Capacity Expansion Planning Through Augmented Lagrangian Optimization and Scenario Decomposition

Zukui Li and Marianthi G. Ierapetritou
Dept. of Chemical and Biochemical Engineering, Rutgers University, Piscataway, NJ 08854

DOI 10.1002/aic.12614
Published online June 16, 2011 in Wiley Online Library (wileyonlinelibrary.com).

Stochastic programming is a typical method for addressing the uncertainties in capacity expansion planning problem. However, the corresponding deterministic equivalent model is often intractable with considerable number of uncertainty scenarios especially for stochastic integer programming (SIP) based formulations. In this article, a hybrid solution framework consisting of augmented Lagrangian optimization and scenario decomposition algorithm is proposed to solve the SIP problem. The method divides the solution procedure into two phases, where traditional linearization based decomposition strategy and global optimization technique are applied to solve the relaxation problem successively. Using the proposed solution framework, a feasible solution of the original problem can be obtained after the first solution phase whereas the optimal solution is obtained after the second solution phase. The effectiveness of the proposed strategy is verified through a numerical example of two stage stochastic integer program and the capacity expansion planning examples. © 2011 American Institute of Chemical Engineers *AIChE J,* 58: 871–883, 2012
Keywords: augmented Lagrangian relaxation, stochastic integer programming, capacity expansion

Introduction

Capacity planning is a major strategic decision in the process industry. The optimal planning comprises of selection of processes from among competing alternatives and timing of capacity expansions in a way that maximizes the net present value (NPV) of the project over a planning horizon. In early work on the subject,[1,2] an general assumption is made that the predictions for prices, as well as demands and availabilities of chemical products and raw materials are known with certainty. However, a more realistic scenario is one where those key parameters are only partially known, and typically there is significant uncertainty regarding their future values. Thus, the capacity planning decisions have to be made in the face of uncertainty, and as a result, the consideration of uncertainty transforms the deterministic capacity expansion planning to a stochastic optimization problem.

Models dealing with uncertainties can be mainly divided into the following groups: two stage (multistage) stochastic programming, probabilistic constraint programming, fuzzy programming, robust optimization, and parametric programming. Among them, the two-stage (multistage) stochastic programming with recourse are ideally suited for addressing the problem of capacity expansion planning since it combines "here and now" with "wait and see" decisions.

In the traditional stochastic programming approaches for capacity planning,[3,4] it is assumed that the capacity expansion decisions are made before the uncertainty is realized, and only some recourse operational actions (operating levels,

Correspondence concerning this article should be addressed to M. G. Ierapetritou at marianth@soemail.rutgers.edu.

purchases, sales) can be taken to correct any infeasibilities. Thus, the binary capacity expansion decisions are modeled as first stage variables and the second or later stage problem correspond to linear program. Since all integrality capacity expansion decisions are restricted to the first stage of the problem, standard stochastic programming decomposition methods (such as L-shaped method) can be used to solve these problems.

Multi-stage models extend the two-stage stochastic programming models by allowing revised decisions in each time stage based upon the uncertainty realized so far.[5,6] The uncertainty information in a multistage stochastic program is generally modeled as a multilayered scenario tree, and the optimization problem consists of determining an expansion schedule that takes into consideration these scenarios. However, the multistage stochastic program with integer variables in all stages cannot be addressed using a nested Benders decomposition, which is applicable to its continuous counterpart.

Treating all the capacity expansion decisions as first-stage decision variables could lead to sub-optimal solution since it only allows operational decisions as recourse correction actions facing uncertainty. To address this disadvantage, Ahmed and Sahinidis[7] study the stochastic capacity expansion problem where the economies-of-scale in expansion costs are handled via fixed-charge cost functions, the resulting formulation is a multistage stochastic mixed integer program with binary variables in all stages. With the more realistic modeling method, binary decision variables have to be defined for each possible second or later stage scenario, thus the combinatorial complexity of the problem increases exponentially. The corresponding stochastic integer programming (SIP) problem is computationally demanding since it combines the difficulties from both the stochastic and combinatorial nature of the problems. When discrete (integer) restrictions are imposed on some of the second or later stage variables, the stochastic programming problem loses some desirable properties such as convexity and continuity of the recourse cost function. Solution methods that are successful for stochastic linear programming cannot simply be adapted to the stochastic program with integer second-stage variables. Indeed, the expected recourse function is discontinuous and the set of first-stage decisions that yield second-stage feasible solutions is known to be general nonconvex.

