is designed in control systems in the following way. The control of system (2) is directed towards attainability of some performance. The performance metrics (or functionals, in terms of optimal control) can be grouped into terminal (i.e., flow-oriented metrics, e.g., work-in-process inventory or planned production volume subject to a customer demand) and integral (e.g., due dates) functionals (Eq. 8):

$$J = \sum_{i=1}^{n} (a_i^{(o)} - x_i^{(o)}(t_f))^2 + \int_{t_0}^{t_f} \varphi(\mathbf{u}(\tau)) d\tau \qquad (8)$$

Performance reachability depends on the selection of the trajectory $\mathbf{x}(t)$ and control $\mathbf{u}(t)$. In general, scheduling problems in terms of control have been formulated for searchingoptimal program control for dynamic system (2) subject to minimization of the functional (8) subject to constraints (3)-(7). The first term in Eq. (8) characterizes the relationship between planned job execution volume and the volume that can be realized in the computed schedule. The second term depends on the interpretation of the function $\varphi(\mathbf{u}(\tau))$ and frequently plays the role of delay penalties.

2.3. *Major elements of optimal control computational procedures*

Necessary optimality conditions can be derived from the maximum principle (Sethi 1978, Hartl et al. 1995, Afanasiev et al. 1996, Khmelnitsky et al. 1997, Sethi and Thompson 2000). Consider control system (9):

$$\dot{\mathbf{x}}(t) = f(t, \mathbf{x}(t), \mathbf{u}(t)), \ t_0 \leq t \leq t_f, \ \mathbf{x}(t_0) = \mathbf{x}_0, \ \mathbf{u}(t) \in U, \ J = F(\mathbf{x}(t_f)) \rightarrow \mathbf{min} \qquad (9)$$

Let us introduce a scalar Hamiltonian function H and adjoint vector system $\boldsymbol{\psi} \in R^n$ in Eq. (10):

$$H(t, \mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\psi}(t)) = \boldsymbol{\psi}^{\mathbf{T}}(t) \mathbf{f}(t, \mathbf{x}(t), \mathbf{u}(t)),$$

$$\dot{\psi}(t) = -\frac{\partial H}{\partial \mathbf{x}}(t, \mathbf{x}(t), \mathbf{u}(t), \psi(t)), \tag{10}$$

$$\psi(t_f) = -\frac{\partial F(\mathbf{x}(t))}{\partial \mathbf{x}}\bigg|_{t=t_f}, \tag{11}$$

Adjoint vector system plays the role of dual models in linear programming. Coefficients of the adjoint systems can be interpreted as Lagrange multipliers. Under assumptions that $\mathbf{u}(t)$ is optimal control and $\mathbf{x}(t)$ and $\psi(t)$ are the trajectory and adjoint system satisfying (10) and (11), the function $H(t, \mathbf{x}(t), \mathbf{u}(t), \psi(t))$ reaches its maximum for $\mathbf{x}(t)$ at the point $\mathbf{u}(t)$. Then Eq. (12) holds:

$$\mathbf{u} = \mathbf{u}(t, \mathbf{x}(t), \psi(t)) \tag{12}$$

Subsequently, Eq. (12) is brought into correspondence with (10) and (11). As a result, a two-point boundary problem for a system of ordinary differential equations involving $\mathbf{x}(t)$ and $\psi(t)$ is formed. The optimal solution is now bounded by this differential system. Note that Eqs. (9)-(12) in general provide only the necessary conditions for the existence of an optimal solution. For linear control systems, these maximum principles provide both optimality and the necessary conditions (Ivanov and Sokolov 2010). This fact requires further optimality study for each concrete case of application.

Typically, computational procedures start with a nominal solution that satisfies both main and adjointdifferential systems. Then this solution is modified by integrating main and adjoint systems by control variations towards a Hamiltonian increase. During this procedure, at $t = T_f$ transversality conditions are evaluated. Transversality conditions are the end conditions of the adjoint system. The adjoint variables can be interpreted as dynamic priorities of jobs and play here the role of "shadow" prices in linear programming models. However, in contrast to those canonical forms where "shadow" prices are fixed, the adjoint variables change dynamically. These changes are subject to the contribution of a particular operations assignment and scheduling (i.e., machine and time windows) to the change in performance assessment functions. Consequently, at each point in time the dynamic priorities of jobs may be changed if a newly arrived job provides a better contribution to the performance functional.

Various algorithms based on Pontryagin's maximum principle have been developed in the aforementioned area. In essence, these algorithms reduce the non-classical calculus variation problem to a two-point boundary problem (Zimin and Ivanilov 1971, Moiseev 1974). Regarding large-scale problems, two major shortcomings of this approach have been observed. First, the two-point problem became a multiple-point problem with jumps in the adjoint variables at the end of job processing because of terminal constraints (4). The explanation of this effect can be seen in step function differentiating. The step functions became general delta-functions. The second problem was related to the numerical computation of the boundary problems for deriving the initial conditions of the adjoint variables which are needed to compute the optimal schedule. Questions of convergence, optimality, and the necessary conditions remained open. Special heuristic algorithms were developed in this area to overcome these problems (Chernousko & Lyubushin (1982).

In order to resolve the problems with the step functions and the respective non-linearity in the right-hand parts of Eq. (2), Moiseev (1974) and Kalinin and Sokolov (1985)] developed another variant to describe the right-hand parts of Eq. (2) as terminal constraints in the control and functional space (Eq 13), whereby the functional (8) was modified as shown in Eq. (14).

$$d_i = x_i^{(o)} \left[\prod_{\bar{\alpha} \in \Gamma_{i1}^-} (a_{\bar{\alpha}}^{(o)} - x_{\bar{\alpha}}^{(o)}) + \sum_{\bar{\beta} \in \Gamma_{i2}^-} (a_{\bar{\beta}}^{(o)} - x_{\bar{\beta}}^{(o)}) \right] = 0 \qquad (13)$$

$$J_p = J + \int_{t_0}^{t_f} \sum_{i=1}^{n} f_i d_i(\tau) d\tau , \qquad (14)$$

where f_i are given penalty functions if the conditions (13) are not fulfilled (in case of objective function minimization). d_i are the given terminal constraints.

The additional component in (14) describes the penalty for non-fulfillment of the condition (13). However, the search complexity for λ_i was similar to the search complexity for the jump values in the adjoint variables. In addition, the classical optimal control models for scheduling do not consider aspects such as setups, indivisibility of resources for job execution at any point of time, and bans on interruptions of the job execution. A special group of problems can be considered regarding uncertainty and risks, schedule stability, flexibility and robustness analysis, as well as multi-criteria resolution. The following Section 3 will describe how the given problems can be resolved through several modifications of the model (2)-(8).

3. Optimal control models for job scheduling in production, supply chain, and Industry 4.0

Scheduling problem statements differ in terms of fixed or variable process and operations sequences. In this regards, we structure the analysis in this section in accordance to Fig. 2.

Fig. 2. Classification of scheduling problems solved by optimal program control

Let us analyse the problem statements, models, and algorithms in Sects 3.1-3.3 in detail.

3.1. Model with terminal constraints

According to Kimemia and Gershwin (1983), Kogan and Khmelnitsky (1996), Khmelnitsky et al. (1997) and Maimon et al. (1998), the optimal control model with terminal constraints can be applied to production scheduling in the following settings.

Problem description

The task consists of scheduling jobs on machines subject to cost minimization and on-time demand fulfillment regarding quantity and due date with limited machine capacity considerations.

The problem formalization is presented as follows. Note that, in line with the referenced studies Kogan and Khmelnitsky (1996), Khmelnitsky et al. (1997) and Maimon et al. (1998), Sect. 3.1 contains original notations which are different from notations in other sections of this paper and are valid for Sect. 3.1 only.

Set indexes

t_f time of the planning horizon

i product index, $i = 1....,l$

k machine index, $k = 1,..., J(k)$;

j machine state index, $j = 1,....J(k)$;

m job index, $m = 1,...M(i)$;

Parameters

D_{im} due date of job

X_{im} demand

T_{kj} minimum setup time

v_{ikj} capacity of machine k being in state j when producing product i (in case of product consumption, this parameter is negative)

c_i^c subcontracting cost per unit product c_i^r ;

c_i^r purchasing cost of all raw materials required per unit product i ;

c_{im}^p penalty for violating the due date of job m

Decision variables

$X_i(t)$ cumulative buffer level of i-product at time t ;

$V_{kj}(t)$ a dimensionless continuous function that reflects the current state of machine; $V_{kj}(t) = [0;1]$

$w_{kj}(t)$ control variable characterizing actual loading of machine k relative to its production capacity $w_{kj}(t) \in [0,1]$;

$u_{kj}(t)$ control variable, characterizing the rate of state j on machine k, i.e., the setup rate

Process control model

$$\dot{X}_i(t) = \sum_{kj} w_{kj}(t) v_{ikj} , \tag{15}$$

$$\dot{V}_{kj}(t) = u_{kj}(t) , \tag{16}$$

Eq. (15) describes production process. Eq. (16) describes the setup process.

Constraints

The setup process is characterized by two types of constraints. First, the rate of transformation from current state is equal to the transformation rate into a new state (Eq. 17)

$$\sum_{j} u_{kj}(t) = 0, \forall_{k}.\tag{17}$$

Second, the rate of transformation from and to state j is bounded (Eq. 18)

$$-\overline{T_{kj}}^{-1} \le u_{kj}(t) \le T_{kj}^{-1},\tag{18}$$

where T_{kj} is the minimum time of setting up machine k on state j, and $\overline{T}_{kj}$ is the minimum time of setting down machine k on state j. Eq. (17) characterizes the balance between the setup intensities of the j-machine by state transition. Eq. (18) constraints the maximum intensity (i.e., processing speed) of the operation processing at the j-machine.

