Computers and Chemical Engineering 37 (2012) 214–226

Contents lists available at SciVerse ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/compchemeng

Integrated production planning and scheduling optimization of multisite, multiproduct process industry

Nikisha K. Shah, Marianthi G. Ierapetritou *

Department of Chemical and Biochemical Engineering, Rutgers University, 98 Brett Road, Piscataway, NJ 08854, USA

ARTICLE INFO

Article history:
Received 10 May 2011
Accepted 30 August 2011
Available online 7 September 2011

Keywords:
Planning and scheduling integration
Multisite production
Continuous time formulation
Decomposition method
Augmented Lagrangian relaxation

ABSTRACT

The current manufacturing environment for process industry has changed from a traditional single-site, single market to a more integrated global production mode where multiple sites are serving a global market. In this paper, the integrated planning and scheduling problem for the multisite, multiproduct batch plants is considered. The major challenge for addressing this problem is that the corresponding optimization problem becomes computationally intractable as the number of production sites, markets, and products increases in the supply chain network. To effectively deal with the increasing complexity, the block angular structure of the constraints matrix is exploited by relaxing the inventory constraints between adjoining time periods using the augmented Lagrangian decomposition method. To resolve the issues of non-separable cross-product terms in the augmented Lagrangian function, we apply diagonal approximation method. Several examples have been studied to demonstrate that the proposed approach yields significant computational savings compared to the full-scale integrated model.

1. Introduction

Modern process industries operate as a large integrated complex that involve multiproduct, multipurpose, and multisite production facilities serving a global market. The process industries supply chain is composed of production facilities and distribution centers, where the final products are transported from the production facilities to distribution centers and then to retailers to satisfy the customers demand. In current global market, spatially distributed production facilities across various geographical locations can no longer be regarded as independent from each other and interactions between the manufacturing sites and the distribution centers should be taken into account when making decisions. In this context, the issues of enterprise planning and coordination across production plants and distribution facilities are important for robust response to global demand and to maintain business competitiveness, sustainability, and growth (Papageorgiou, 2009). As the pressure to reduce the costs and inventories increases, centralized approaches have become the main policies to address supply chain optimization. An excellent overview of the enterprise-wide optimization (EWO) and the challenges related to process industry supply chain is highlighted by Grossmann (2005). Varma, Reklaitis, Blau, and Pekny (2007) described the main concepts of EWO and presented the potential research opportunities in addressing the problem of EWO models and solution approaches.

Supply chain optimization can be considered an equivalent term for describing the enterprise-wide optimization (Shapiro, 2001) although supply chain optimization places more emphasis on logistics and distribution, whereas enterprise-wide optimization is aimed at manufacturing facilities optimization. Key issues and challenges faced by process industry supply chain are highlighted by Shah (2004, 2005). Traditional supply chain management planning decisions can be divided into three levels: strategic (long-term), tactical (medium-term), and operational (short-term). The long-term planning determines the infrastructure (e.g. facility location, transportation network). The medium-term planning covers a time horizon between few months to a year and is concerned with decisions such as production, inventory, and distribution profiles. Finally, short-term planning decision deals with issues such as assignment of tasks to units and sequencing of tasks in each unit. The short-term planning level covers time horizon between days to a few weeks and at production level, is typically refer to as scheduling. Wassick (2009) proposed a planning and scheduling model based on resource task network for an integrated chemical complex. He considered the enterprise-wide optimization of the liquid waste treatment network with their model. Kreipl and Pinedo (2004) discussed issues present in modeling the planning and scheduling decisions for supply chain management. For a multisite facilities, the size and level of interdependences between these sites present unique challenges to the integrated tactical production planning and day-to-day scheduling

* Corresponding author. Tel.: +1 732 445 2971; fax: +1 732 445 2421.
E-mail address: marianth@rutgers.edu (M.G. Ierapetritou).

0098-1354/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compchemeng.2011.08.007

Nomenclature

Indices

i	task
j	units
m	distribution market
n	event point
p	production site
s	material state
t	planning period

Sets

I	tasks
I^p	task that can be performed at site p
I_j^p	tasks that can be performed in unit j at site p
J	units
J^p	units that are located at site p
J_i^p	units that can perform task i at site p
M	distribution markets
N	event points
P_s	sites that can produced final product s
PS	production sites
S	material states
S_f^m	products that can be sold at market m
S_f^p	products that can be produced at site p
T	planning periods

Parameters

$d_s^{p,m}$	unit transport cost of material s from site p to market m
$Dem_s^{m,t}$	demand of product s at market m for period t
$FixCost_i^p$	fixed production cost of task i at site p
h_s^p	holding cost of product s at production site p
$stcap_s^p$	available maximum storage capacity for state s at site p
u_s^m	backorder cost of product s at distribution market m
$v_{i,j,p}^{min}$, $v_{i,j,p}^{max}$	minimum and maximum capacity of unit j when processing task i at site p
$VarCost_i^p$	unit variable cost of task i at site p
$\alpha_{i,j}^p$, $\beta_{i,j}^p$	constant, variable term of processing time of task i in unit j at site p
$\rho_{s,i}^{c,p}$, $\rho_{s,i}^{p,p}$	proportion of state s consumed, produced by task i respectively at site p

Variables

$b_{i,j,n}^{p,t}$	amount of material processed by task i in unit j at event point n at site p during period t
$D_s^{p,m,t}$	transportation of product s from site p to market m at period t
$Inv_s^{p,t}$	inventory level of state s at the end of the planning period t for site p
$stin_s^{p,t}$	initial inventory for state s in planning period t
$Tf_{i,j,n}^{p,t}$	finish time of task i in unit j at event point n in site p during period t
$Ts_{i,j,n}^{p,t}$	start time of task i in unit j at event point n in site p during period t
$U_s^{m,t}$	backorder of product s at market m in planning period t
$wv_{i,j,n}^{p,t}$	binary variable, task i active in unit j at event point n at site s during period t

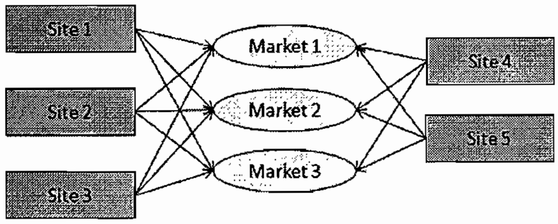

Fig. 1. Multisite production and distribution network.

problem and these challenges are highlighted by Kallrath (2002a). For further elucidation of various aspects of planning, the reader is directed to the work of Timpe and Kallrath (2000) and Kallrath (2002b).

A simple network featuring the multisite facilities is given in Fig. 1, where multiple products may be produced in individual process plants at different locations spread across geographic region and then transported to distribution centers to satisfy customers demand. These multisite plants produce a number of products driven by market demand under operating conditions such as sequence dependent switchovers and resource constraints. Each plant within the enterprise may have different production capacity and costs, different product recipes, and different transportation costs to the markets according to the location of the plants. To maintain economic competitiveness in a global market, interdependences between the different plants, including intermediate products and shared resources need be taken into consideration when making planning decisions. Furthermore, the planning model should take into account not only individual production facilities constraints but also transportation constraints because in addition to minimizing the production cost, it's important to minimize the costs of products transportation from production facilities to the distribution center. Thus, simultaneous planning of all activities from production to distribution stage is important in a multisite process industry supply chain (Shah, 1998).

Wilkinson, Cortier, Shah, and Pantelides (1996) proposed an aggregated planning model based on the resource task network framework developed by Pantelides (1994). Their proposed planning model considers integration of production, inventory, and distribution in multisite facilities. Lin and Chen (2007) developed a multistage, multisite planning model that deals with routings of manufactured products demand among different production plants. They simultaneously combine two different time scales (i.e. monthly and daily) in their formulation by considering varying time buckets. Verderame and Floudas (2009) developed an operational planning model which captures the interactions between production facilities and distribution centers in multisite production facilities network. Their proposed multisite planning with product aggregation model (Multisite-PPDM) incorporates a tight upper bound on the production capacity and transportation cost between production facilities and customers distribution centers in the supply chain network under consideration. A multisite production planning and distribution model is proposed by Jackson and Grossmann (2003) where they utilized nonlinear process models to represent production facilities. They have exploited two different decomposition schemes to solve the large-scale nonlinear model using Lagrangian decomposition. In temporal decomposition, the inventory constraints between adjoining time periods are dualized in order to optimize the entire network for each planning time period. In spatial decomposition technique, interconnection constraints between the sites and markets are dualized in order to optimize each facility individually. They conclude that temporal decomposition technique performs far better than spatial decomposition technique.