In the literature, the systematic investigation of stochastic integer programs only started during the 1990s. The first use of decomposition methods in stochastic programs having discrete decisions in the second stage is the integer L-shaped method.[8] As a dual to the L-shaped methods, the scenario decomposition method[9] considers subproblems corresponding to scenarios (i.e., realizations in the two-stage setting) and is governed by finding good dual multipliers. Loketangen and Woodruff[10] developed a scenario decomposition method employing augmented Lagrangian for the dualization and tabu search for the resulting quadratic mixed integer subproblems. Sen and Higle[11] developed the disjunctive decomposition method which uses cuts in a lower dimension space of the scenario subproblem and generate cuts based on disjunctive programming for one scenario subproblem and uses a simple translation to make it valid for different scenarios. Except from the above primal and dual decomposition methods, there also exists several enumerative algorithms. Ruszczynski et al.[12] proposed a stochastic version of the deterministic branch and bound algorithm which uses statistical estimates of function values instead of using exact computation of the objective function at many possible solutions, since exact computations are in general far too expensive. Schultz et al. [13] proposed a solution procedure based on enumeration and bounding in the first-stage while handling the second-stage by algebraic methods exploiting problem similarities.

Although several methods have been proposed as reviewed above, the solution of SIP problems is still identified as a challenging problem for realistic applications. In this article, we present a scenario decomposition approach for the SIP model of the capacity expansion planning problem. The method is based on augmented Lagrangian relaxation (ALR) technique, which is an approach that is often applied to models with a block diagonal structure. The augmented Lagrangian optimization method has been used in several applications in areas such as power generation scheduling,[14] multidisciplinary design,[15] etc. Here, we extended its application to the capacity expansion planning problem since the model takes similar structure. In such models, distinct blocks of variables and constraints can be identified that are linked with a few ''linking'' constraints and variables. The ALR method avoids the major drawback of classical Lagrangian relaxation method, i.e., there exists the duality gap between the solution of the Lagrangian dual problem and the solution of original problem, and also the feasibility of the solution often needs to be recovered through heuristic steps. One drawback of the ALR method is the inseparable issue of the relaxed problem, which avoids the solution of the relaxation problem in a decomposed way. To address this issue, different methods for the solution of the relaxation problem have been proposed in the literature, such as Diagonal Quadratic Approximation, Auxiliary Problem Principle, Block Coordinate Descent, etc.[16–18] However, these methods only ensure that feasible solution of original problem is obtained for the SIP based capacity planning problem, which is often not enough for realistic application. So in this work, we proposed a hybrid solution framework consisting of two phases: in the first phase, a feasible solution for original problem is obtained, whereas in the second phase, global optimization is applied to solve the relaxation problem, which is reformulated as a mixed integer nonlinear optimization problem with implicit objective functions. The evaluation of the objective function can be performed through parallel solution of a set of subproblems. In such a way, the quality of the final solution can be further improved and optimal solution for the original problem can be obtained.

The content of this article is organized as follows. A two stage SIP model for the capacity expansion problem is presented in the ''Capacity Expansion Planning Model'' section, and a general solution framework based on augmented Lagrangian optimization is given in the ''Augmented Lagrangian Optimization'' section followed by several traditional decomposition strategies for the solution of the relaxation problem. A hybrid solution framework for the stochastic capacity planning problem is presented in the ''Hybrid Solution Strategy'' section. The proposed method is further studied through a numerical example and the application in chemical process planning problems.

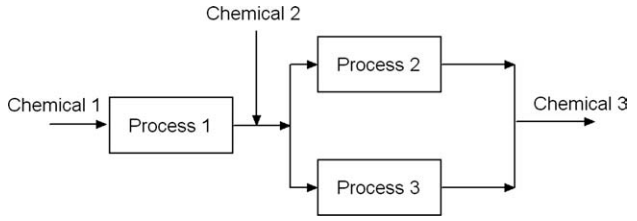

Figure 1. Example of a network of chemical processes.

Capacity Expansion Planning Model

Generally, the capacity expansion planning problem for a chemical process involves a network of chemical processes that are interconnected by raw materials, intermediate and final products that may be purchased from and/or sold to different markets as shown in Figure 1 which consists of three chemical processes and three chemicals. Also given are the prices, demands, and availabilities of chemicals as well as investment and operating cost data over a time period, whereas their exact values are not known for the future, described by a set of discrete scenarios with known probability value.

The decision problem is to determine the timing and the level of capacity acquisitions for a set of production facilities, along with a policy for allocating the available capacity to satisfy the demand of a set of product families, while minimizing the expected total discounted investment and allocation cost for a planning horizon divided into a number of periods.