A restriction on sequence-dependent setup process interaction (19) states that the j-machine in the setup state cannot execute processing operations at the same time.

$$0 \le \Theta(w_{kj}(t)) \le V_{kj}(t),\tag{19}$$

where $\Theta(w)$ is the step (Heaviside) function, i.e., $\Theta(w) = 1$ if $w > 0; \Theta(w) = 0,$ if $w \le 0$.

Objective function

$$\min \int_{0}^{t_f} \left[\frac{1}{2} \sum_{kj} c_{kj}^{s} (u_{kj}(t))^2 + \sum_{i} C(X_i(t)) \right] dt + \sum_{i} \left[(c_i^c X_i(0) + c_i^r (X_i(T) - X_i(0)) \right] + \frac{1}{2} \sum_{im} c_{im}^{p} (\max\{0, X_{im} - X_i(D_{im})\})^2, \tag{20}$$

where $X_i(t)$ is positive, and the function $C(X)$ reflects the inventory carrying cost. Otherwise, it reflects the stock out cost for subproducts (internal shortages) and penalties for violating demands for end products (external shortages)

$$C(X_i(t)) = \frac{1}{2} b_i^1 [X_i(t) - h_i(t)], \text{ if } X_i(t) - h_i(t) \ge 0,$$

$$C(X_i(t)) = \frac{1}{2} b_i^2 [X_i(t) - h_i(t)], \text{ if } X_i(t) - h_i(t) < 0,$$

and h_i is written

$$h_i(t) = \max_{m} \min\{X_{im}, X_i(D_{im})\Theta(t - D_{im})\}.$$

The scheduling problem is now formulated in the following way: minimize the functional constraints (20) to transit the dynamic object (14)-(15) from the given initial state $X_i(0), V_{k_i}(0)$ to the required final state (defined in Eq. (20)) subject to constraints (17)-(18).

With the help of the maximum principle, Kogan and Khmelnitsky (1996), (2000) formulated the Hamiltonian function (21) and the respective adjoint equation systems (22).

$$H = -\frac{1}{2} \sum_{kj} c_{kj}^{s} (u_{kj}(t))^2 - \sum_{i} C(X_i(t)) + \sum_{kj} \psi_{kj}^{v}(t) u_{kj}(t) + \sum_{i} \psi_{i}^{x}(t) \sum_{kj} w_{kj}(t) v_{ikj},\tag{21}$$

$$d\psi_{kj}^{v}(t) = C_{x}'(X_i(t)) dt - \sum_{m} c_{im}^{p} (\max\{0, X_{im} - X_i(D_{im})\}) d\Theta(t - D_{im}), \ \psi_{i}^{x}(T + 0) = 0,\tag{22}$$

$$d\psi_{kj}^{V}(t) = -a_{kj}(t) dt - d\mu_{kj}(t), \ \psi_{kj}^{V}(T + 0) = 0.$$

As a result of Hamiltonian maximization, it becomes possible to compute the optimal schedule for material flow processing at a machine complex. The schedule is described using three logical conditions regarding control w (Eq. 23):

$$v_{kj}(t), \text{ if } \sum_i \psi_i^x(t)v_{vikj} > 0$$

$$f_n(w_{kj}(t)) = 0, \text{ if } \sum_i \psi_i^x(t)v_{vikj} < 0$$

$$\left[0, V_{kj}(t)\right], \text{ if } \sum_i \psi_i^x(t)v_{vikj} = 0 \tag{23}$$

Eq. (23) was derived in the studies by Kogan and Khmelnitsky (1996) and (2000). It underlines the advantages of the maximum principle as compared to other optimal control approaches. The developed approach allows explicit determination of information about the structure of optimal operating modes of work stations prior to computational experiments. In addition, existence conditions were found, i.e., conditions where the Hamiltonian function equals Eq. (21). Kogan and Khmelnitsky (1996) provided evidence of the practical application of the developed approach for a flexible manufacturing system with four machines and six jobs for the case of fixed job processing technology.

A number of questions have arisen from the analysis of the aforementioned approach. First, the computation error estimation as a consequence of replacing the step function in Eq. (19) with a sequence of approximated functions needs to be named. Second, the convergence in the numerical computation procedure needs to be analysed. The Hulquist's (1988) modification does not guarantee convergence for the considered form of non-linear systems in general. The selection of the initial conditions for the adjoint system is an important precondition for algorithmic convergence. In addition, optimality proofs need to be addressed. As known, the maximum principle for non-linear dynamic systems is a necessary optimality condition. The sufficiency condition needs to be analysed separately (Athaus and Falb 1966, Bryson and Ho 1975, Pontryagin et al. 1964).

At the same time, it can be observed that the initial model (2)-(8) has been extensively extended by the model (15)-(23). Another extension in Sect. 3.2 will be made regarding the dynamic system (15)-(16) itself and the technical and technological constraints (Eq. 17 and 18) subject to practical needs of real-time job and flow shop environments.

3.2 Model with hybrid terminal-logical constraints: applications to job shop and flow shop scheduling

The model in Sect. 3.1 has been proved to be a working tool for production scheduling when flow control is considered and no precedence operation relations exist in jobs. For job shop scheduling, precedence relations exist. The studies by Kalinin and Sokolov (1985), (1987), Sokolov and Yusupov (2002) resolved the aforementioned problems with multiple criteria, bans on execution interruptions and non-stationarity for large-scale scheduling problem. The following studies by Ivanov and Sokolov (2013), Ivanov et al. (2013), and Ivanov et al. (2016a,b,c) developed a special form of hybrid terminal-logical constraints that allows the application of maximum principle-based optimal control models to job and flow shop scheduling for flexible manufacturing systems, supply chains, and Industry 4.0 networks.

Consider the following simplified example that unites the model of operation execution control and flow control (Eqs. 24-29).

$$\dot{x}_i^{(o)} = u_i^{(o)}; \dot{x}_i^{(n)} = u_i^{(f)} \ ; \tag{24}$$

$$u_i^{(o)} \left[\prod_{\overline{\alpha} \in \Gamma_{i1}^-} (a_{\overline{\alpha}}^{(o)} - x_{\overline{\alpha}}^{(o)}) + \sum_{\overline{\beta} \in \Gamma_{i2}^-} (a_{\overline{\beta}}^{(o)} - x_{\overline{\beta}}^{(o)}) \right] = 0 , \tag{25}$$

$$u_i^{(o)}(t_0) \in \{0,1\} , \tag{26}$$

$$0 \le u_i^{(f)} \le c_i^{(o)} u_i^{(o)} , \tag{27}$$

$$x_i^{(o)}(t_0) = 0; \; x_i^{(o)}(t_f) = a_i^{(o)} , \tag{28}$$

$$J = \sum_{i=1}^{n} (a_i^{(o)} - x_i^{(o)}(t_f))^2 + \int_{t_0}^{t_f} \varphi(\mathbf{u}(\tau)) d\tau , \tag{29}$$

In contrast to constraints (5) and (13), Eq. (25) contains both control and state variables from Eq. (24). In this form, constraints (25) are new in optimal control applications to scheduling and allow the use of optimal control for assignment and sequencing tasks. In the following part of this paper, we will show how to transit from impulse controls in Eq. (26) to interval [0,1] for these controls in the class of piecewise continuous functions. The controls will take binary values that will allow their usage in assignment problems. In terms of sequencing, we observe that in Eq. (27), the controls from Eq. (24) are contained. These controls activate Eq. (27) that describes the flow distribution control problem and sequencing subject to processing intensity of the machines. The given constraint system unifies two large model classes: the flow models in system dynamics and operation control and resource distribution models in network planning theory.

The basic *idea* of this approach is that operation execution and machine availability are dynamically distributed in time over the planning horizon. As such, not all operations and machines are involved in decision making at the same time (Fig. 3).
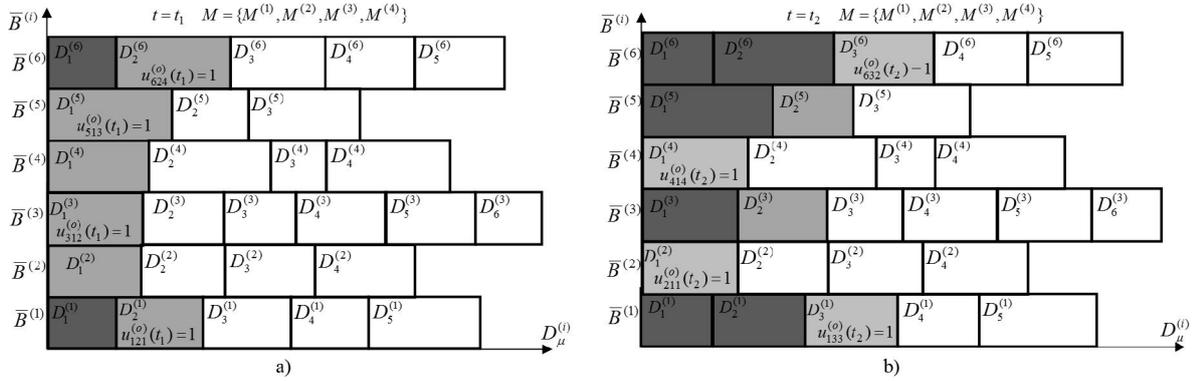

Fig. 3. Schedule execution dynamics at two points in time (based on Ivanov et al. 2016c)