The traditional strategy to address planning and scheduling level decisions is to follow a hierarchical approach in which planning decisions are made first and then scheduling decisions are made using planning demand targets. However, this approach does not consider any interactions between the two decision making levels and thus the planning decisions may result in suboptimal or even infeasible scheduling problems. Due to significant interactions between planning and scheduling decisions levels in order to determine the global optimal solution it is necessary to consider the simultaneous optimization of the planning and scheduling decisions. However, this simultaneous optimization problem leads to a large problem size and the model becomes intractable when typical planning horizon is considered. For an overview of issues, challenges and optimization opportunities present in production planning and scheduling problem, the reader is referred to the work of Maravelias and Sung (2009).

In recent years, the area of integrated planning and scheduling for single site has received much attention. Different decomposition strategies are developed to effectively deal with a large scale integrated model. One of the existing approaches follows a hierarchical decomposition method, where the upper level planning problem provides a set of decisions such as production and inventory targets to the lower level problem to determine the detailed schedule. If the solution of lower level problem is infeasible, an iterative framework is used to obtain a feasible solution (Bassett, Pekny, & Reklaitis, 1996). To further improve this approach, tight upper bounds on production capacity are implemented in upper level problem in presence of an approximate scheduling model or aggregated capacity constraints (Shapiro, 2001; Shah, 2005). Another related idea is the one that follows a hierarchical decomposition within a rolling horizon framework. In this model detailed scheduling models are used for a few early periods and aggregated models are used for later periods (Dimitriadis, Shah, & Pantelides, 1997; Li & Ierapetritou, 2010a; Verderame & Floudas, 2008; Wu & Ierapetritou, 2007). A different decomposition strategy is based on the special structure of the large-scale mathematical programming model. The integrated planning and scheduling model has a block angular structure which arises when a single scheduling problem is used over multiple planning periods. The constraints matrix of the integrated problem has complicating variables that appear in multiple constraints. By making copies of the complicating variables, the complicating variables are transformed into complicating constraints (linking constraints) and these complicating constraints can be relaxed using the Lagrangian relaxation method. One major drawback of the Lagrangian relaxation (LR) is that there is duality gap between the solution of the Lagrangian relaxation method and original problem and to resolve this issue, augmented Lagrangian relaxation (ALR) method should be used (Li, Lu, & Michalek, 2008; Tosserams, Etman, Papalambros, & Rooda, 2006; Tosserams, Etman, & Rooda, 2008). Li and Ierapetritou (2010b) applied augmented Lagrangian optimization method to integrated planning and scheduling problem for single site plants. One disadvantage of ALR method is the non-separability of the relaxed problem which arises due to the quadratic penalty terms present in the objective function. To resolve the issue of the non-separability, Li and Ierapetritou (2010b) studied two different approaches. The first approach is based on linearization the cross-product terms using diagonal quadratic approximation (DQA) (Li et al., 2008). However, in this approach, an approximation of the relaxation problem is solved and it may not lead to a global optimal solution of the original problem. In the second approach, Li and Ierapetritou (2010b) proposed a two-level optimization method which solves an exact relaxation problem. However, the proposed two-level optimization strategy requires a non-smooth quadratic problem to be optimized at every iteration. They conclude that DQA-ALO method is more effective than the two-level optimization method for the integrated planning and scheduling problems.

Even though most companies operate in a multisite production manner, very limited attention has been paid on integrating planning and scheduling decisions for multisite facilities. The integrated planning and scheduling model for multisite facilities is important to ensure the consistency between planning and scheduling level decisions and to optimize production and transportation costs. Since the production planning and scheduling level deals with different time scales, the major challenge for the integration using mathematical programming methods lies in addressing large scale optimization models. The full-scale integrated planning and scheduling optimization model spans the entire planning horizon of interest and includes decisions regarding all the production sites and distribution centers. When typical planning horizon is considered, the integrated full-scale problem becomes intractable and a mathematical decomposition solution approach is necessary.

In this work, we apply augmented Lagrangian relaxation method to solve the multisite production and distribution optimization problem. The paper is organized as follow. The problem statement is given in Section 2, whereas Section 3 presents the problem formulation. The general augmented Lagrangian method and its application to multisite facility is given in Section 4. The results of examples studied are shown in Section 5 and the paper concludes with Section 6.

2. Problem statement

The supply chain network (Fig. 1) under investigation contains multiple batch production facilities which supply products to multiple distribution centers. Every production site may supply all distribution centers but all the products may not be produced at every production site. In the proposed model, we assume that we cannot sell more products to a market than the market forecast demand. Thus the requested demand acts as an upper bound on finished product sales. The proposed model tries to satisfy the market demands; however it allows for unsatisfied demand to be carried over to the next planning period (backorder) and also allows for partial order fulfillment. The unsatisfied demand and backorders are penalized on a daily basis in order to maximize the degree of customer demand order fulfillment. We assume that unlimited supply of raw materials is available and fixed and variable production, storage, and backorder costs are known for the planning horizon under consideration. The transportation costs are also assumed to be known. We further assume that there are no shipping delays in the network and the length of time of the planning horizon is such that the effects of transportation delays are neglected.

Given the daily demand profiles for each distribution center, the goals of the integrated planning and scheduling problem is to ascertain the daily production target profiles for each production facilities and product shipment profiles from production facilities to distribution centers so that demand is satisfied over the planning horizon under considerations (several months up to a year). The objective of the integrated problem is to minimize inventory, backorder, transportation, and production costs.

3. Multisite integrated model formulation

The multisite model includes production site constraints and distribution center (market) constraints. The set of products are to be produced at various production sites ($p \in PS$) and are to be distributed to a global market ($m \in M$) over planning horizon ($t \in T$). To formulate the integrated planning and scheduling model, an integrated modeling approach is proposed in which planning and scheduling decisions level constraints are incorporated. The planning horizon is discretized into fixed time length (daily production periods) and for each planning period, a detailed scheduling model for each batch production facilities is considered. The detailed scheduling model is based on continuous time representation and notion of event points (Ierapetritou & Floudas, 1998). The planning and scheduling decisions levels are inter-connected via production and inventory constraints. We extend the integrated planning and scheduling model for a single-site proposed by Li and Ierapetritou (2010b) to accommodate multisite production facilities serving multiple markets. The extended integrated multisite model is as follows.

$$\min \sum_{t}\sum_{p}\sum_{s\in S_f^p} h_s^p Inv_s^{p,t} + \sum_{t}\sum_{m}\sum_{s\in S_f^m} u_s^m U_s^{m,t} + \sum_{t}\sum_{m}\sum_{p}\sum_{s\in S_f^p\cap S_f^m} d_s^{p,m} D_s^{p,m,t} + \sum_{t}\sum_{p}\sum_{j\in J^p}\sum_{i\in I_j^p}\sum_{n}(FixCost_i^p wv_{i,j,n}^{p,t} + VarCost_i^p b_{i,j,n}^{p,t}) \quad (1a)$$

$$\text{s.t.}\quad Inv_s^{p,t} = Inv_s^{p,t-1} + P_s^{p,t} - \sum_{m} D_s^{p,m,t}, \quad \forall s\in S_f^p,\ p\in PS,\ t\in T \quad (1b)$$

$$U_s^{m,t} = U_s^{m,t-1} + Dem_s^{m,t} - \sum_{p\in P_s} D_s^{p,m,t}, \quad \forall s\in S_f^m,\ m\in M,\ t\in T \quad (1c)$$

$$st_{s,n=N}^{p,t} - stin_s^{p,t} = P_s^{p,t}, \quad \forall s\in S_f^p,\ p\in PS,\ t\in T \quad (1d)$$

$$stin_s^{p,t} = Inv_s^{p,t-1}, \quad \forall s\in S_f^p,\ p\in PS,\ t\in T \quad (1e)$$

$$\sum_{i\in I_j^p} wv_{i,j,n}^{p,t} \le 1, \quad \forall j\in J^p,\ n\in N,\ p\in PS,\ t\in T \quad (1f)$$

$$v_{i,j,p}^{\min} wv_{i,j,n}^{p,t} \le b_{i,j,n}^{p,t} \le v_{i,j,p}^{\max} wv_{i,j,n}^{p,t}, \quad \forall i\in I_j^p,\ j\in J^p,\ n\in N,\ p\in PS,\ t\in T \quad (1g)$$

$$st_{s,n}^{p,t} \le stcap_s^p, \quad \forall s\in S^p,\ n\in N,\ p\in PS,\ t\in T \quad (1h)$$