In this work, we modeled the capacity expansion planning problem as two stage stochastic programming problem based on the formulation proposed by Liu and Sahinidis.[4] The difference from the original model is that we treat the capacity expansion decisions for the first time period $Y1_i$, $E1_i$, $Q1_i$ as first stage decisions and treat the other capacity expansion decisions (in the rest periods) Y_{its}, E_{its}, Q_{its} ($t = 2,\ldots,T$) and all the operational decisions (in all periods) W_{its}, P_{jlts}, S_{jlts} ($t = 1,\ldots,T$) as the second-stage decision variables. The objective of the proposed model is to address a problem with both integer first stage and integer second stage problem, which cannot be addressed with the traditional Benders decomposition strategy. In this kind of framework, only the first stage decision is considered important and could be implemented in a rolling horizon framework. Under the discrete (scenario) description of the uncertainty, the capacity expansion planning problem is modeled as the following two stage SIP problem:

$$\min -NPV = -\sum_s p_s \left\{ \sum_j \sum_l \sum_t \gamma_{jlts} S_{jlts} - \sum_i \sum_t (\alpha_{it} E_{its} + \beta_{it} Y_{its}) - \sum_i \sum_t \delta_{its} W_{its} - \sum_j \sum_l \sum_t \Gamma_{jlts} P_{jlts} \right\} \quad (1a)$$

s.t.

$$Y_{its} E^L_{it} \le E_{its} \le Y_{its} E^U_{it} \qquad \forall i,t,s \quad (1b)$$

$$Q_{its} = Q_{i,t-1,s} + E_{its} \qquad \forall i,t,s \quad (1c)$$

$$\sum_t Y_{its} \le NEXP(i) \qquad \forall i,s \quad (1d)$$

$$\sum_i (\alpha_{it} E_{its} + \beta_{it} Y_{its}) \le CI_t \qquad \forall t,s \quad (1e)$$

$$W_{its} \le Q_{its} \qquad \forall i,t,s \quad (1f)$$

$$\sum_l P_{jlts} + \sum_i \eta_{ij} W_{its} = \sum_l S_{jlts} + \sum_i \mu_{ij} W_{its} \qquad \forall j,t,s \quad (1g)$$

$$a^L_{jlts} \le P_{jlts} \le a^U_{jlts} \qquad \forall j,l,t,s \quad (1h)$$

$$d^L_{jlts} \le S_{jlts} \le d^U_{jlts} \qquad \forall j,l,t,s \quad (1i)$$

$$Y_{i,1,s} = Y1_i \qquad \forall i,s \quad (1j)$$

$$E_{i,1,s} = E1_i \qquad \forall i,s \quad (1k)$$

$$Q_{i,1,s} = Q1_i \qquad \forall i,s \quad (1l)$$

$$Y_{its} \in \{0,1\}, E_{its}, Q_{its}, W_{its}, P_{jlts}, S_{jlts} \ge 0 \quad (1m)$$

$$Y1_i \in \{0,1\}, E1_i, Q1_i \ge 0 \quad (1n)$$

In the above model, the objective (1a) is to maximize the expected NPV over the two stages of the capacity expansion project. For the sake of simplicity, the NPV is defined here as the expectation of the sales revenue minus the sum of the expectation of the capacity expansion investment, operating cost and material cost. Constraints (1b) correspond to the lower and upper bounds for the capacity expansions which also forces capacity expansion values to zero whenever the corresponding binary variable is zero. Constraints (1c) are the capacity balance equations. Constraints (1d) express the limit on the number of expansions of processes. Constraints (1e) express the capital available for investment. Constraints (1f) ensure that the operating level of a process does not exceed the installed capacity. Constraints (1g) express the material balances for each chemical, which means that the total purchase amount plus production amount should be equal to the total sale amount plus the consumption amount, where η,μ represent the production recipe and is specific to the chemical processes ($\eta_{ij}\mu_{ij} = 0$). Constraints (1h) and (1i) express the lower and upper bounds for raw material availabilities and product demands. Constraints (1j), (1k), and (1l) ensure that the first stage decisions are the same under all possible scenarios, which are also called nonanticipativity constraints. Detail information of all the variables and model parameters can be referred from the nomenclature section.