The multidimensionality and the combinatorial explosion of the problem results in decreasing connectivity under the network diagram of operations. Consider four machines and six jobs $\overline{B}^{(i)}$ each of which is composed of 3–6 operations $D_\mu^{(i)}$. At any time instant, only one operation can be processed on one machine. Different colors describe current execution states of operations. The operations marked in black have already been completed. The operations marked in gray may be executed subject to machine availability and precedence relations. The operations marked in white cannot be executed yet because of the precedence relations. For example, at $t = t_2$, the operation $D_2^{(4)}$ cannot be assigned since the operation $D_1^{(4)}$ is still being processed with the use of machine $M^{(4)} (u_{12j}^{(o)}(t_2) = 1)$. From Figure 3, it can be

observed that at any time instant, the assignment decisions consider only the gray colored operations subject to some available ("competing") machines, i.e., the large-scale multi-dimensional combinatorial matrix is decomposed. The assignment of a machine $M^{(j)}$ to the execution of the operation $D_\mu^{(i)}$ can be described by the piecewise continuous function $u_{i\mu j}^{(o)}(t)$ that becomes equal to 1 in the case of an assignment. It can be observed from Figure 3 that the current dimensionality of the considered scheduling problem is determined by the dimensionality of the gray colored area. The operations in the black and white areas are not being considered at the given points of time , and, therefore, will not influence the assignment matrix size. This is the principal advantage of the proposed dynamic decomposition as compared to mathematical programming and combinatorial optimization theory (Kalinin and Sokolov 1985, 1987, Sokolov and Yusupov 2002, Ivanov and Sokolov 2010, 2012, Ivanov et al. 2012).

Problem description

The task consists of sequencing jobs and scheduling operations with precedence relations on machines subject to a multi-criteria objective function (makespan, due dates, lead time, and throughput) with limited machine capacity and non-stationary capacity availability considerations (Ivanov et al. 2016c).

Consider the problem formalization. Note that some of the notations have been defined in Sect. 2. Additional notations are defined as follows.

Set indexes

j is the machine index,
μ is the operation index (i.e., number of the operation in the job),
o is the index of parameters and variables in the model of operation control,
k is the index of parameters and variables in the model of machine control, and
f is the index of parameters and variables in the model of flow control.
For further consideration and a more convenient comparison of approaches from different articles, we assume $t \in (t_0, t_f] = (T_0, T_f]$

Parameters

$a_{i\mu}^{(o)}$ is the planned processing volume of $D_\mu^{(i)}$ operation,

$\tilde{\tilde{R}}_j^{(f)}$ is the total $M^{(j)}$ machine capacity,

$c_{i\mu j}^{(f)}$ is the processing intensity of $D_\mu^{(i)}$ execution on $M^{(j)}$ machine,

$\xi_1(t) \in [0,1]$, $\xi_2(t) \in [0,1]$ are the vectors of perturbation impacts on the notes and links,
$\alpha(\tau)$ is the penalty function in the mathematical model of the operation control processes,
$\beta(\tau)$ is the penalty function in the mathematical model for the flow control,
$\mathbf{q}^{(1)}$ and $\mathbf{q}^{(2)}$ are vector-functions, defining the main spatio-temporal, economic, technical, and technological conditions for the operation processing process, and
λ are the weight coefficients of the performance indicators in the multi-criteria objective function.

Continuous decision variables

$x_{i\mu}^{(o)}$ is a state variable characterizing the state of an operation $D_\mu^{(i)}$,

$\varepsilon(t)$ is the given preset matrix time function of time-spatial constraints,

$x_j^{(k)}$ is a state variable characterizing the total employment time of machine $M^{(j)}$,

$x_{i\mu j}^{(f)}$ is a state variable characterizing the processed flow volume,

$u_{i\mu j}^{(f)}(t)$ is a control variable that is equal to the processed flow volume $x_{i\mu j}^{(f)}$ at any point of time t,

$\mathbf{u}(t)$ is a feasible schedule, and

$\mathbf{u}*(t)$ is the optimal schedule.

Binary decision variable

$u_{i\mu j}^{(o)}(t) \in \{0,1\}$ is the assignment decision control action at time t.

The impact of the processing intensity $c_{i\mu j}^{f}(t)$ is that the machine $M^{(j)}$ can process $a_{i\mu j}$ units subject to the planned processing volume $a_{i\mu}^{(o)}$ and $c_{i\mu j}^{(f)}(t)$. An operation $D_{\mu}^{(i)}$ may start only after the previous operation $D_{\alpha}^{(i)}$ has been completed. The *problem* consists of scheduling the operations while taking into account flow dynamics control subject to three *objectives*: J_1 – minimization of total lateness (subject to t_f), J_2 – throughput maximization subject to $a_{i\mu}^{(o)}$ and $a_{i\mu j}^{(f)}$; i.e., in the ideal case $x_{i\mu}^{(o)}(t_f) = a_{i\mu}^{(o)}$, $x_{i\mu j}^{(f)}(t_f) = a_{i\mu j}^{(f)}$ for all jobs subject to $c_{i\mu j}(t)$ and $\varepsilon_{ij}(t)$; and J_3 – equal utilization of the machines (subject to $\widetilde{\widetilde{R}}_{j}^{(f)}$).

Process control model

The simplified form of *processing dynamics of the operation* $D_{\mu}^{(i)}$ can be expressed as follows in Eq. (30)-(32):

$$\frac{dx_{i\mu}^{(o)}}{dt} = \dot{x}_{i\mu}^{(o)} = \sum_{j=1}^{m} \varepsilon_{ij}(t) u_{i\mu j}^{(o)}(t) \tag{30}$$

$$\frac{dx_{j}^{(k)}}{dt} = \dot{x}_{j}^{(k)} = \sum_{i=1}^{n} \sum_{\mu=1}^{s_i} u_{i\mu j}^{(o)}(t) \tag{31}$$

$$\frac{dx_{i\mu j}^{(f)}}{dt} = \dot{x}_{i\mu j}^{(f)} = u_{i\mu j}^{(f)} \tag{32}$$

For simplification, we introduce only three equations of process dynamics control. The overall modelling complex also contains additional models such as setup and material delivery models (Ivanov 2016b). Eq. (30) represents the progress of operation execution at a machine whereas $\varepsilon_{ij}(t) = 1$, if machine $M^{(j)}$ is available, and $\varepsilon_{ij}(t) = 0$. $u_{i\mu j}^{(o)}(t)$ is a decision variable. $u_{i\mu j}^{(o)}(t) = 1$ at the any point of time t, if the operation $D_{\mu}^{(i)}$ is assigned to the machine $M^{(j)}$, and $u_{i\mu j}^{(o)}(t) = 0$ otherwise. Equation (31) represents machine utilization dynamics. The variable $x_{j}^{(k)}$ characterizes the total employment time of machine $M^{(j)}$. Eq. (32) corresponds to Eq. (30) subject to the control variable $u_{i\mu j}^{(o)}(t)$. $u_{i\mu j}^{(o)}(t) = 1$ at each t, if the operation $D_{\mu}^{(i)}$ is assigned to machine $M^{(j)}$ and $u_{i\mu j}^{(o)}(t) = 0$ otherwise (cf Fig. 2).

Recall that the task times may differ depending upon different speeds $c_{i\mu j}^{(f)}(t)$ and machine availabilities $\varepsilon_{ij}(t)$. For this reason, the assignments from the model (30) (made on the basis of the volumes $a_{i\mu}$) are now subject to further optimization regarding flow dynamics control.

The assignment of an operation to a machine and the starting execution of operations causes dynamic flows of processed products. The economic sense of Eq. (32) consists of the dynamic representation of the material flows resulting from the execution of the operations on machine $M^{(j)}$. The meaning of Eq. (32) is quite similar to a system dynamics model for balancing the flows in a system. The proposed ap-

proach also considers the strictly defined logic of the operation execution. Moreover, the models of operations and flow control are interlinked linearly by precedence constraints and the adjoint system. In contrast to model (30), the control variable $u_{i\mu j}^{(f)}(t)$ is not a binary variable, but is equal to the processed flow volume $x_{i\mu j}^{(f)}$ at each t.

Constraints

$$\sum_{j=1}^{m} u_{i\mu j}^{(o)} [\sum_{\overline{\alpha} \in \Gamma_{i\mu_1}^-} (a_{i\overline{\alpha}}^{(o)} - x_{i\overline{\alpha}}^{(o)}) + \prod_{\overline{\beta} \in \Gamma_{i\mu_2}^-} (a_{i\overline{\beta}}^{(o)} - x_{i\overline{\beta}}^{(o)})] = 0, \qquad (33)$$

$$0 \le u_{i\mu j}^{(f)}(t) \le c_{i\mu j}^{(f)} \cdot u_{i\mu j}^{(o)} \cdot \xi_1^{(f)}(t), \qquad (34)$$

$$\sum_{i=1}^{n} \sum_{\mu=1}^{s_i} u_{i\mu j}^{(f)}(t) \le \widetilde{\widetilde{R}}_j^{(f)} \cdot \xi_2^{(f)}(t), \qquad (35)$$

Constraints (33) depict the processing logic of operations in the jobs and determine precedence relations by blocking the operation $D_\mu^{(i)}$ until the previous operations $D_{\overline{\alpha}}^{(i)}$, $D_{\overline{\beta}}^{(i)}$ have been completed. Constraints (34) and (35) are capacity constraints in terms of processing intensity and maximum capacity, respectively.