$$st_{s,n}^{p,t} = st_{s,n-1}^{p,t} - \sum_{i\in I_s^p}\rho_{s,i}^{c,p}\sum_{j\in J_i^p} b_{i,j,n}^{p,t} + \sum_{i\in I_s^p}\rho_{s,i}^{p,p}\sum_{j\in J_i^p} b_{i,j,n-1}^{p,t}, \quad \forall s\in S^p,\ n\in N,\ p\in PS,\ t\in T \quad (1i)$$

$$st_{s,n=1}^{p,t} = stin_s^{p,t} - \sum_{i\in I_s^p}\rho_{s,i}^{c,p}\sum_{j\in J_i^p} b_{i,j,n=1}^{p,t}, \quad \forall s\in S^p,\ p\in PS,\ t\in T \quad (1j)$$

$$Tf_{i,j,n}^{p,t} = Ts_{i,j,n}^{p,t} + \alpha_{i,j}^p wv_{i,j,n}^{p,t} + \beta_{i,j}^p b_{i,j,n}^{p,t}, \quad \forall i\in I_j^p,\ j\in J^p,\ n\in N,\ p\in PS,\ t\in T \quad (1k)$$

$$Ts_{i,j,n+1}^{p,t} \ge Tf_{i,j,n}^{p,t} - H(1 - wv_{i,j,n}^{p,t}), \quad \forall i\in I_j^p,\ j\in J^p,\ n\in N,\ p\in PS,\ t\in T \quad (1l)$$

$$Ts_{i,j,n+1}^{p,t} \ge Tf_{i',j,n}^{p,t} - H(1 - wv_{i',j,n}^{p,t}), \quad \forall i\in I_j^p,\ i'\in I_j^p,\ j\in J^p,\ n\in N,\ p\in PS,\ t\in T \quad (1m)$$

$$Ts_{i,j,n+1}^{p,t} \ge Tf_{i',j',n}^{p,t} - H(1 - wv_{i',j',n}^{p,t}), \quad \forall i,\ i'\in I_j^p,\ i'\ne i,j,\ j'\in J^p,\ n\in N,\ p\in PS,\ t\in T \quad (1n)$$

$$Ts_{i,j,n+1}^{p,t} \ge Ts_{i,j,n}^{p,t}, \quad \forall i\in I_j^p,\ j\in J^p,\ n\in N,\ p\in PS,\ t\in T \quad (1o)$$

$$Tf_{i,j,n+1}^{p,t} \ge Tf_{i,j,n}^{p,t}, \quad \forall i\in I_j^p,\ j\in J^p,\ n\in N,\ p\in PS,\ t\in T \quad (1p)$$

$$Ts_{i,j,n}^{p,t} \le H, \quad \forall i\in I_j^p,\ j\in J^p,\ n\in N,\ p\in PS,\ t\in T \quad (1q)$$

$$Tf_{i,j,n}^{p,t} \le H, \quad \forall i\in I_j^p,\ j\in J^p,\ n\in N,\ p\in PS,\ t\in T \quad (1r)$$

The objective function shown in Eq. (1a) minimizes the total costs of the integrated model, which includes variable inventory costs, backorder costs, transportation costs, and production costs and fixed production costs. The planning level is modeled by constraints (1b) and (1c). Eq. (1b) predicts the production targets ($P_s^{p,t}$), inventory targets ($Inv_s^{p,t}$), and shipping targets ($D_s^{p,m,t}$) for each product. The constraint describing each distribution center (market) is given in Eq. (1c). As shown in Eq. (1c), the backorder balance is performed for each customer market by considering the demand forecast ($Dem_s^{m,t}$) of that market and the sales all the shipments of the product from one or combination of all production sites to that market ($D_s^{p,m,t}$). Constraints (1d) assign production targets ($P_s^{p,t}$) of planning level solutions to scheduling level problem for each product to each production facility for different planning periods. Eq. (1e) represents the connection between the inventory level requirements for the scheduling problems to that of the different planning periods for each product. In addition to constraints (1a)–(1e), the model also includes detailed scheduling constraints (1f)–(1r) for each production site ($p\in PS$) and for each planning period ($t\in T$). These scheduling level constraints are allocation constraints (1f), production capacity constraints (1g), storage capacity constraints (1h), material balance constraints (1i) and (1j), and sequence constraints (1k)–(1r). Eqs. (1a)–(1r) comprise the complete multisite batch production facilities and multisite distribution centers considering planning and scheduling decisions.

4. Solution method

The full-scale integrated model gives rise to a large scale optimization problem which requires the use of decomposition methods to be solved effectively. The appropriate mathematical decomposition approach is decided by analyzing the constraints matrix of the full-scale model. If we denote the planning decision variables ($Inv_s^{p,t}$, $P_s^{p,t}$, $D_s^{p,m,t}$, $U_s^{m,t}$) as $X^{t,p}$ and scheduling decision variables as $Y^{t,p}$, then the structure of the integrated model can be illustrated as shown in a constraints matrix (Fig. 2). As it is seen in Fig. 2, the matrix has a block angular structure and these blocks are linked through the planning decisions variables, inventory and production targets for each production facilities ($Inv_s^{p,t}$, $P_s^{p,t}$). These complicating variables can be handled using augmented Lagrangian relaxation method described in the next section to obtain a decomposable structure.

4.1. Augmented Lagrangian decomposition

In order to obtain a decomposable structure, the complicating variables need to be transformed into complicating constraints and then, we can relax the model by eliminating complicating constraints from the total constraints set. The first step in obtaining the relaxation problem is to duplicate the planning inventory and production targets variables, using different variables in planning and scheduling problems and incorporate the coupling constraints (2f)–(2g) into the full-scale model. The production and inventory scheduling targets constraints are rewritten as (2q)–(2r). This

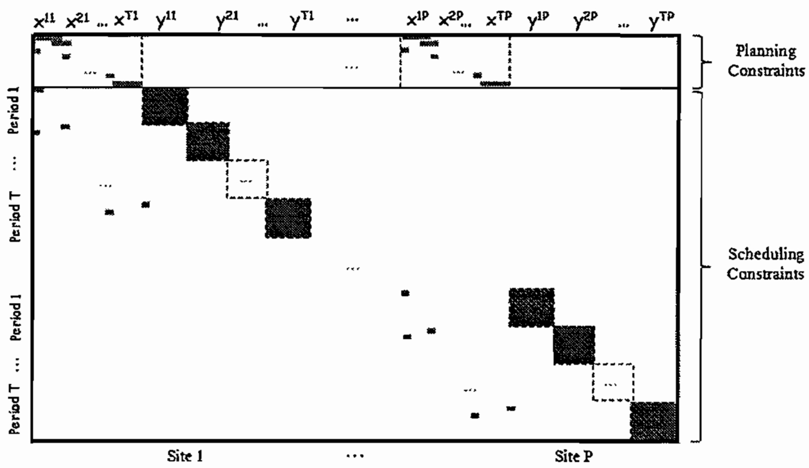

Fig. 2. Constraints matrix structure of an integrated multisite model.

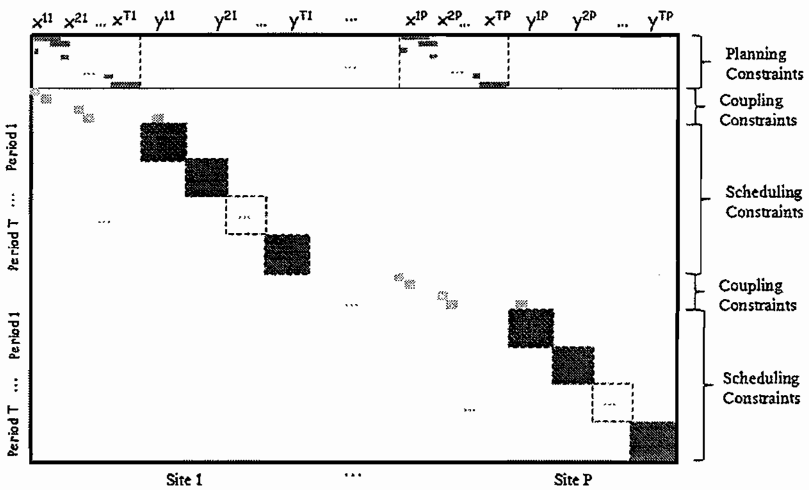

Fig. 3. Constraints matrix structure of a reformulated model.

transforms the constraints matrix (Fig. 2) into the matrix with complicating constraints shown in Fig. 3.