It is also worth pointing out that the following redundant constraints related to the first stage variables can also be incorporated into the model, which will be used in the proposed hybrid optimization framework.

$$Y1_i E^L_{i,1} \le E1_i \le Y1_i E^U_{i,1} \qquad \forall i \quad (1o)$$

$$Q1_i = Q_{i0} + E1_i \quad \forall i \quad (1p)$$

$$\sum_i (\alpha_{i,1} E1_i + \beta_{i,1} Y1_i) \le CI_1 \tag{1q}$$

Notice that the proposed formulation (1) is only one possible formulation for capacity expansion planning problem. Without comparing the advantage of this model to other possible stochastic programming based formulations, we are going to use the above model to illustrate the proposed algorithm, which is rather general and can be also applied to solve other stochastic programming based formulations.

Augmented Lagrangian Optimization

Algorithm

For the capacity planning problem (1), it is observed that the model takes a block angular structure and those blocks are linked through the nonanticipativity constraints (1j–1l). The model is decomposable once the constraints (1j–1l) are relaxed. In this work, augmented Lagrangian function is applied to relax those nonanticipativity constraints and the resulted ALR problem is as follows:

$$\begin{aligned} L(\lambda,\sigma) = \min & -\sum_s p_s \Bigg\{ \sum_j \sum_l \sum_t \gamma_{jlts} S_{jlts} \\ & - \sum_i \sum_t (\alpha_{it} E_{its} + \beta_{it} Y_{its}) - \sum_i \sum_t \delta_{its} W_{its} \\ & - \sum_j \sum_l \sum_t \Gamma_{jlts} P_{jlts} \Bigg\} + \sum_s \sum_i \lambda^Y_{is} (Y1_i - Y_{i,1,s}) \\ & + \sum_s \sum_i \lambda^E_{is} (E1_i - E_{i,1,s}) + \sum_s \sum_i \lambda^Q_{is} (Q1_i - Q_{i,1,s}) \\ & + \sigma \sum_s \sum_i (Y1_i - Y_{i,1,s})^2 + \sigma \sum_s \sum_i (E1_i - E_{i,1,s})^2 \\ & + \sigma \sum_s \sum_i (Q1_i - Q_{i,1,s})^2 \end{aligned} \tag{2}$$

s.t. Eqs. 1b–1i, 1m–1q where λ^Y, λ^E, λ^Q represent the Lagrangian multipliers and σ is a positive penalty parameter. To simplify the presentation of the solution algorithm, we express the above problem (2) in the following general compact form:

$$\begin{aligned} L(\lambda,\sigma) = & \min_{x\in\Omega_0, (\chi^s, y^s)\in\Omega_s} \sum_s p_s q y^s + \sum_s \lambda_s (x - \chi^s) \\ & + \sigma \sum_s (x - \chi^s)^2 \end{aligned} \tag{3}$$

where x represents the first stage decisions ($Y1_i$, $E1_i$, $Q1_i$), y^s the second stage decisions (W_{its}, P_{jlts}, $S_{jlts}(t = 1,\ldots,T)$, E_{its}, Q_{its}, Y_{its} $(t = 2,\ldots,T)$), χ^s represent the copy of first stage variables ($Y_{i,l,s}$, $E_{i,l,s}$, $Q_{i,l,s}$), λ_s represent (λ^Y, λ^E, λ^Q), $\Omega_0 = \{x|(1n) \sim (1q)\}$, $\Omega_s = \{(\chi^s, y^s)|(1b) \sim (1i), (1m)\}$, $s = 1,\ldots,S$, q represent the constant coefficients.

Based on the above definition of the relaxation problem, the augmented Lagrangian dual problem can be expressed as

$$\max_{\lambda,\sigma} L(\lambda,\sigma) \tag{4}$$

The solution of the original problem (1) can be transformed to the solution of this augmented Lagrangian dual problem (4).

Augmented Lagrangian optimization algorithm or method of multipliers can be applied to solve the above dual problem. Generally, the algorithm consists of an outer loop and an inner loop. In the outer loop, the method of multipliers is used to update the multipliers and penalty parameters in such a way to find solution of the augmented Lagrangian dual problem, while the inner loop is used to compute a solution of the ALR problem. The detail steps of the algorithm can be described as follows:

Method of Multipliers(outer loop)

Step 0: Initialization. Set $k = 1$, initial multipliers $\lambda^{(1)i} \in [\lambda_{\min}, \lambda_{\max}]$, $i = 1,\ldots,m$, penalty parameter $\sigma^{(1)} > 0$, tolerance $\varepsilon_{\text{outer}} > 0$, bounds for the multipliers $\lambda \in [\lambda_{\min}, \lambda_{\max}]$, constant $\gamma > 1$ and $0 < \tau < 1$.