The flow control model (32) uses the assignment results from the operations control model (30) in the form of the control variables $u_{i\mu j}^{(o)}(t)$ and extends them by the actual processing speed of the machines subject to the constraints (34) and (35). Inequalities (34) use the assignment decisions ($u_{i\mu j}^{(o)}(t)$) from the model (30) and the processing speed $c_{i\mu j}^{(f)}(t)$ of the machines $M^{(j)}$. Constraints (35) reflect that the processing speed is constrained by $\widetilde{\widetilde{R}}_j^{(f)}$, taking into account the lower and upper bounds of some perturbation impacts $0 \le \xi_2^{(f)}(t) \le 1$ which may decrease capacity availability.

Objective function

$$J_1^{(o)} = \frac{1}{2} \sum_{i=1}^{\overline{n}} \sum_{\mu=1}^{s_i} (a_{i\mu}^{(o)} - x_{i\mu}^{(o)}(t_f))^2, \qquad (36)$$

$$J_2^{(o)} = \sum_{i=1}^{\overline{n}} \sum_{\mu=1}^{s_i} \sum_{j=1}^{n} \int_{T_0}^{t_f} \alpha_{i\mu}^{(o)}(\tau) u_{i\mu j}^{(o)}(\tau) d\tau. \qquad (37)$$

$$J_1^{(k)} = \frac{1}{2} \sum_{j=1}^{n} (t_f - x_j^{(k)}(t_f))^2. \qquad (38)$$

$$J_1^{(f)} = \frac{1}{2} \sum_{i=1}^{n} \sum_{\mu=1}^{s_i} \sum_{j=1}^{m} (a_{i\mu j}^{(f)} - x_{i\mu}^{(f)}(t_f))^2, \qquad (39)$$

$$J_2^{(f)} = \frac{1}{2} \sum_{i=1}^{n} \sum_{\mu=1}^{s_i} \sum_{j=1}^{m} \int_{T_0}^{t_f} \beta_{i\mu}^{(f)}(\tau) u_{i\mu j}^{(f)}(\tau) d\tau. \qquad (40)$$

A good model should provide decision makers with alternatives. The performance indicators may be weighted differently depending on decision-maker's preferences. The preference relations (minmax,

maxmin, etc.) form the Pareto space and allow the calculation of a general relative quality index (31) within the corresponding schedule $\mathbf{u}(t)$.

$$\mathbf{J}(\mathbf{x}(t), \mathbf{u}(t), \xi(t), t) = \left\| (J_1^{(o)}, J_2^{(o)}, J_1^{(k)}, J_1^{(f)}, J_2^{(f)}) \right\|^{\mathrm{T}}, \tag{41}$$

The transition from the vector form $\mathbf{J}$ to a scalar form J_G can be performed on the basis of the weight coefficients $\lambda_1^{(o)}, \lambda_2^{(o)}, \lambda_3^{(k)}, \lambda_4^{(f)}, \lambda_5^{(f)}$.

The performance indicator $J_1^{(o)}$ (36) characterizes the accuracy of the end condition accomplishment, i.e., the volume of the completed operations by the time t_f. Maximization of (36) is throughput maximization. Minimization of the function (37) relates to total maximum lateness using penalties and depicts due date objective. The indicator $J_1^{(k)}$ (38) helps to estimate the uniformity of the machine capacity utilization at the end point $t = t_f$ of the planning period. The economic meaning of the objectives (39) and (40) is identical to the objectives (36) and (37).

The scheduling problem can now be formulated as the following optimal program control problem. This is necessary to find an allowable control $\mathbf{u}(t)$, $t \in (t_0, t_f]$ that ensures the model (30)-(32) meets the requirements $\mathbf{q}^{(1)}(\mathbf{x}, \mathbf{u}) = \mathbf{0}$, $\mathbf{q}^{(2)}(\mathbf{x}, \mathbf{u}) \le \mathbf{0}$ (33)–(34) and guides the dynamic system $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}, t)$ from the initial state $\mathbf{h}_0$ to the specified final state $\mathbf{h}_1$. If there are several allowable controls (schedules), then the best one (optimal) should be selected to maximize (minimize) J_G.

This scheduling model is a linear, non-stationary, finite-dimensional controlled differential system with a convex area of admissible control and a reconfigurable structure that allows the use of Boolean assignment control variables. According to optimal control theorems (Lee and Markus 1967, Moiseev 1974), the optimal control exists and can be calculated (Pontryagin *et al.* 1964) for the model class considered.

The main idea of the model is to implement non-linear technological constraints in the sets of allowable control inputs rather than in the right parts of differential equations. In this case, Lagrange coefficients, keeping the information about economic and technological constraints, are defined via the local-sections method (Pontryagin et al. 1964). The recurrence description of models allows parallel computations, accelerated problem solving. Furthermore, the model proposed use of interval constraints instead of relay ones. Nevertheless, the control inputs take on Boolean values as caused by the linearity of differential equations and the convexity of the set of alternatives.

The first model feature is that the right parts of the differential equations undergo discontinuity at the beginning of interaction zones. The considered problems can be regarded as control problems with intermediate conditions. The second feature is the multi-criteria nature of the problems. The third feature is concerned with the influence of uncertainty factors. The fourth feature is the form of time-spatial, technical, and technological non-linear conditions that are mainly considered in control constraints and boundary conditions.

3.3 Model with hybrid structure-terminal-logical constraints: applications to customized assembly systems and Industry 4.0

Problem description

Industry 4.0 principles give rise to new requirements for scheduling techniques. In classical manufacturing systems, scheduling is performed for a predefined production system and process design. Since both

production system and the process design become dynamic in Industry 4.0, a new two-level problem of simultaneous process design and job scheduling arises. Traditionally, these two tasks were solved separately. Let us illustrate how the optimal control models with hybrid *structure-terminal-logical constraints* allow the handling of scheduling tasks in Industry 4.0 systems.

Consider a customized, reconfigurable assembly system that is controlled on the basis of Industry 4.0 principles. This means that manufacturing processes for different customer orders may have individual process structures, whereby the flexible stations are able to execute different process steps subject to individual sets of operations within the jobs. Therefore, a problem of simultaneous, structural-functional synthesis of the customized assembly system arises. The task consists of process structure synthesis, sequencing jobs and scheduling operations with precedence relations on machines subject to a multi-criteria objective function (makespan, due dates, lead time, and throughput) with limited machine capacity and non-stationary capacity availability considerations (Ivanov et al. 2016b,c, Ivanov et al. 2017b). Consider a simplified example of a system with *hybrid structural-terminal-logical constraints* in Fig. 4.
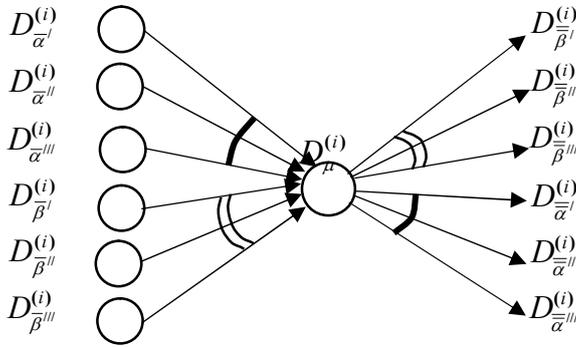

Fig. 4. Precedence relations of operation $D_\mu^{(i)}$

Let us illustrate a general logical-dynamic mathematical model of production process structure and process control. This model implicitly allows the description of all possible interconnections of the production structures and processes on the basis of operation execution control (Okhtilev et el. 2006).

Operation $D_\mu^{(i)}$ follows the operations $D_{\overline{\alpha}'}^{(i)}$ $D_{\overline{\alpha}''}^{(i)}$ $D_{\overline{\alpha}'''}^{(i)}$ according to the "and" precedence rule and operations $D_{\overline{\beta}'}^{(i)}$ $D_{\overline{\beta}''}^{(i)}$ $D_{\overline{\beta}'''}^{(i)}$ according to the precedence rule "or". Analogously, six operations follow the operation $D_\mu^{(i)}$ according to either "and" or "or" rules. For the considered system, the following modified optimal control model can be presented (Eqs. 42-44).

$$\dot{x}_{i\mu}^{(o)} = \sum_{j=1}^m \varepsilon_{ij}(t)\Theta_{i\mu j}(t)(u_{i\mu j}^{(o)}(t) + w_{i\mu j}^{(o)}(t)) \tag{42}$$

$$\dot{x}_{i\mu}^{(f)} = \sum_{j=1}^m u_{i\mu j}^{(f)}(t) \tag{43}$$

$$\dot{x}_j^{(k)} = \sum_{i=1}^n \sum_{\mu=1}^{s_i} u_{i\mu j}^{(o)} \tag{44}$$

The interpretation of this model is similar to the model (30)-(32). The constraint system (33)-(35) can now be modified as shown in Eqs. (45)-(53).