$$\min \sum_t \sum_p \sum_{s\in S_f^p} h_s^p Inv_s^{p,t} + \sum_t \sum_m \sum_{s\in S_f^m} u_s^m U_s^{m,t} + \sum_t \sum_m \sum_p \sum_{s\in S_f^p \cap S_f^m} d_s^{p,m} D_s^{p,m,t} + \sum_t \sum_p \sum_{i\in I^p} \sum_{j\in J_i^p} \sum_n (FixCost_i^p wv_{i,j,n}^{p,t} + VarCost_i^p b_{i,j,n}^{p,t}) \quad (2a)$$

$$\text{s.t.}\quad Inv_s^{p,t} = Inv_s^{p,t-1} + P_s^{p,t} - \sum_m D_s^{p,m,t}, \quad \forall s \in S_f^p,\ p \in PS,\ t \in T \quad (2b)$$

$$U_s^{m,t} = U_s^{m,t-1} + Dem_s^{m,t} - \sum_{p\in P_s} D_s^{p,m,t}, \quad \forall s \in S_f^m,\ m \in M,\ t \in T \quad (2c)$$

$$st_{s,n=N}^{p,t} - stin_s^{p,t} = PP_s^{p,t}, \quad \forall s \in S_f^p,\ p \in PS,\ t \in T \quad (2q)$$

$$stin_s^{p,t} = II_s^{p,t-1}, \quad \forall s \in S_f^p,\ p \in PS,\ t \in T \quad (2r)$$

$$II_s^{p,t} = Inv_s^{p,t}, \quad \forall s \in S_f^p,\ p \in PS,\ t \in T \quad (2f)$$

$$PP_s^{p,t} = P_s^{p,t}, \quad \forall s \in S_f^p,\ p \in PS,\ t \in T \quad (2g)$$

$$y^{p,t} \in Y, \quad \forall p \in PS,\ t \in T \quad (2h)$$

The complicating constraints (2f)–(2g) link the planning level constraints with the scheduling level constraints. The constraints (2h) express a compact representation of the scheduling level constraints (1f)–(1r) for each production site and planning period.

The reformulated model (2) is still not decomposable since the planning and scheduling problems are interconnected via coupling constraints thus we need to apply the Augmented Lagrangian relaxation method by dualizing the complicating constraints in Eqs. (2f) and (2g), which involves removing them from the reformulated model constraints set and adding them to the objective function multiplied by the Lagrange multipliers ($\lambda_s^{p,t}$, $\mu_s^{p,t}$) and quadratic penalty parameters (σ) as shown in Eq. (3a). Constraints

(3a)–(3h) correspond to the augmented Lagrangian relaxation problem.

$$\begin{aligned} f(\lambda,\mu,\sigma) &= \min \sum_t \sum_p \sum_{s\in S_f^p} h_s^p Inv_s^{p,t} + \sum_t \sum_m \sum_{s\in S_f^m} u_s^m U_s^{m,t} \\ &+ \sum_t \sum_m \sum_p \sum_{s\in S_f^p \cap S_f^m} d_s^{p,m} D_s^{p,m,t} \\ &+ \sum_t \sum_p \sum_{i\in I^p} \sum_{j\in J_i^p} \sum_n (FixCost_i^p wv_{i,j,n}^{p,t} \\ &+ VarCost_i^p b_{i,j,n}^{p,t}) + \sum_t \sum_{s\in S_f^p} \lambda_s^{p,t}(P_s^{p,t} - PP_s^{p,t}) \\ &+ \sum_t \sum_p \sum_{s\in S_f^p} \mu_s^{p,t}(Inv_s^{p,t} - II_s^{p,t}) \\ &+ \sum_t \sum_p \sum_{s\in S_f^p} \sigma\left[(P_s^{p,t} - PP_s^{p,t})^2 + (Inv_s^{p,t} - II_s^{p,t})^2\right] \end{aligned} \quad (3a)$$

$$s.t.\quad Inv_s^{p,t} = Inv_s^{p,t-1} + P_s^{p,t} - \sum_m D_s^{p,m,t}, \quad \forall s\in S_f^p,\ p\in PS,\ t\in T \quad (3b)$$

$$U_s^{m,t} = U_s^{m,t-1} + Dem_s^{m,t} - \sum_{p\in P_s} D_s^{p,m,t}, \quad \forall s\in S_f^m,\ m\in M,\ t\in T \quad (3c)$$

$$st_{s,n=N}^{p,t} - stin_s^{p,t} = PP_s^{p,t}, \quad \forall s\in S_f^p,\ p\in PS,\ t\in T \quad (3q)$$

$$stin_s^{p,t} = II_s^{p,t-1}, \quad \forall s\in S_f^p,\ p\in PS,\ t\in T \quad (3r)$$

$$y^{p,t}\in Y, \quad \forall p\in PS,\ t\in T \quad (3h)$$

To improve the convergence to the feasible solution and to avoid duality gap that may result with just the Lagrangian terms ($\lambda_s^{p,t}(P_s^{p,t} - PP_s^{p,t})$ and $\mu_s^{p,t}(Inv_s^{p,t} - II_s^{p,t})$), the quadratic penalty term ($\sigma\{(P_s^{p,t} - PP_s^{p,t})^2 + (Inv_s^{p,t} - II_s^{p,t})^2\}$) is applied to the relaxation formulation as given in (3a). However, the quadratic penalty term in the objective function of the relaxation problem has non-separable bilinear terms $P_s^{p,t}\cdot PP_s^{p,t}$ and $Inv_s^{p,t}\cdot II_s^{p,t}$. To resolve the non-separabilility issue, we apply the diagonal quadratic approximation (DQA) method to linearize the cross-product terms around the tentative solution $\overline{P}_s^{p,t}, \overline{PP}_s^{p,t}, \overline{Inv}_s^{p,t}, \overline{II}_s^{p,t}$ as shown in the following equations (Li et al., 2008).

$$(Inv_s^{p,t} - II_s^{p,t})^2 \approx (Inv_s^{p,t} - \overline{II}_s^{p,t})^2 + (\overline{Inv}_s^{p,t} - II_s^{p,t})^2 - (\overline{Inv}_s^{p,t} - \overline{II}_s^{p,t})^2$$

$$(P_s^{p,t} - PP_s^{p,t})^2 \approx (P_s^{p,t} - \overline{PP}_s^{p,t})^2 + (\overline{P}_s^{p,t} - PP_s^{p,t})^2 - (\overline{P}_s^{p,t} - \overline{PP}_s^{p,t})^2$$

The objective function (3a) can be thus be rewritten in decomposable form given by Eq. (4a').

$$f(\lambda,\mu,\sigma) = f_{pl} + \sum_{t\in T} \sum_{p\in PS} f_{sc}^{p,t} \quad (4a')$$

where the f_{pl} represents the objective function of the planning problem (4a)–(4c).

$$\begin{aligned} f_{pl}(\lambda,\mu,\sigma) &= \min_{P,Inv,D,U} \sum_t \sum_p \sum_{s\in S_f^p} h_s^p Inv_s^{p,t} + \sum_t \sum_m \sum_{s\in S_f^m} u_s^m U_s^{m,t} \\ &+ \sum_t \sum_m \sum_p \sum_{s\in S_f^p \cap S_f^m} d_s^{p,m} D_s^{p,m,t} \\ &+ \sum_t \sum_{s\in S_f^p} \lambda_s^{p,t} P_s^{p,t} + \sum_t \sum_p \sum_{s\in S_f^p} \mu_s^{p,t} Inv_s^{p,t} \\ &+ \sum_t \sum_p \sum_{s\in S_f^p} \sigma\left[\left(P_s^{p,t} - \overline{PP}_s^{p,t}\right)^2 + \left(Inv_s^{p,t} - \overline{II}_s^{p,t}\right)^2 \right. \\ &\left. - \left(\overline{P}_s^{p,t} - \overline{PP}_s^{p,t}\right)^2 - \left(\overline{Inv}_s^{p,t} - \overline{II}_s^{p,t}\right)^2\right] \end{aligned} \quad (4a)$$

$$s.t\quad Inv_s^{p,t} = Inv_s^{p,t-1} + P_s^{p,t} - \sum_m D_s^{p,m,t}, \quad \forall s\in S_f^p,\ p\in PS,\ t\in T \quad (4b)$$

$$U_s^{m,t} = U_s^{m,t-1} + Dem_s^{m,t} - \sum_{p\in P_s} D_s^{p,m,t}, \quad \forall s\in S_f^m,\ m\in M,\ t\in T \quad (4c)$$

where $f_{sc}^{p,t}$ represents the objective function of the scheduling sub-problems (5a)–(5d). The scheduling sub-problem is defined at each production site and for each planning period.