Step 1: Solve the augmented Lagrangian relaxation problem (3), obtain a solution x^k which satisfies the approximate KKT conditions.

Step 2: Update multipliers' and penalty parameter's value

$$\lambda_i^{(k+1)} = \min\left\{\max\left(\lambda_{\min}, \lambda_i^{(k)} + \sigma^{(k)}\left(x^{(k)} - \chi^{s(k)}\right)_i\right), \lambda_{\max}\right\}, \tag{5}$$

if $\|x^{(k)} - \chi^{s(k)}\| \le \varepsilon_{\text{outer}}$, stop.

if $\|x^{(k)} - \chi^{s(k)}\|_\infty \le \tau \|x^{(k-1)} - \chi^{s(k-1)}\|_\infty$, set $\sigma^{(k+1)} = \sigma^{(k)}$, otherwise set $\sigma^{(k+1)} = \gamma\sigma^{(k)}$.

Set $k = k + 1$, go to step 1.

The method of multipliers has been widely studied and analyzed since it was first proposed in the 1960s.[19,20] The theoretical advantages of this approach include its natural link with penalty and Lagrangian methods and also its interpretation in connection with the proximal point algorithm as a regularized dual technique for convex programming.[21] Detail convergence properties of the above algorithm are given in the next subsection.

Convergence property

To present the convergence theorems of the above augmented Lagrangian optimization algorithm, consider the following general problem:

$$\begin{aligned} & \min f(x) \\ & \text{s.t. } h(x) = 0, \ \ g(x) \le 0, \\ & \qquad x \in \Omega = \{x | H(x) = 0, G(x) \le 0\} \end{aligned} \quad , \tag{6}$$

and its augmented Lagrangian dual problem generated by relaxing the upper level constraints $h(x) = 0$, $g(x) \ge 0$:

$$\begin{aligned} \max_{\lambda,v\ge 0,\sigma} \min_{x\in\Omega} L(x,\lambda,v,\sigma) = & f(x) + \frac{\sigma}{2} \sum_i \left(h_i(x) + \frac{\lambda_i}{\sigma}\right)^2 \\ & + \frac{\sigma}{2} \sum_j \left(g_j(x) + \frac{v_j}{\sigma}\right)_+^2, \end{aligned} \tag{7}$$

where λ, v are the Lagrangian multipliers, σ is the penalty parameter and $(A)_+ = \max\{A, 0\}$.

Assume all these functions admit continuous first derivatives on a sufficiently large and open domain. Different definitions

of augmented Lagrangian with respect to first group of constraints ($h(x) = 0$, $g(x) \geq 0$) of (6) are as follows:

$$L_1(x,y,\lambda,\mu,\sigma) = f + \sum_i \lambda_i h_i + \frac{\sigma}{2}\sum_i h_i^2 + \sum_j \mu_j(g_j + y_j^2) + \frac{\sigma}{2}\sum_j (g_j + y_j^2)^2$$

$$L_2(x,y,\lambda,\mu,\sigma) = f + \frac{\sigma}{2}\sum_i \left(h_i + \frac{\lambda_i}{\sigma}\right)^2 + \frac{\sigma}{2}\sum_j \left(g_j + y_j^2 + \frac{\mu_j}{\sigma}\right)^2$$

$$L_3(x,\lambda,\mu,\sigma) = f + \frac{\sigma}{2}\sum_i \left(h_i + \frac{\lambda_i}{\sigma}\right)^2 + \frac{\sigma}{2}\sum_j \left(g_j + \frac{\mu_j}{\sigma}\right)_+^2,$$

$$\left(g_j + \frac{\mu_j}{\sigma}\right)_+ = \max\left\{g_j + \frac{\mu_j}{\sigma}, 0\right\}$$

First, notice that the definitions L_1 and L_2 are equivalent because the difference between them is a constant term, thus the same solution x^* can be achieved when they are minimized. Furthermore, to minimize L_2, we have the following conclusion:

(a) if $g_j + \frac{\mu_j}{\sigma} \geq 0$, we must have $y_j^* = 0$, then $L_2 = f + \frac{\sigma}{2}\sum_i \left(h_i + \frac{\lambda_i}{\sigma}\right)^2 + \frac{\sigma}{2}\sum_j \left(g_j + \frac{\mu_j}{\sigma}\right)^2 = L_3$

(b) if $g_{1j} + \frac{\mu_j}{\sigma} \leq 0$, we must have $y_j^{*2} = -(g_j + \frac{\mu_j}{\sigma})$, then $L_2 = f + \frac{\sigma}{2}\sum_i \left(h_i + \frac{\lambda_i}{\sigma}\right)^2 = L_3$

So, L_2 and L_3 are also equivalent. Thus all the above three definitions are equivalent.