$$\sum_{j=1}^m u_{i\mu j}^{(f)}(t) \le 1; \forall i, \forall \mu ; \quad \sum_{i=1}^n \sum_{\mu=1}^{s_i} u_{i\mu j}^{(o)}(t) \le 1 \cdot \xi_j^{(1)}; \forall j \tag{45}$$

$$\sum_{j=1}^{m} u_{i\overline{\alpha}j}^{(o)} \cdot \sum_{\mu \in \Gamma_{i\overline{\alpha}}^{+}} x_{i\mu}^{(o)} = 0 \tag{46}$$

$$\sum_{j=1}^{m} u_{i\overline{\beta}j}^{(o)} \cdot \prod_{\mu \in \Gamma_{i\overline{\alpha}}^{+}} x_{i\mu}^{(o)} = 0 \tag{47}$$

$$\sum_{j=1}^{m} u_{i\overline{\overline{\alpha}}j}^{(o)} \cdot \sum_{\mu \in \Gamma_{i\overline{\overline{\alpha}}}^{-}} (a_{i\mu}^{(f)} - x_{i\mu}^{(f)}) = 0 \tag{48}$$

$$\sum_{j=1}^{m} u_{i\overline{\overline{\beta}}j}^{(o)} \cdot \prod_{\mu \in \Gamma_{i\overline{\overline{\beta}}}^{-}} (a_{i\mu}^{(f)} - x_{i\mu}^{(o)}) = 0 \tag{49}$$

$$\sum_{j=1}^{m} w_{i\mu j}^{(o)} [a_{i\mu}^{(f)} - x_{i\mu}^{(f)}] = 0 \tag{50}$$

$$0 \le \xi_j^{(1)}(t) \le 1 \tag{51}$$

$$0 \le \xi_j^{(2)}(t) \le 1 \tag{52}$$

$$u_{i\mu j}^{(o)}(t), \, w_{i\mu j}^{(o)}(t) \in \{0,1\} \tag{53}$$

Constraints (46) and (47) define precedence relations for operation $D_\mu^{(i)}$ with regards to the predecessor operations $D_{\overline{\alpha}}^{(i)}$, and $D_{\overline{\beta}}^{(i)}$. Constraints (48) and (49) define precedence relations for operation $D_\mu^{(i)}$ with regards to the following operations $D_{\overline{\overline{\alpha}}}^{(i)}$, and $D_{\overline{\overline{\beta}}}^{(i)}$. Constraint (50) defines the logic for the auxiliary control variable $w_{i\mu j}^{(o)} \in \{0,1\}$ which equals 1 if $x_{i\mu}^{(f)}(t) = a_{i\mu}^{(f)}$ at timet t and $x_{i\mu}^{(o)} \ne a_{i\mu}^{(o)}$. In other words, the flow is interrupted. To compensate for this, the auxiliary control $w_{i\mu j}^{(o)}$ is introduced in Eq. (42) that differentiates it from Eq. (30). Constraint (48) is used to avoid overproduction regarding the operation $D_\mu^{(i)}$, i.e., $x_{i\mu}^{(o)}(t_f) = a_{i\mu}^{(o)}$, which means that $x_{i\mu}^{(o)}(t_f)$ cannot exceed $a_{i\mu}^{(o)}$. In order to assess schedule robustness (e.g., using attainable sets; see Section 4), we introduce constraints on perturbations (52) and (53).

End conditions can be written as shown in Eqs (54) and (55).

$$t = t_o: \ x_{i\mu}^{(o)}(t_o) = x_{i\mu}^{(f)}(t_o) = x_j^{(k)}(t_o) = u_{i\mu j}(t_o) = 0 \tag{54}$$

$$t = t_f: \ x_{i\mu}^{(o)}(t_f) = a_{i\mu}^{(o)}; \quad x_{i\mu}^{(f)}(t_f) = a_{i\mu}^{(f)}; \ x_i^{(f)}(t_f), \in \mathfrak{R}^1, \ \mathfrak{R}^1 \in [0,...,\infty) \tag{55}$$

In Section 4, we describe the method of attainable sets for schedule robustness. In Section 5, computational procedures for the models from Section 3 will be presented.

4 Attainable (reachable) sets as a method for qualitative analysis of optimal control performance.

An advantage of optimal control application to scheduling is the possibility of attracting a rich variety of qualitative performance analysis methods. For example, robustness and resilience objectives can be integrated as non-stationary performance indicators in scheduling decisions. (Ivanov et al. 2016 a,b) proposed application of the method of attainable sets to calculate the robustness index for schedules and to obtain the attainable sets for interval data with no a priori information about perturbation impacts, i.e., for the most severe case of non-stationary perturbations.

The attainable set (also known as reachable sets) approach determines a range of operating policies (the union of which is called an attainable set) during the scheduling stage over which the system's current performance can be guaranteed to meet certain targets, i.e., output (Ivanov and Sokolov, 2010). The attainable set characterizes all possible states of a schedule subject to different variations of control parameters (e.g., machine capacity availability).

The attainable set is calculated from the main optimal program control vector. Here, the impacts of perturbations play the role of control variables. In varying these perturbations at each instant of time over the schedule within the time interval and setting these variations into the initial differential system, a set of points where the schedule can be steered to is generated. In other words, a set of alternative optimal program control vectors is generated through admissible variations of perturbations and forms herewith the attainable set of a schedule subject to disturbances. With the help of attainable sets, it becomes possible to create a dynamic projection of schedule execution for different uncertainty scenarios. Therefore, attainable sets can be used to calculate a corresponding robustness metric for schedules. This metric can be used for ranging alternative schedules subject to the individual risk perceptions of decision-makers.

The application of the attainable set method to schedule robustness analysis can be found in studies by Ivanov et al. (2016a,b). In particular, Eqs. (34) and (35) (cf Sect. 3) set up the attainable set of the optimal program in a dynamic system, i.e., all possible states of schedule execution subject to different variations of the parameters (e.g., the capacity availability). The introduction of Eqs (52) and (63) allows analysis of feasible schedule executions under conditions of non-stationary perturbations. If so, an attainable set can be used to analyse schedule robustness i.e., the ability to continue schedule execution according to the specified objectives despite the presence of perturbations.

5 Computational algorithms

Analytical methods for optimal control have been proved for small-dimensional systems only (Moiseev 1974, Afanasiev et al. 1996, Maimon et al. 1998, Thompson and Sethi 2000, Kogan and Khmelnitsky 2000). In control engineering practice, numerical methods have been applied. A methodological challenge in applying the maximum principle is to find the coefficients of the adjoint system which change over time. Another methodical challenge of boundary problems is that the initial conditions for the adjoint variables $\psi(t_0)$ are not given. At the same time, optimal program control should be calculated subject to the end conditions.

Numerical methods can be classified into several groups. General and specialized methods and algorithms can be distinguished. General methods can be classified into three groups: the state space (so called direct methods, e.g., gradient methods), control space (so called indirect methods based on control variations such as the method of successive approximations), and trajectory space (e.g., dynamic programming method) methods. The dynamic programming method was applied to production scheduling through optimal control in the work by Giglio (2015). Specialized methods are valid for special control system classes such as linear systems where methods and algorithms for quadratic linear problems are applied (Ivanov and Sokolov 2012) or when the optimal control problem is presented in terms of mathematical programming (Tabak and Kuo 1971, Ivanov et al. 2017a).

The first group of general methods (i.e., system space methods) is based on optimal control problem presentation as a two-point boundary problem using the maximum principle and based on the necessary conditions of optimal control. Two systems of differential equations are solved: the main system and the conjugate one. This provides the optimal program control vector $\mathbf{u}^*(t)$ and the state trajectory $\mathbf{x}^*(t)$.

The vector $\mathbf{u}^*(t)$ at time $t = T_0$ under the conditions $\mathbf{h}_0(\mathbf{x}(T_0)) \leq \mathbf{O}$ and for the given value of $\boldsymbol{\psi}(T_0)$ returns the maximum to the Hamiltonian function (10). The missing initial conditions of the adjoint system are computed based on different algorithms.

In the second group and regarding the constraint type from Sect. 3.1 (i.e., control space methods), Maimon et al. (1998) analysed time-decomposition methods based on multi-iteration control variations. These authors along with Ivanov and Sokolov (2010) pointed out the importance of approximation accuracy and the quality of the initial nominal heuristic solution. Ivanov et al. (2016c) proved that computational complexity of the method of successive approximations in one iteration is polynomial. Regarding the considered scheduling problem, the complexity of the proposed algorithm in one iteration is determined by the complexity of an integer programming assignment problem.

Let us limit detailed analysis of computational procedures to two-point boundary problems with fixed ends of the state trajectory $\mathbf{x}(t)$ and a fixed time interval $t \in (t_0, t_f] = (T_0, T_f]$ (Lee and Markus 1967, Moiseev 1974, Ivanov and Sokolov 2010, Krylov & Chernousko 1972, Chernousko & Lyubushin 1982). Newton's method and its modifications allow for simple realization (there is no need to integrate the conjugate system), a fast convergence (if the initial choice of control is good), and high accuracy of the solution. The main disadvantage is the dependency of a convergence upon the choice of $\boldsymbol{\psi}(T_0)$. In the case of absence of a good heuristic plan, these methods can be divergent. The method of penalty functionals is rather simple, but it does not provide an exact solution. Therefore, it is advisable to combine it with other methods. The main advantage of these algorithms over the classical gradient algorithms is that it has a simpler calculation for the direction vector at all iterations. However, this results in a slower convergence (sometimes in divergence). The convergence of all the gradient methods depends upon the initial approximation $\boldsymbol{\psi}_{(0)}(T_0)$.

Ivanov and Sokolov (2012) applied the Krylov-Chernousko method (1972). For production flow scheduling, Kogan and Khmelnitsky (1996), Khmelnitsky et al. (1997) and Maimon et al. (1998) related space methods as shooting methods. The authors, along with Kalinin and Sokolov (1985, 1987) and Ivanov and Sokolov (2012), noted that the methods of this group can be efficiently applied to linear main and adjoint systems whereby non-linearity should be transferred to the constraints. Ivanov et al. (2016c) applied a method of successive approximations with a free right end that can be used in combination with the branch & bound method.