$$\begin{aligned} f_{sc}^{p,t}(\lambda,\mu,\sigma) &= \min_{PP^{p,t},II^{p,t},y^{p,t}} \sum_{i\in I^p} \sum_{j\in J_i^p} \sum_n (FixCost_i^p wv_{i,j,n}^{p,t} \\ &+ VarCost_i^p b_{i,j,n}^{p,t}) - \sum_{s\in S_f^p} \lambda_s^{p,t} PP_s^{p,t} \\ &- \sum_{s\in S_f^p} \mu_s^{p,t} II_s^{p,t} + \sum_{s\in S_f^p} \sigma\left[\left(\overline{P}_s^{p,t} - PP_s^{p,t}\right)^2 \right. \\ &\left. + \left(\overline{Inv}_s^{p,t} - II_s^{p,t}\right)^2\right] \end{aligned} \quad (5a)$$

$$s.t.\quad st_{s,n=N}^{p,t} - stin_s^{p,t} = PP_s^{p,t}, \quad \forall s\in S_f^p,\ p\in PS,\ t\in T \quad (5b)$$

$$stin_s^{p,t} = II_s^{p,t-1}, \quad \forall s\in S_f^p,\ p\in PS,\ t\in T \quad (5c)$$

$$y^{p,t}\in Y, \quad \forall p\in PS,\ t\in T \quad (5d)$$

These quadratic problems are solved using a general augmented Lagrangian optimization and diagonal quadratic approximation (ALO-DQA) algorithm which is outlined in Fig. 4.

The ALO-DQA algorithm can provide an optimal solution if the objective functions and constraints function are convex, feasible region is bounded and closed, and the step size is sufficiently small. The algorithm has the following parameters: the initial Lagrange multipliers ($\lambda_s^{p,t,0}, \mu_s^{p,t,0}$) which are chosen to be zero, the initial penalty parameter ($\sigma^0 > 1$), and the iteration counter k which is set to 1. The convergence tolerance ($\varepsilon > 0$) for the coupling constraints is 1 and the parameters $\beta\in(0,1)$ (e.g., 0.4) and $\alpha > 1$.

The augmented Lagrangian multipliers are updated at every iteration as shown in Fig. 4 while the quadratic penalty parameters are updated only if the improvement of the current iteration is not large enough. The ALO-DQA method alternates between solving an optimization planning problem (4) and solving optimization scheduling sub-problems (5). The solution of each problem is used to linearize the non-separable terms and the algorithm terminates

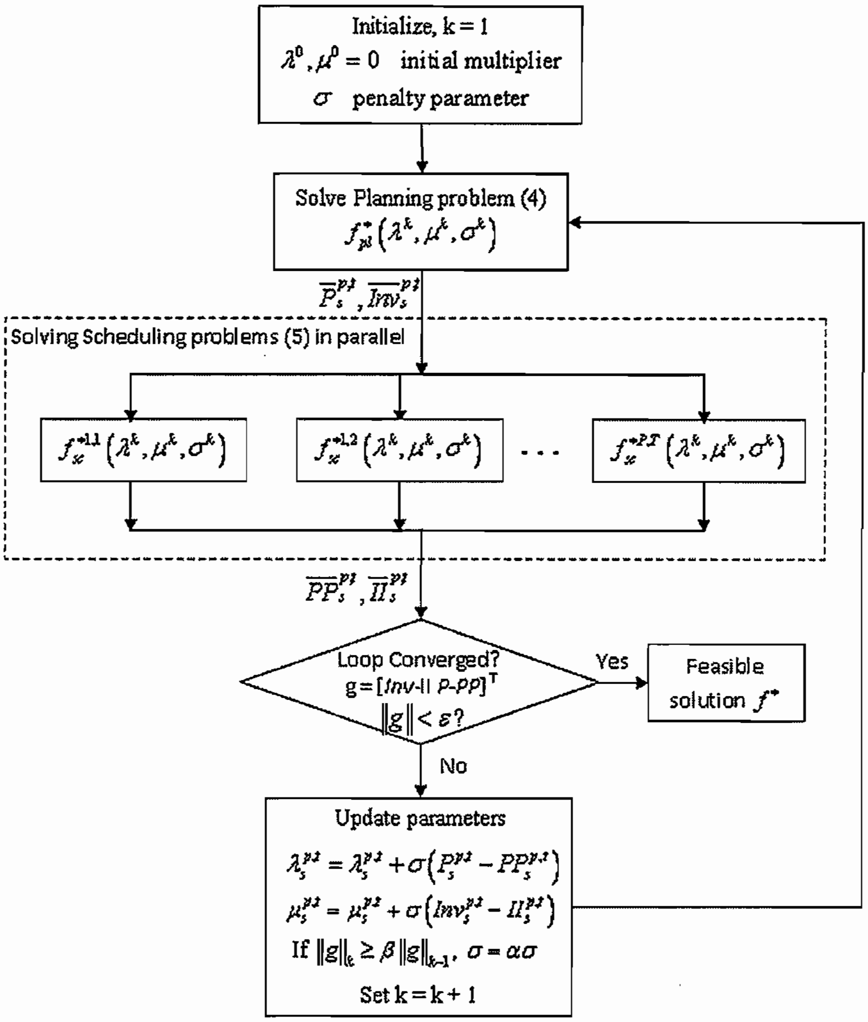

Fig. 4. Augmented Lagrangian-DQA decomposition algorithm.

when the consistency constraints (g) have met the pre-defined tolerance or when the given iteration limit is reached. In the next section, three numerical examples are solved using the ALO-DQA method.

5. Numerical examples

The proposed multiproduct and multisite production facility model is applied to the two examples of supply chain with a planning horizon of 3 months (90 days) and scheduling horizon of 1 production shift (8 h). The full-scale integrated planning and

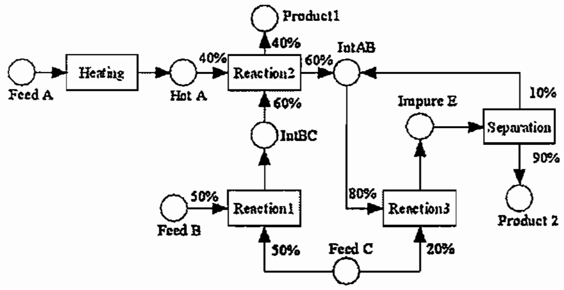

Fig. 5. Production facility state and task network (STN) representation (example 1).

scheduling problem corresponds to a mixed integer linear programming (MILP) problem while in the ALO-DQA method, the planning problem is quadratic programming (QP) problem and scheduling sub-problems are mixed integer quadratic programming (MIQP) problem. The multisite models were implemented

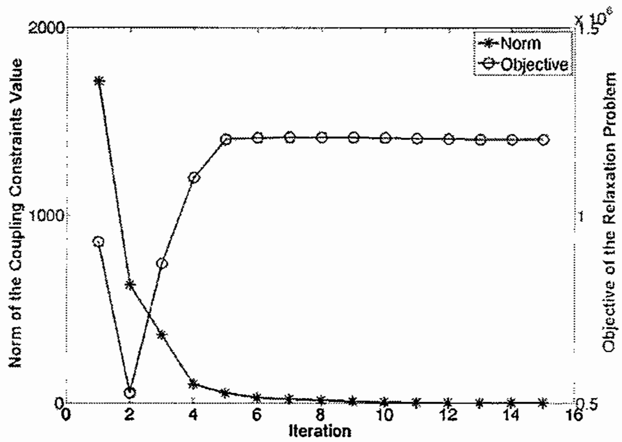

Fig. 6. Solution procedure of the ALO-DQA method (example 1 and 90 planning periods).

Table 1
Model statistics of full-space integrated problem.

Period (T)	Example 1			Example 2			Example 3		
	Binary var.	Cont. var.	Const.	Binary var.	Cont. var.	Const.	Binary var.	Cont. var.	Const.
5	720	4006	10,465	720	5011	10,654	1440	8851	21,178
10	1440	8011	20,935	1440	10,021	21,319	2880	17,701	42,373
15	2160	12,016	31,405	2160	15,031	31,984	4320	26,551	63,568
30	4320	24,031	62,815	4320	30,061	63,979	8640	53,101	127,153
45	6480	36,046	94,225	6480	45,091	95,974	12,960	79,651	190,738
60	8640	48,061	125,635	8640	60,121	127,969	17,280	106,201	254,323
90	12,960	72,091	188,455	12,960	90,181	191,959	25,920	159,301	381,493

Table 2
Computational results for example 1.