In the article, problem (3) is taking the form of L_1 (notice that $g(x) \leq 0$ does not appear in our case), and problem (7) is in the form of L_3, but they are equivalent description of augmented Lagrangian.

The convergence of the augmented Lagrangian optimization algorithm is shown using the following theorems. The details of the proofs are not shown but the reader is referred to the literature where proofs are given.[22,23]

Theorem 1. The solution sequence $\{x^{(k)}\}$ generated by the augmented Lagrangian algorithm (method of multipliers) admits a limit point x^* if the solution $x^{(k)}$ satisfy approximate KKT conditions for $\lim_{k\to\infty} \varepsilon^k = 0$:

a)

$$\left\|\nabla L(x^{(k)}, \lambda^{(k)}, v^{(k)}, \sigma^{(k)}) + v^{(k)}\nabla H(x^{(k)}) + u^{(k)}\nabla G(x^{(k)})\right\| \leq \varepsilon^k$$

b) $G_j(x^{(k)}) \leq \varepsilon^k, u_j^{(k)} \geq 0, \ \forall j$

c) if $G_j(x^{(k)}) < -\varepsilon^k$, then $u_j^k = 0, \ \forall j$

d) $\left\|H_i(x^k)\right\| \leq \varepsilon^k, \ \forall i$

It is proved[22] that under the conditions that *f, g, h* admit continuous first derivatives and $\Omega = \{x | H(x) = 0, G(x) \leq 0\}$ is a closed set, a limit point x^* of the sequence $\{x^k\}$ generated by the augmented Lagrangian optimization algorithm exists under the sufficient condition that there exists $\varepsilon > 0$ such that the set $\{x \mid \| H(x) \| < \varepsilon, G(x) < \varepsilon\}$ is bounded.

It is easy to prove that the above theorem is also valid for the problem (1) due to the following facts. First, problem (1) can be written in the form of (6), where $h(x) = 0$ represents the nonanticipativity constraints (1j)–(1l), $g(x) \leq 0$ does not appear and $\Omega = \{(1b) - (1i), 1(m) - 1(q)\}$. Notice that the augmented Lagrangian optimization framework is still valid for the case that both inequality and equality constraints exists (inequality can also be transformed into equality constraints with slack variables). Second, the mixed integer problem can be always transformed into its equivalent continuous counterpart because any binary variable $x \in \{0,1\}$ can be always replaced by its continuous relaxation $0 \geq x \geq 1$ and the complementarity constraints $x(1 - x) = 0$. Furthermore, Ω is bounded in problem (1). So when the algorithm is applied onto the mixed integer programming problem (1), the convergence property of Theorem 1 can still be ensured. Similarly, the following theorem also applies to problem (1).

Theorem 2. The limit point x^* is a feasible solution of the original problem (6).

Regarding the feasibility of the solution, it is proved[22] that if the sequence of penalty parameters $\{\sigma^k\}$ is bounded (i.e., from some iteration on, the penalty parameters are not updated, or there exists k_0 such that $\sigma^k = \sigma^{k_0}$ for all $k \geq k_0$), the limit point x^* is a feasible solution of problem (6). Furthermore, in terms of the optimality of the solution, it is proved that if the limit point x^* is feasible and also satisfies the Constant Positive Linear Dependence constraint qualification condition[24] with respect to lower level constraints $\{x \in \Omega\}$, then x^* is a KKT (stationary) point of the original problem (6).

To ensure convergence for the method of multipliers, it is proved[22] that if approximate KKT conditions (defined by conditions in Theorem 1 and a tolerance ε^k) is satisfied and $\lim_{k\to\infty} \varepsilon^k = 0$, the convergence is ensured. So, in the performed calculations, we tightened the stopping criterion (decreasing its value) for the solution of the KKT point to satisfy the requirement $\varepsilon^k \to 0$.

Solution of the Augmented Lagrangian Relaxation Problem

The major step in the augmented Lagrangian optimization algorithm is to solve the ALR problem (3). Obviously, this problem involves large number of constraints and integer variables and may still be hard to solve directly unless it is decomposed. Since the quadratic term in the objective function $(x - \chi^s)^2$ is not separable, different strategies have been proposed in the literature as presented in the follows.