Regarding the constraints from Sect. 3.2, Ivanov and Sokolov (2010), Ivanov and Sokolov (2012), Ivanov et al. (2013 and Ivanov et al. (2016a,c) applied a combination of optimal control and mathematical programming. Optimal control is not used for solving the combinatorial problem, but rather for enhancing the existing mathematical programming algorithms regarding non-stationarity, flow control, and continuous material flows. Since the control variables are presented as binary variables, methods of discrete optimization are applied to combinatorial tasks within certain time intervals.

The basic *computational idea* of this approach is that operation execution and machine availability are dynamically distributed in time over the planning horizon. As such, not all operations and machines are involved in decision making at the same time (cf Fig. 3). Therefore, the solutions at each point of time for small dimensionalities of assignment and flow distribution problems are calculated with the help of mathematical programming.

For a more general case in Sect. 3.3 (Eqs. 42-44), the following Hamiltonian can be written (Eqs. 56-58).

$$H(\overset{p}{x}(t), \overset{p}{u}(t), \overset{p}{\psi}(t)) = H_1 + H_2 + H_3 + H_4 + H_5 + H_6 + H_7, \text{ where}$$

$$H_1 = \sum_i \sum_j \sum_\mu [\psi_{i\mu}^{(o)} \varepsilon_{ij} \Theta_{i\mu j} + \psi_j^{(k)} + \varphi_{i\mu j}^{(1)} + \alpha_{i\mu} - \beta_{i\mu}] \cdot u_{i\mu j}^{(o)}, \tag{56}$$

$$H_2 = \sum_i \sum_\mu \sum_j [\psi_{i\mu}^{(f)}] \cdot u_{i\mu j}^{(f)} \tag{57}$$

$$H_3 = \sum_i \sum_\mu \sum_j [\varphi_{i\mu j}^{(1)} + \psi_{i\mu}^{(o)} \cdot \varepsilon_{ij} \Theta_{i\mu j}] \cdot w_{i\mu j}^{(o)} \tag{58}$$

Transversality conditions for the case from Sect. 3.3 are written as Eqs (59)-(61):

$$\psi_{i\mu}^{(o)}(t_f) = -(a_{i\mu}^{(o)} - x_{i\mu}^{(o)})\big|_{t=t_f} \tag{59}$$

$$\psi_{i\mu}^{(f)}(t_f) = (a_{i\mu}^{(f)} - x_{i\mu}^{(f)})\big|_{t=t_f} \tag{60}$$

$$\psi_j^{(k)}(t_f) = (\tilde{t}_f - x_j^{(k)}(t_f)) \tag{61}$$

Let us consider the algorithmic realization of the above-described modified maximum principle. We consider a general algorithm for the general model (42)-(55) and the model (30)-(41). We refer to the general model (42)-(55) equations as ( ) and to the model (30)-(41) equations as [ ]. Note that the constraints (45)-(53) and [(21)–(23)] are identical to those in the mathematical programming models. However, at each point of time t, the number of variables in the calculation procedure is determined by the operations, which are currently in the "active zone" of scheduling, i.e., the operations marked in grey in Figure 3. For the problem sizes subject to the "active zone", there are known methods for the solution of the mathematical programming.

After transforming into a boundary problem, a relaxed problem can be solved to receive an optimal program control, for the computation of which the main and adjoint systems are integrated, i.e., the optimal program control vector $\mathbf{u}^*(t)$ and the state trajectory $\mathbf{x}^*(t)$ are obtained. The optimal program control vector at time $t = t_{00}$ and for the given value of $\psi(t)$ should return the maximum to (56)–(58), which have been transformed to a general performance index and are expressed in scalar form J_G (Eq. 41). To obtain the adjoint system vector, the Krylov–Chernousko method of successive approximations for an optimal program control problem with a free right end which is based on the joint use of a modified successive approximation method (Krylov & Chernousko, 1972) has been used.

Step 1 An initial solution $\overline{\mathbf{u}}(t)$, $t \in (t_0, t_f]$ (a feasible control, in other words, a feasible schedule) is selected and $r=0$.

Step 2 As a result of the dynamic model run, $\mathbf{x}^{(r)}(t)$ is received. Besides, if $t = t_f$ then the record value $J_G = J_G^{(r)}$ can be calculated. Then, the transversality conditions (59)–(61) are evaluated.

Step 3 The adjoint system is integrated subject to $\mathbf{u}(t) = \overline{\mathbf{u}}(t)$ and over the interval from $t = t_f$ to $t = t_0$. For the time $t = t_0$, the first approximation $\psi_i^{(r)}(t_0)$ is obtained as a result. Here, the iteration number $r=0$ is completed.

Step 4 From the time point $t = t_0$ onwards, the control $\mathbf{u}^{(r+1)}(t)$ is determined ($r = 0, 1, 2, ...$ denotes the number of the iteration). In parallel with the maximization of the Hamiltonian, the main system of equations and the conjugate one are integrated. The maximization involves the solution of several mathematical programming problems at each point of time.

Note that the method of successive approximations in its initial form has not guaranteed convergence. Specific modifications are needed. One such example is shown in the study by Ivanov et al. (2016c). Finally, we note the problem with operation preemptions while applying the algorithm described above. Since operations have different processing durations, the machines become available for the processing of the next operations at different points in time. This can result in a situation in which Hamiltonian maximization can lead to an improvement in the performance functional if operation processing at one machine is interrupted and the operation is reassigned to the new machine that became available. For some practical problem, it can be possible through technology. In other settings, such preemptions can be prohibited. The study by Ivanov et al. (2013) showed how to resolve this problem at the modelling and the algorithmic levels. The monotony and convergence of the modified successive approximations method for problems with the non-preemption condition has been previously proved for two-point linear boundary problems with convex control areas and goal function (Lybushin 1979, Kalinin and Sokolov 1987).

6 Contributions, application areas, and limitations

The performed analysis allows classification of optimal control models with applications to scheduling in the following three classes (Fig. 5).
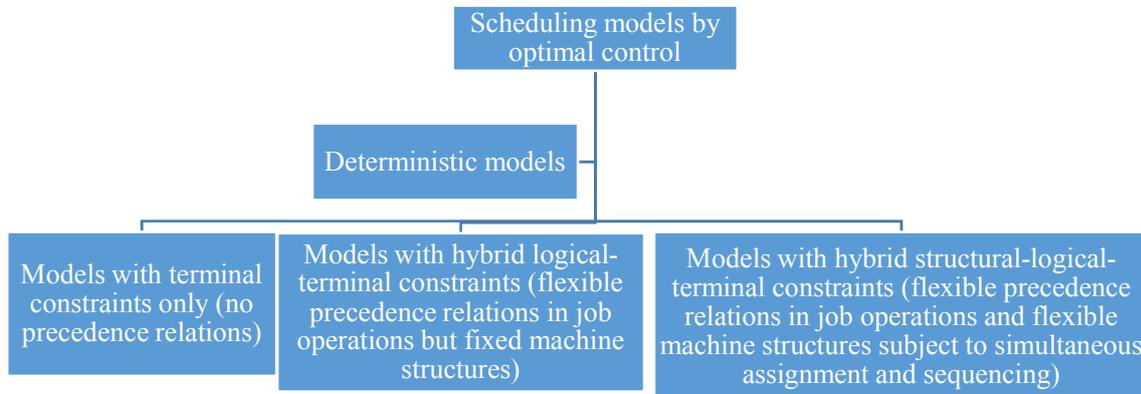

Fig. 5. Classification of deterministic optimal control scheduling models

Models with terminal constraints only (i.e., no precedence relations in jobs) are frequently applied to master production scheduling and flexible manufacturing system domains. Models with hybrid logical-terminal constraints (i.e., with flexible precedence relations in job operations, but fixed machine structures) have their place in job and flow shop scheduling. Models with hybrid structural-logical-terminal constraints are characterized by flexible precedence relations in job operations and flexible machine structures. Such models decide simultaneously on assignment and sequencing with applications to customized assembly systems, where two-stage scheduling procedures are applied containing process design and scheduling levels. An example of such systems is Industry 4.0 (Ivanov et al. 2016c).

Applications of optimal control theory to the scheduling of operational systems and supply chains are summarized in Table 1.

Table 1 - Applications of optimal control theory to scheduling

The main results of optimal control theory	Application to scheduling of operational systems and supply chains
Criteria for existence of a solution	Model verification for schedule control
Criteria for controllability and attainability	Control processes verification for a given time interval / Determination of the constraints on operational goals
Criteria for uniqueness of optimal program control	Analysis of possibility to obtain an optimal plan
Necessary and sufficient conditions of optimality	Preliminary analysis of optimal program controls; generation scheduling algorithms
The program control and feedback control	Planning, scheduling, and execution control models on a united methodological basis
Criteria for stability and sensitivity	Evaluation of schedule robustness, stability, and sensitivity for execution risks and alteration of input data

Advantages of optimal control methods can be considered regarding the treatment of specific scheduling problems with complex constraints, consideration of non-stationary process execution dynamics, representation in differential equations of complex interrelations between process execution, capacity evolution, and machine setups (Kinemia and Gershwin 1983, Kogan and Khmelnitsky 1996, Powell and Chen 1997, Maimon et al. 1998, Yang et al. 1999, Ivanov and Sokolov 2012, Song 2012, Ivanov et al. 2013, Giglio 2015, Ivanov et al. 2016a, Ye and Liu 2016). In addition, accuracy of continuous time and accurate presentation of continuous flows (e.g., in process industry or energy systems) with the help of continuous state variables belong to optimal control advantages. Finally, and probably most important, schedule presentation in terms of optimal control opens constructive ways to incorporate the rich variety of control theoretic axioms regarding feedback adaptive control (most applied in the framework of production-inventory control models, e.g., Dejonckheere et al. 2004, Ortega and Lin 2004, Wang et al. 2012, Spiegler et al. 2012, Spiegler et al. 2016, Dolgui et al. 2018), and use of control tools for qualitative performance analysis such as attainable (reachable) sets (Ivanov and Sokolov 2013b, Ivanov et al. 2016b,c).