T	Full-space model			ALO-DQA method $\sigma^0 = 1.0$, $\alpha = 2.0$				
	CPU sec	f^*	Gap (%)	Iter. k	CPU sec	f^*	$\lambda g + \sigma\|\|g\|\|^2$	$\|\|g\|\|$
5	3600	58,073	0.63	10	34.68	59,500	0.62	0.22
10	3600	119,227	1.91	9	53.53	122,903	34.80	0.66
15	3600	194,567	2.07	12	101.97	199,983	−157.42	0.89
30	3600	390,528	2.27	12	183.21	403,208	−17.96	0.85
45	3600	592,649	2.44	12	292.25	608,059	−166.23	0.76
60	3600	791,399	3.28	13	436.03	803,867	−128.27	0.59
90	3600	1,225,376	6.93	15	760.50	1,203,194	−324.10	0.92

Table 3
Computational results for example 2.

T	Full-space model			ALO-DQA method $\sigma^0 = 0.2$, $\alpha = 1.2$				
	CPU sec	f^*	Gap (%)	Iter k	CPU sec	f^*	$\lambda g + \sigma\|\|g\|\|^2$	$\|\|g\|\|$
5	52.15	249,097	0.00	36	49.37	250,052	99.31	0.56
10	3600	502,983	0.07	36	66.85	504,687	14.18	0.65
15	3600	756,527	0.25	40	112.49	762,006	−211.75	0.89
30	3600	1,381,017	0.50	40	209.29	1,390,602	−216.91	0.95
45	3600	1,964,466	1.05	40	262.49	1,968,542	−234.33	0.98
60	3600	2,594,945	0.75	41	499.90	2,608,491	−1.59	0.70
90	3600	4,009,291	1.04	41	709.48	4,024,795	24.81	0.75

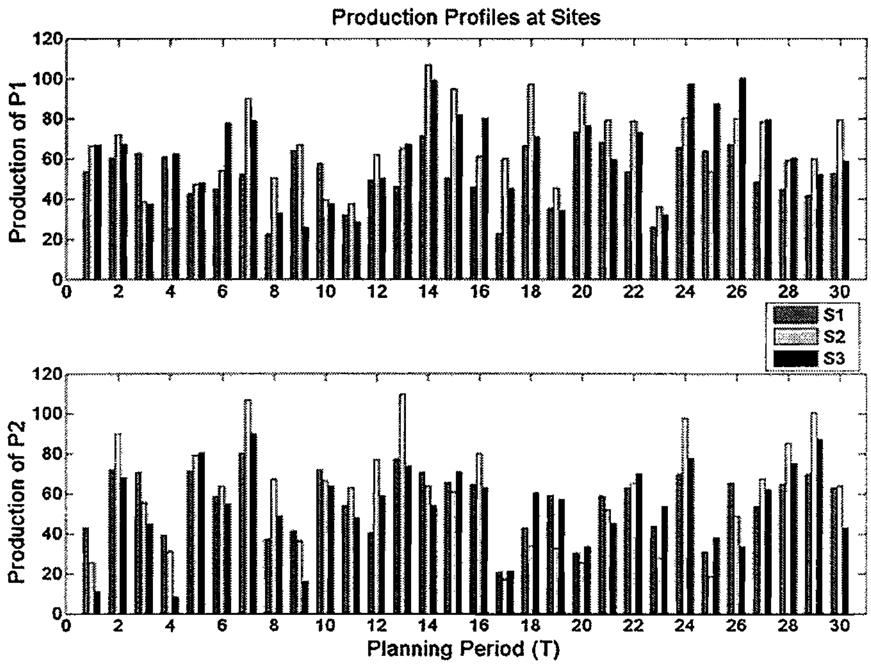

Fig. 7. Production profiles of products (P1 and P2) at production sites (S1, S2, and S3) obtained using the ALO-DQA method (Example 1 and 30 planning periods).

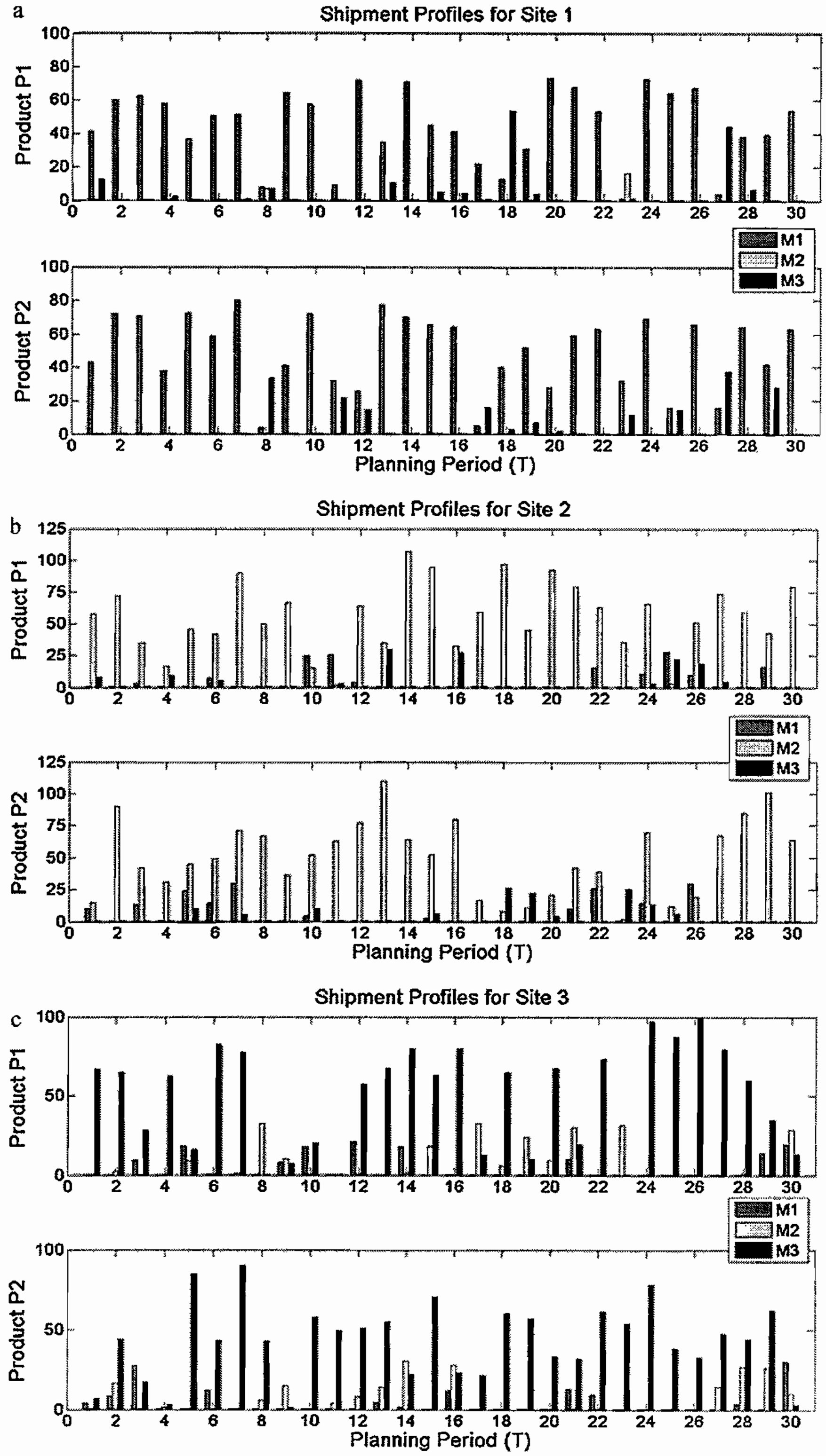

Fig. 8. (a) Shipment profiles of products (P1 and P2) for production site 1 to markets (M1, M2, and M3) obtained using the ALO-DQA method (example 1 and 30 planning periods). (b) Shipment profiles of products (P1 and P2) for production site 2 to markets (M1, M2, and M3) obtained using the ALO-DQA method (example 1 and 30 planning periods). (c) Shipment profiles of products (P1 and P2) for production site 3 to markets (M1, M2, and M3) obtained using the ALO-DQA method (example 1 and 30 planning periods).