Diagonal quadratic approximation (DQA)

Diagonal quadratic approximation was originally proposed by Ruszczynski[16] to solve block angular structure problems. The cross-product term $x\ \chi^s$ can be replaced with its first

order approximation around a certain reference point ($\bar{x}^{(n)}$, $\bar{\chi}^{s(n)}$) (where n represents the iteration index of the inner loop for the solution of relaxation problem) as follows:

$$x\chi^s \approx x\bar{\chi}^{s(n)} + \bar{x}^{(n)}\chi^s - \bar{x}^{(n)}\bar{\chi}^{s(n)} \tag{8}$$

then the term $(x - \chi^s)^2$ is approximated as follows:

$$(x-\chi^s)^2 = x^2 + (\chi^s)^2 - 2x\chi^s \approx (x-\bar{\chi}^{s(n)})^2 + (\bar{x}^{(n)} - \chi^s)^2 - (\bar{x}^{(n)} - \bar{\chi}^{s(n)})^2 \tag{9}$$

and correspondingly, the original ALR problem is approximated by:

$$\min_{x\in\Omega_0,(\chi^s,y^s)\in\Omega_s} \sum_s p_s q y^s + \sum_s \lambda_s^T(x-\chi^s) + \sigma\sum_s \left((x-\bar{\chi}^{s(n)})^2 + (\bar{x}^{(n)} - \chi^s)^2 - (\bar{x}^{(n)} - \bar{\chi}^{s(n)})^2\right) \tag{10}$$

which is decomposable. Steps of the DQA method for solving the above ALR problem is as follows:

DQA method for the solution of the augmented Lagrangian relaxation problem (inner loop)

Step 0: Set $n = 1$, and set initial value for reference point $\bar{x}^{(1)}, \bar{\chi}^{s(1)}, s = 1, \ldots, S$. It is generally set as the value obtained from last outer loop iteration. Set tolerance $\varepsilon_{\text{inner}}$, inner iteration limit N.

Step 1: Solve the following decomposed problem

$$x^{(n)} := \arg\min_{x\in\Omega_0} \sum_s \lambda_s^T x + \sigma\sum_s (x-\bar{\chi}^{s(n)})^2 - \sigma\sum_s (\bar{x}^{(n)} - \bar{\chi}^{s(n)})^2, \tag{11}$$

$$\chi^{s(n)} := \arg\min_{(\chi^s,y^s)\in\Omega_s} p_s q y^s - \lambda_s^T\chi^s + \sigma(\bar{x}^{(n)} - \chi^s)^2, \quad s = 1, \ldots, S \tag{12}$$

Step 2: With some step size $\tau \in (0,1)$, the reference point is updated as follows

$$\bar{x}^{(n+1)} = \bar{x}^{(n)} + \tau(x^{(n)} - \bar{x}^{(n)}), \tag{13}$$

$$\bar{\chi}^{s(n+1)} = \bar{\chi}^{s(n)} + \tau(\chi^{s(n)} - \bar{\chi}^{s(n)}), \qquad s = 1, \ldots, S \tag{14}$$

Step 3: Set $n = n + 1$, go to step 1 until $\|x^{(n)} - x^{(n-1)}\| \le \varepsilon_{\text{inner}}$ or $n > N$.

The subproblems (11) and (12) are mixed integer quadratic programming (MIQP) problems. Problems (12) can be solved in parallel to improve computation efficiency. The inner loop is terminated when the relative change in the solution of the relaxation problem for two consecutive inner loop iterations is smaller than some user-defined termination tolerance or given iteration number limit reaches.

Auxiliary problem principle (APP)

Auxiliary problem principle was introduced by Cohen[17] in the context of the decomposition of the optimization problems, it allows to determine the solution of a problem by solving a sequence of auxiliary problems. Informally, the APP method involves linearizing the inseparable quadratic terms in the Augmented Lagrangian and adding convex/concave terms which can be chosen to be separable. The chief utility of the added term is that it regularizes the problem in the sense that it makes the partially linearized augmented Lagrangian strictly convex. Thus, first order approximation of the whole quadratic term is made first around a certain reference point ($\bar{x}^{(n)}$, $\bar{\chi}^{s(n)}$)

$$(x-\chi^s)^2 \approx -(\bar{x}^{(n)} - \bar{\chi}^{s(n)})^2 + 2(x-\chi^s)(\bar{x}^{(n)} - \bar{\chi}^{s(n)}) \tag{15}$$