Limitations of control applications should also be addressed. They comprise of a conceptual restriction for modelling mostly continuous flows and computational restrictions in terms of possible numerical instability, Heaviside step function jumps, nonexistence and non-convexity of gradients, gradient step, and quality of nominal control vector (Zimin and Ivanilov 1971, Sethi 1978, Afanasiev et al. 1996, Maimon et al. 1998, Sethi et al. 2002, Ivanov and Sokolov 2010, Ivanov et al. 2012). The algorithmic method selection strongly depends on the form of the process model and constraints. This is why the state-of-the art in optimal control mathematics does not provide a general framework for all application areas but rather serves to develop concrete solutions for concrete application cases using complexes of proved theorems and axioms.

7   Conclusions

Specific scheduling problems with complex structural, logical, and terminal constraints, process execution non-stationary (i.e., interruptions in machine availability) and complex interrelations between process dynamics, capacity evolution, and dynamic setups require further investigation and the attraction of a broad range of methodical approaches. One of these approaches can be optimal control. Optimal control strategies, as functions of the system and control state, allow generation of optimal decisions with consideration of system evolution in time in the presence of perturbations which results in different system states. In this paper, we provided a survey on the applications of optimal control to scheduling in production, supply chain, and Industry 4.0 systems with a focus on maximum principle-based studies. Optimal control approaches differ in perspective from mathematical programming methods representing schedules as trajectories.

In the paper, we considered optimal control models, performance analysis qualitative methods, and computational methods for optimal control. We provided a brief historic overview and clarified major mathematical fundamentals to bridge the control engineering terms and the industrial engineering and management terms. Optimal control models for job scheduling in production, supply chain and Industry 4.0 systems have been presented and compared regarding their analytical content and application areas. The survey grouped models with only terminal constraints with application to master production scheduling, models with hybrid terminal-logical constraints with applications to short term job and flow shop scheduling, and models with hybrid structural-terminal-logical constraints with applications to Industry 4.0. We also presented attainable (reachable) sets as a method for qualitative analysis of optimal control performance. Computational algorithms regarding state, control, and adjoint variable spaces have been analysed.

We point out that the advantages of optimal control methods belong treatment of scheduling problems, especially in process industry, with complex constraints, consideration of non-stationary process execution dynamics, representation in differential equations of complex interrelations between process execution, capacity evolution, and machine setups. In addition, accuracy of continuous time and accurate presentation of continuous flows (e.g., in process industry or energy systems) with the help of continuous state variables are also advantages of optimal control. Other applications of control theoretic models and algorithms to scheduling in shipyards, aerospace systems, logistics systems, and service systems can be found at http://litsam.ru.

It is important to note that scheduling presentation in terms of optimal control makes it possible to incorporate the rich variety of control theoretic axioms regarding feedback adaptive control (most applied in the framework of production-inventory control models) as well as to use control tools for qualitative performance analysis such as attainable (reachable) sets. Limitations of control applications are conceptual and algorithmic restrictions regarding continuous process applications and specific (i.e., non-generalized) forms of constructing algorithms with necessary requirements on optimality, convergence, and numerical stability.

Regarding future research and in light of the revealed methodical shortcomings and application limitations of optimal control methods using deterministic maximum principle, the following research avenues can be stated. First, concrete application cases need to be considered for which specific control models and algorithms will be developed. The construction of models and computational procedures within proved axioms of control theory is important. Second, application of qualitative performance analysis methods for control policy dynamic investigation under uncertainty, such as attainable sets, should be further explored. These tools might be helpful regarding analysis of production schedule ro-

bustness, supply chain resilience, and Industry 4.0 system flexibility. Third, computational methods themselves need to be further investigated and modified for concrete application. Fourth, the works on the stochastic maximum principle can be applied in stochastic scheduling environments. Fifth, new domains such as energy supply scheduling and information flow scheduling belong to optimal control application areas. A closer collaboration of control and industrial engineers is therefore important for future applications of control methods to operations and supply chain management.

ACKNOWLEDGMENT

References

Afanasiev V.N., Kolmanovskii V.B., Nosov V.R. Mathematical Theory of Control Systems Design. Dordrecht: Kluwer, 1996.

Albright S.C., Collins R.S. (1977). A Bayesian approach to the optimal control of continuous industrial processes. International Journal of Production Research, 15(1), 37-45.

Athaus, M., Falb, P.L. (1966). Optimal control: An introduction to the theory and its applications. New York, San Francisco, CA, Sidney, NSW: McGraw-Hill.

Bedini R., Toni P. (1980). The planning of a manufacturing system: a dynamic model. International Journal of Production Research, 18(2), 189-199.

Bellmann, R. (1972). Adaptive control processes: a guided tour. Princeton Univ. Press: Princeton, New Jersey.

Blazewicz, J., Ecker K., Pesch E., Schmidt G., Weglarz J. (2001). Scheduling Computer and Manufacturing Processes, 2nd Ed., Springer Verlag, Berlin.

Boltyanskiy, V. (1969). Mathematical methods of optimal control [In Russian]. Moscow: Nauka

Bożek, A., and M. Wysocki. 2015. Flexible Job Shop with Continuous Material Flow. International Journal of Production Research, 53(4), 1273–1290.

Bryson, A.E., Ho, Y.-C. (1975). Applied Optimal Control. Hemisphere, Washington.

Chernousko, F.L., Lyubushin, A.A. (1982). Method of successive approximations for solution of optimal control problems. Optimal Control Applications and Methods, 3(2), 101–114.

Choi, T.-M., Yeung, W.-K., Cheng, T.C.E. (2013). Scheduling and co-ordination of multi-suppliers single-warehouse-operator single-manufacturer supply chains with variable production rates and storage costs. International Journal of Production Research, 51(9), 2593–2601.

Dejonckheere, J., Disney, S.M., Lambrecht, M.R., Towill, D.R., (2004). The impact of information enrichment on the bullwhip effect in supply chains: A control engineering perspective. European Journal of Operational Research, 153(3), 727–750.

Disney, S.M, Towill, D.R, Warburton, R.D.H. (2006). On the equivalence of control theoretic, differential, and difference equation approaches to modeling supply chains. International Journal of Production Economics, 101, 194-208.

Dolgui, A., Proth, J.-M. (2010). Supply Chains Engineering: Useful Methods and Techniques. Springer.

Dolgui, A., Ivanov, D., Sokolov, B. (2018) Ripple Effect in the Supply Chain: An Analysis and Recent Literature. International Journal of Production Research, 56(1-2), published online.

Eilon S. (1961). Two inventory control policies. International Journal of Production Research, 1(1), 48-55.

Feng, Y. Y., H. Yan. 2000. Optimal production control in a discrete manufacturing system with unreliable machines and random demands. IEEE Transactions on Automatic Control, 35, 2280-2296.

Gershwin, S. B. (1994). Manufacturing Systems Engineering. PTR Prentice Hall, Englewood Cliffs, New Jersey.

Giglio D. (2015). Optimal control strategies for single-machine family scheduling with sequence-dependent batch setup and controllable processing times. Journal of Scheduling, 18(5), 525-543.

Harjunkoski, I., Maravelias C.T., Bongers P., Castro P.M., Engell, S., Grossmann, I.E., Hooker J., Méndez, C., Sand G., Wassick J. (2014). Scope for industrial applications of production scheduling models and solution methods. Computers and Chemical Engineering, 62, 161-193.

Hartl, R.F., S. P. Sethi (1984). Optimal control problems with differential inclusions: Sufficiency conditions and an application to a production inventory model. Optimal Control Applications and Methods 5, 289-307.

Hartl, RF., Sethi, SP., Vickson RG (1995). A survey of the maximum principles for optimal control problems with state constraints. SIAM Review, Vol. 37, No. 2, 181-218.

Hultquispt, F . (1988). Numerical Methods for Engineers and Computer Scientists, Menlo Park

Hwang, C.L., Fan, L.T., Erikson L.E. (1967). Optimum production planning by the maximum principle. Management Science, 13(9), 751–755.

Hwang, C.L., Fan, L.T., Tillman, F.A., Sharma, R. (1969). Optimal production planning and inventory control. International Journal of Production Research, 8(1), 75-83.

Ivanov D., Dolgui A., Sokolov B. (2016b). Robust dynamic schedule coordination control in the supply chain. Computers and Industrial Engineering, 94(1), 18-31.

Ivanov D., Sokolov B. (2013a) Dynamic coordinated scheduling in the supply chain under a process modernization, International Journal of Production Research. 51(9), 2680-2697.

Ivanov D., Sokolov B. (2013b) Control and system-theoretic identification of the supply chain dynamics domain for planning, analysis, and adaptation of performance under uncertainty, European Journal of Operational Research, 224(2), 313-323.

Ivanov D., Sokolov B., Dolgui, A. (2013) Multi-stage supply chain scheduling in petrochemistry with non-preemptive operations and execution control, International Journal of Production Research, 52(13), 4059-4077.

Ivanov, D., Dolgui A., Sokolov B., Werner F. (2016a). Schedule robustness analysis with the help of attainable sets in continuous flow problem under capacity disruptions, International Journal of Production Research, 54(1), 3397-3413.