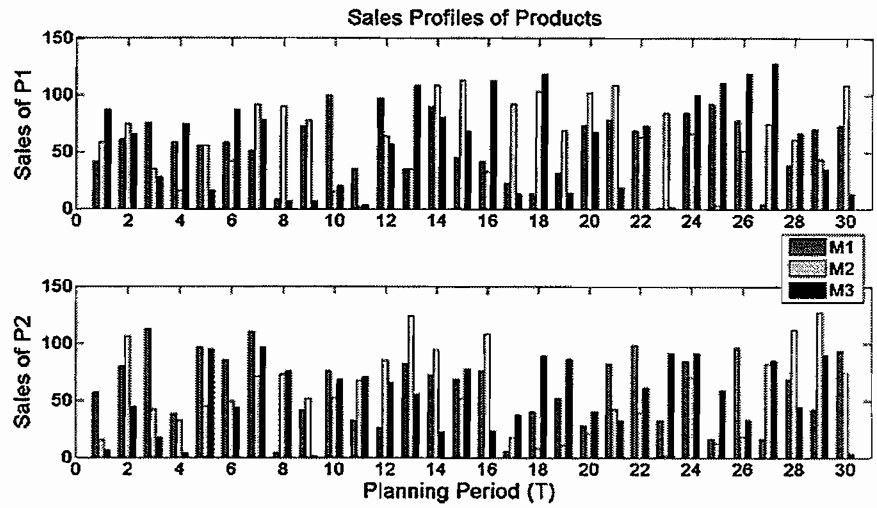

Fig. 9. Sales profiles of products (P1 and P2) for markets (M1, M2, and M3) obtained using the ALO-DQA method (example 1 and 30 planning periods).

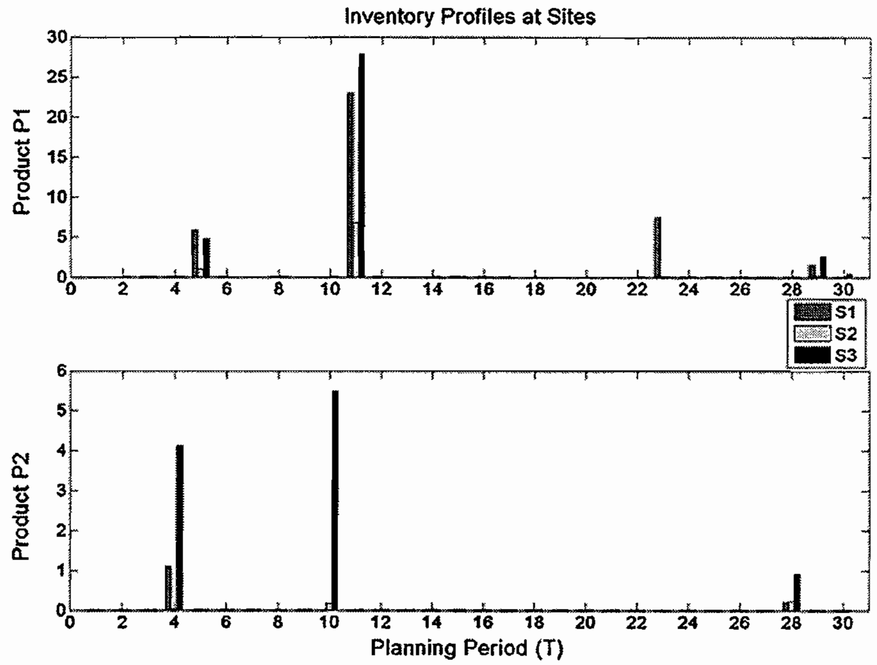

Fig. 10. Inventory profiles of products (P1 and P2) at production sites (S1, S2, and S3) obtained using the ALO-DQA method (example 1 and 30 planning periods).

using GAMS 23.6 (2010) and solved using CPLEX 12.2 on 2.53 GHz Precision T7500 Tower Workstation with 6 GB RAM. The scheduling sub-problems (MIQP) in the ALO-DQA method are solved in parallel for each planning period and each production site thus improving the efficiency of the algorithm. In all the examples studied in our work, we consider limited storage capacity for final products and intermediate materials.

Example 1: We studied a small example that has 3 production sites serving 3 markets. Each production site contains a multiproduct, multitask batch process plant that produces two products, P1 and P2 (Kondili, Pantelides, & Sargent, 1993). The state and task representation (STN) of the production plant is given in Fig. 5 and the data for the example are given in Appendix. The process parameters, production and inventory costs, and shipping and backorder costs are given in Tables A1–A3, respectively. Continuous time scheduling problem is solved using 6 event points and 8 h time horizon. The daily demand data for the example 1 is given in Fig. A1 for planning horizon of 90 days.

The full-scale model statistics for example 1 is shown in Table 1 and results are given in Table 2 for time periods 5–90. As the time periods increases, the problem becomes difficult to solve to optimality as observed by the optimality gap (%) in Table 2 for the full-space model. We compared the performance and quality of the full-scale model to that of the ALO-DQA in Table 2. From Fig. 6, it can be observed that the augmented Lagrangian algorithm converges to a feasible solution of the original problem as the norm value of the coupling constraints ($||g||$) converges to zero. The quality of the feasible solution (f^*) obtained using the ALO-DQA method may be inferior to the full-scale model since the ALO-DQA strategy solves an approximation version of the relaxation problem. The key information that the integrated model solution provides is production, shipping, sales, inventory, and backorder profiles. The profiles in the solution time and relative gap (%) given in Table 4. The progress of the solution procedure of the ALO-DQA method for 90 time periods is shown in Fig. 13. Table 4 shows the solutions of the integrated full-scale problem and the ALO-DQA decomposition method. Similar to results of examples 1 and 2, significant computational savings are observed when decomposition is applied compared to the full-scale model. Furthermore, for example 3, the ALO-DQA method is able to provide a better solution than the one reported by the full-scale model for planning periods of $T = 45$, 60, and 90.

6. Conclusions

This work addresses the problem of integrated planning and scheduling for multisite, multiproduct and multipurpose batch plants using the augmented Lagrangian method. The integrated multisite model is proposed by extending single site formulation of Li and Ierapetritou (2010b). The shipping costs from the production sites to distribution markets are taken into account explicitly in the integrated problem. Given the fixed demand forecast the model optimizes the production, inventory, transportation, and backorder costs. Temporal decomposition scheme was developed to address the large scale model resulting from multiperiod planning and scheduling problem. The augmented Lagrangian relaxation with diagonal approximation method allowed solution of the scheduling optimization problems into parallel. Three example problems solved to illustrate the advantages of applying the augmented Lagrangian decomposition scheme. With the proposed decomposition method, faster solution times were realized.

Acknowledgements

The authors gratefully acknowledge financial support from the National Science Foundation under Grants CBET 0966861 and GAANN.

Appendix A.

Fig. A1.
Tables A1–A3.

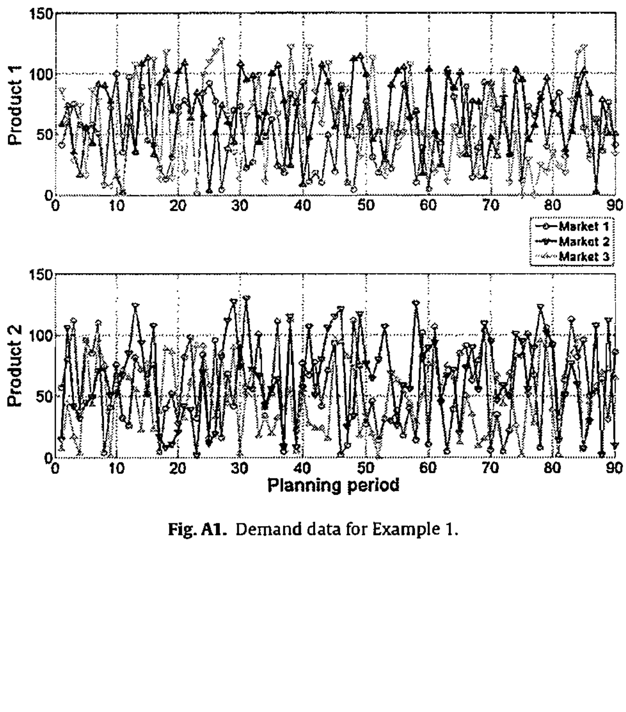

Fig. A1. Demand data for Example 1.

Table A1
Example 1: Process data for production sites.