Ivanov, D., Dolgui A., Sokolov B., Ivanova M. (2017a). Optimal control representation of the mathematical programming model for supply chain dynamic reconfiguration, Proceedings of IFAC 2017 World Congress.

Ivanov, D., Dolgui, A., Sokolov, B. (2017b). A dynamic approach to multi-stage job shop scheduling in an Industry 4.0-based flexible assembly system. In: Lödding, H., Riedel, R., Thoben, K.-D., Kiritsis, D., von Cieminski, G. (Eds.) Advances in Production Management Systems. The Path to Intelligent,

Collaborative and Sustainable Manufacturing. IFIP WG 5.7 International Conference, APMS 2017, Hamburg, Germany, September 3-7, 2017, Proceedings, Part I, pp. 475-482.

Ivanov, D., Sokolov, B., Dolgui, A., Werner, F., Ivanova, M. (2016c). A dynamic model and an algorithm for short-term supply chain scheduling in the smart factory Industry 4.0. International Journal of Production Research, 54(2), 386-402.

Ivanov, D., Sokolov B. (2012) Dynamic supply chain scheduling, Journal of Scheduling, 15(2), 201-216.

Ivanov, D., Sokolov B., Dolgui, A. (2012) Applicability of optimal control theory to adaptive supply chain planning and scheduling, Annual Reviews in Control, 36, 73-84.

Ivanov, D., Sokolov, B. (2010), Adaptive Supply Chain Management, Springer, London et al.

Jungwattanakit, J., Reodecha, M., Chaovalitwongse, P., Werner, F. (2009). A comparison of scheduling algorithms for flexible flow shop problems with unrelated parallel machines, setup times, and dual criteria. Computers and Operations Research, 36(2), 358–378.

Jiang, J., S. P. Sethi (1991). A state aggregation approach to manufacturing systems having machines states with weak and strong interactions. Operations Research, 39, 970-978.

Kalinin, V.N., Sokolov, B.V. (1985). Optimal planning of the process of interaction of moving operating objects. International Journal of Difference Equations, 21(5), 502-506.

Kalinin, V.N., Sokolov, B.V. (1987). A dynamic model and an optimal scheduling algorithm for activities with bans of interrupts. Automation and Remote Control, 48(1-2), 88-94.

Khmelnitsky E., Kogan K., Maimom O. (1997). Maximum principle-based methods for production scheduling with partially sequence-dependent setups. Int J Prod Res, Vol. 35, No. 10, 1997, pp. 2701-2712.

Khmelnitsky E, Presman E., Sethi SP (2011) Optimal production control of a failure-prone machine. Ann Oper Res 182:67-86

Kimemia, J.G., Gershwin, T.I. (1983). An algorithm for the computer control of a flexible manufacturing system. IIE Transactions, 15, 353-362.

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

Kogan K., Khmelnitsky E. (2000). Scheduling: control-based theory and polynomial-time algorithms. Dordrecht, Kluwer

Krylov, I.A., Chernousko, F.L. (1972). An algorithm for the method of successive approximations in optimal control problems. Zh. Vychisl. Mat. Mat. Fiz., 12(1), 14–34

Lauff, V., Werner, F. (2004) On the Complexity and Some Properties of Multi-Stage Scheduling Problems with Earliness and Tardiness Penalties. Computers and Operations Research, 31, 317-345.

Lee, E.B., Markus, L. (1967). Foundations of optimal control theory. Wiley, New York.

Lou, S.X.C., Van Ryzin, G. (1989). Optimal control rules for scheduling job shops. Annals of Operations Research, 17(1): 233-248.

Lou, S., S. P. Sethi, Q. Zhang (1994). Optimal feedback production planning in a stochastic two machine flowshop. European Journal of Operational Research{Special Issue on Stochastic Control Theory and Operational Research, 73, 331-345.

Lyubushin, A.A. (1979). Modifications and convergence of successive approximations for optimal control problems. Zh. Vychisl. Mat. Mat. Fiz., 19(6), 1414–1421.

Maimon, O., Khmelnitsky, Eugene, Kogan, K. (1998). Optimal Flow Control in Manufacturing Systems. Springer.

Moiseev, N.N. (1974). Element of the optimal systems theory [In Russian]. Moscow: Nauka.

Okhtilev, M.Yu., Sokolov, B.V., Yusupov, R.M. (2006). Intellectual Technologies for Monitoring and Control of Structure-Dynamics of Complex Technical Objects. Moscow, Nauka

Ortega, M. and Lin, L. (2004). Control theory applications to the production-inventory problem: a review. International Journal of Production Research. 42, 2303–2322.

Pinedo, M. (2008). Scheduling: Theory, Algorithms, and Systems. Springer, New York.

Pinha D., Ahluwalia R., Carvalho A. (2015). Parallel Mode Schedule Generation Scheme. IFAC-PapersOnLine Volume 48, Issue 3, 15th IFAC Symposium on Information Control Problems in Manufacturing INCOM 2015, Edited by Alexandre Dolgui et al.

Pontryagin, L.S., Boltyanskiy, V.G., Gamkrelidze, R.V., and Mishchenko, E.F. (1964). The mathematical theory of optimal processes. Pergamon Press, Oxford.

Powell, WB., Chen Z.L. (1997). An Optimal Control Formulation of Large-Scale Multiclass Machine Scheduling Problems. Vol. 450 of the series Lecture Notes in Economics and Mathematical Systems, 423-440.

Presman, E., S. P. Sethi, Q. Zhang (1995). Optimal feedback production planning in a stochastic N-machine flowshop. Automatica, 31, 1325-1332.

Presman, E. S. P. Sethi, W. Suo (1997). Optimal feedback production planning in a stochastic N-machine flowshop with limited buffers. Automatica,33, 1899-1903.

Presman, E., S. P. Sethi, H. Zhang, A. Bisi (2000). Optimality in two-machine flowshop with limited buffer. Annals of Operations Research, 98, 333-351.Samaratunga, C., S. P. Sethi, X. Y. Zhou (1997). Computational evaluation of hierarchical production control policies for stochastic manufacturing systems. Operations Research, 45, 258-274.

Sarimveis, H., Patrinos, P., Tarantilis, C.D., Kiranoudis, C.T. (2008). Dynamic modeling and control of supply chain systems: A review. Computers & Operations Research, 35, 3530–3561.

Sethi SP (1978) A Survey of Management Science Applications of the Deterministic Maximum Principle. TIMS Studies in the Management Science, 9:33-68.

Sethi SP (1984) Applications of the Maximum Principle to Production and Inventory Problems. Proceedings Third International Symposium On Inventories, Budapest, Aug. 27-31, 753-756.

Sethi, S.P. and Zhang, Q. (1994). Hierarchical Decision Making in Stochastic Manufacturing Systems, in series Systems and Control: Foundations and Applications, Birkhäuser Boston, Cambridge, MA.

Sethi, S. P., X. Y. Zhou (1996). Optimal feedback controls in deterministic dynamic two-machine flowshops. Operations Research Letters, 19, 225-235.

Sethi, S.P., Thompson, G.L. (2000). Optimal Control Theory: Applications to Management Science and Economics, 2nd Edition. Springer, Berlin.

Sethi SP, Yan H, Zhang H, Zhang Q. (2002) Optimal and Hierarchical Controls in Dynamic Sto-chastic Manufacturing Systems: A Survey. Manuf Serv Oper Mant, 4(2):133-170.

Sharifnia, A., M. Caramanis, S. B. Gershwin (1991). Dynamic setup scheduling and flow control in manufacturing systems. J. Discrete Event Dynamic Syst., 1, 149-175.

Soner, H.M. (1986). Optimal control with state space constraints II, SIAM J. Contr. Optim., 24, 1110-1122

Song DP (2012) Optimal Control and Optimization of Stochastic Supply Chain Systems, Springer

Sokolov B.V., Yusupov R.M. (2002) Complex simulation of automated control system of navigation spacecraft operation, Journal of Automation and Information Science 34(9-12), pp.19-30.

Sotskov, Y. N., Lai, T.-C., & Werner, F. (2013). Measures of Problem Uncertainty for Scheduling with Interval Processing Times. OR Spectrum, 35(3), 659–689.

Spiegler, V., Potter AT., Naim MM, Towil DR. (2016). The value of nonlinear control theory in investigating the underlying dynamics and resilience of a grocery supply chain. International Journal of Production Research 54(1):265-286

Spiegler, V., Naim, M. Wikner, J. (2012). A control engineering approach to the assessment of supply chain resilience. International Journal of Production Research 50:6162-6187.

Tabak, D., Kuo B.C. (1971). Optimal control by mathematical programming. Prentice Hall, NY.

Wang, X., Disney, S. M., Wang, J. 2012. Stability analysis of constrained inventory systems with transportation delay. European Journal of Operational Research 223(1), pp. 86-95.

Werner, F., Sotskov, Y. (Eds.) (2014) Sequencing and Scheduling with Inaccurate Data. Nova Publishers.

Yang J, Yan H, Sethi SP (1999). Optimal production planning in pull flow lines with multiple products. European Journal of Operational Research 119 (1999) 582-604

Ye, H., Liu, R. (2016) A multiphase optimal control method for multi-train control and scheduling on railway lines. Transportation Research Part B: Methodological, 93 (Part A), 377-393.

Zimin, I.N., Ivanilov, Yu.P. (1971). Solution of network planning problems by reducing them to optimal control problems. Zh. Vychisl. Mat. Mat. Fiz., 11(3), 632–641.