Production site	Unit	Capacity	Suitability	Processing time
S1	Heater	100	Heating	1.0
	Reactor 1	50	Reactions 1,2,3	2.0,2.0,1.0
	Reactor 2	80	Reactions 1,2,3	2.0,2.0,1.0
	Sill	200	Separation	1 for Product 2 2 for IntAB
S2	Heater	150	Heating	1.0
	Reactor 1	100	Reactions 1,2,3	2.0,2.0,1.0
	Reactor 2	80	Reactions 1,2,3	2.0,2.0,1.0
	Sill	300	Separation	1 for Product 2 2 for IntAB
S3	Heater	100	Heating	1.0
	Reactor 1	75	Reactions 1,2,3	2.0,2.0,1.0
	Reactor 2	100	Reactions 1,2,3	2.0,2.0,1.0
	Sill	150	Separation	1 for Product 2 2 for IntAB

	State	Storage capacity	Initial amount
S1	Feed A	100,000	100,000
	Feed B	100,000	100,000
	Feed C	100,000	100,000
	Hot A	100	0.0
	Int AB	200	0.0
	Int BC	150	0.0
	Impure	200	0.0
	Product 1	400	0.0
	Product 2	375	0.0
S2	Feed A	100,000	100,000
	Feed B	100,000	100,000
	Feed C	100,000	100,000
	Hot A	130	0.0
	Int AB	250	0.0
	Int BC	150	0.0
	Impure	300	0.0
	Product 1	300	0.0
	Product 2	425	0.0
S3	Feed A	100,000	100,000
	Feed B	100,000	100,000
	Feed C	100,000	100,000
	Hot A	115	0.0
	Int AB	250	0.0
	Int BC	150	0.0
	Impure	300	0.0
	Product 1	350	0.0
	Product 2	400	0.0

Table A2
Production and inventory cost data.

Production sites		Fixed cost	Variable costs		Inventory cost
S1	Heating	150	1.0	P1	10
	Reactions 1,2,3	100	0.5	P2	10
	Separation	150	1.0		
S2	Heating	175	1.5	P1	15
	Reactions 1,2,3	125	0.75	P2	15
	Separation	175	1.5		
S3	Heating	100	0.8	P1	8
	Reactions 1,2,3	75	0.4	P2	8
	Separation	100	0.8		

Table A3
Transportation and backorder cost data.

Market		Transportation cost		Backorder cost
M1	S1	5	P1	100
	S2	8	P2	100
	S3	10		
M2	S1	10	P1	150
	S2	5	P2	150
	S3	8		
M3	S1	8	P1	75
	S2	10	P2	75
	S3	8		

Table 4
Computational results for example 3.

T	Full-space model			ALO-DQA method σ^0 = 1.0, α = 2.0				
	CPU sec	f^*	Gap (%)	Iter k	CPU sec	f^*	$\lambda g+\sigma\|\|g\|\|^2$	$\|\|g\|\|$
5	3600	141,129	4.11	11	71.11	151,551	24.28	0.99
10	3600	276,230	4.40	13	144.97	295,756	609.56	0.88
15	3600	408,436	4.97	12	192.53	435,078	200.49	0.58
30	3600	784,535	6.05	13	378.51	822,924	−31.96	0.91
45	3600	1,259,638	12.97	16	749.96	1,233,187	−905.25	0.80
60	3600	8,123,074	82.11	16	984.92	1,629,968	−718.61	0.76
90	3600	8,445,397	74.11	15	1263.73	2,437,213	−1271.17	0.85

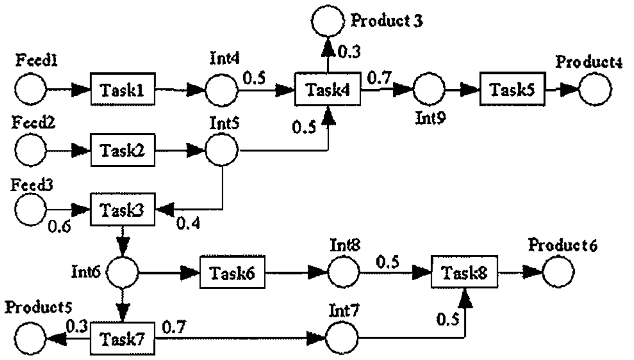

Fig. 11. Production facility state and task network (STN) representation (example 2).

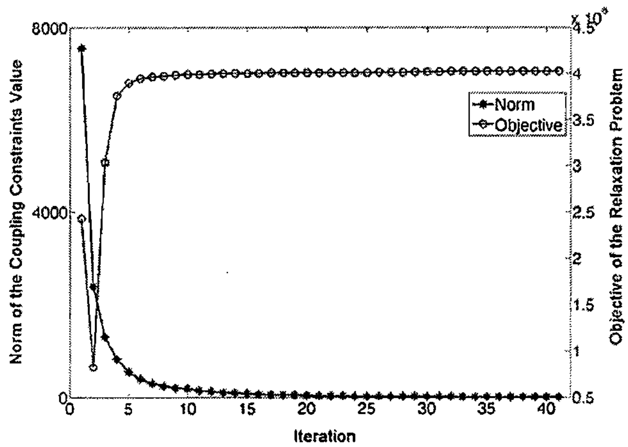

Fig. 12. Solution procedure of the ALO-DQA method (example 2, T = 90 periods).

for example 1 obtained using the ALO-DQA method are shown in Figs. 7–10 for 30 planning periods. The production profiles for production sites (S1, S2, and S3) are shown in Fig. 7. The transportation profile of products (P1 and P2) from production site (S1, S2, and S3) to market place (M1, M2, and M3) is given in Fig. 8a–c. Note that as shown in Fig. 8a–c, the production sites 1, 2, and 3 mainly satisfy the demand of markets 1, 2, and 3, respectively. These transportation profiles are expected based on the shipping cost. The total sale of products (P1 and P2) at markets (M1, M2, and M3) is given in Fig. 9. The variable inventory holding cost is highest at production site 2 and lowest at site 3 and the model solution gives an inventory profiles (Fig. 10) that has highest holdup at site 3 and lowest holdup at site 2. As expected, the advantage of the proposed decomposition approach is shown for bigger problems. So for larger number of time periods (T = 90 periods), better solutions were obtained using the ALO-DQA method than by the full-scale model.

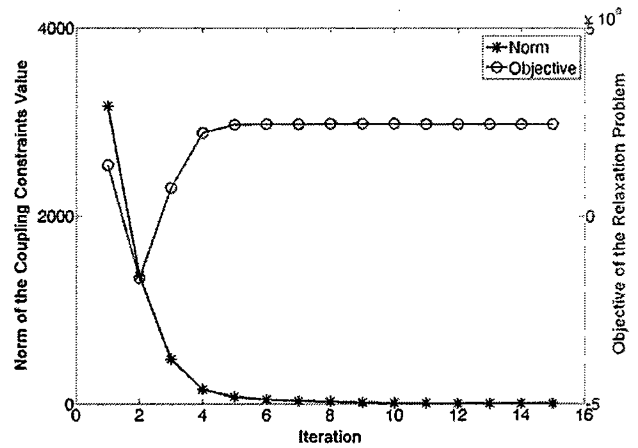

Fig. 13. Solution procedure of the ALO-DQA method (example 3, T = 90 periods).

Example 2: In example 2, we consider a network of 3 production sites producing 4 different products (P3–P6) and serving 3 global markets. All 3 production sites have batch facilities whose STN network is shown in Fig. 11 (Kondili, 1987). This batch facility produces 4 products (P3, P4, P5, and P6) through 8 tasks from three feeds and there are 6 intermediates state in the system. The full-scale model statistics for this example are shown in Table 1 and results are shown in Table 3. The full-scale problem is much easier to solve compared with example 1, even though this example is bigger than example 1. Thus, the production recipe, and the parameters relating to production, capacity, demand, and costs have significant effect on the performance of the full-scale model. To further improve the integrated model performance, we applied the ALO-DQA method and the results are shown in Table 3. The solution convergence profile for planning period 90 is shown in Fig. 12. The performance of the ALO-DQA method depends on the choice of the initial values of Lagrange multipliers, penalty parameters, and other algorithm parameters. By selecting appropriate values of these parameters, we can improve on the quality of the feasible solution obtained by the ALO-DQA. The results with set of parameters $\lambda^0 = 0$, $\mu^0 = 0$, $\sigma^0 = 0.20$, $\alpha = 1.2$, $\beta = 0.4$ are shown in Table 3. Significant CPU time savings is reported when the ALO-DQA method is used compared to the integrated full-scale model.

Example 3: In example 3, we consider a network of 6 production sites producing 6 different products (P1–P6) and serving 9 global markets. Of the 6 production sites, 3 batch production facilities have the network shown in Fig. 5 and produce 2 products (P1 and P2) and 3 production sites have the batch facilities whose STN network is shown in Fig. 11 and produce 4 products (P3, P4, P5, and P6) through multipurpose units.

The model statistics of example 3 are shown in Table 1. As expected, the complexity of the integrated full-scale model increases as the planning horizon increases and this is reflected