and furthermore a regularization term $\gamma\,(x - \bar{x}^{(n)})^2 + \gamma(\chi^s - \bar{\chi}^{s(n)})^2$ is added so that the term $(x-\chi)^2$ is approximated by the following:

$$\gamma(x-\bar{x}^{(n)})^2 + \gamma(\chi^s - \bar{\chi}^{s(n)})^2 - (\bar{x}^{(n)} - \bar{\chi}^{s(n)})^2 + 2(x-\chi^s)(\bar{x}^{(n)} - \bar{\chi}^{s(n)}) \tag{16}$$

and the ALR problem becomes:

$$\min_{x\in\Omega_0,(\chi^s,y^s)\in\Omega_s,\forall s} \sum_s p_s q y^s + \sum_s \lambda_s^T(x-\chi^s) + \sigma\gamma\sum_s (x-\bar{x}^{(n)})^2 + \sigma\gamma\sum_s (\chi^s - \bar{\chi}^{s(n)})^2 - \sigma\sum_s (\bar{x}^{(n)} - \bar{\chi}^{s(n)})^2 + 2\sigma\sum_s (x-\chi^s)(\bar{x}^{(n)} - \bar{\chi}^{s(n)}) \tag{17}$$

Then the APP method for solving the above ALR problem can be expressed as the following:

APP method for the solution of augmented Lagrangian relaxation problem (inner loop)

Step 0: Set $n = 1$, and set initial value for reference point $\bar{x}^{(1)}, \bar{\chi}^{s(1)}, s = 1, \ldots, S$. Set tolerance $\varepsilon_{\text{inner}}$, inner iteration limit N.

Step 1: Solve the following decomposed problem

$$x^{(n)} := \arg\min_{x\in\Omega_0} \sum_s \lambda_s^T x + \sigma\gamma\sum_s (x-\bar{x}^{(n)})^2 - \sigma\sum_s (\bar{x}^{(n)} - \bar{\chi}^{s(n)})^2 + 2\sigma x\sum_s (\bar{x}^{(n)} - \bar{\chi}^{s(n)}), \tag{18}$$

$$\chi^{s(n)} := \arg\min_{(\chi^s,y^s)\in\Omega_s} p_s q y^s - \lambda_s^T\chi^s + \sigma\gamma(\chi^s - \bar{\chi}^{s(n)})^2 - 2\sigma\chi^s(\bar{x}^{(n)} - \bar{\chi}^{s(n)}), \qquad s = 1, \ldots, S \tag{19}$$

Step 2: With some step size $\tau \in (0,1)$, the reference point is updated as follows

$$\bar{x}^{(n+1)} = \bar{x}^{(n)} + \tau(x^{(n)} - \bar{x}^{(n)}), \tag{20}$$

$$\bar{\chi}^{s(n+1)} = \bar{\chi}^{s(n)} + \tau(\chi^{s(n)} - \bar{\chi}^{s(n)}), \qquad s = 1, \ldots, S \tag{21}$$

Step 3: Set $n = n + 1$, go to step 1 until $x^{(n)} - x^{(n-1)}\| \le \varepsilon_{\text{inner}}$ or $n > N$.

The APP method is similar to the DQA method. Actually, it is easy to prove that DQA method is a special case of APP method when $\gamma = 1$ because the differences between (11) and (18), (12) and (19) are constant in this case. Furthermore, for the APP method, convergence is guaranteed when $\gamma \ge 1$.[17] The subproblems (18) and (19) are both

MIQP problems with convex quadratic objective function and linear constraints.

Block coordinate descent (BCD)

Instead of solving the relaxation problem (3) as a whole, the BCD algorithm, also known as the "nonlinear Gauss-Seidel" method,[18] splits variables into two subsets (x and χ^s) and iterates between solving the relaxed problem for a subset of variables, while holding the remaining variables fixed at their latest value. The detailed algorithm can be expressed as the following:

BCD method for the solution of augmented Lagrangian relaxation problem (inner loop)

Step 0: Set $n = 1$, initial value $x^{(1)}$, $\chi^{s(1)}$, $s = 1,\ldots,S$. Set tolerance $\varepsilon_{\text{inner}}$, inner iteration limit N.

Step 1: Solve the following decomposed problem

$$x^{(n+1)} := \arg\min_{x\in\Omega_0} \sum_s \lambda_s^T x + \sigma \sum_s (x - \chi^{s(n)})^2 \tag{22}$$

$$\chi^{s(n+1)} := \arg\min_{(\chi^s,y^s)\in\Omega_s} p_s q y^s - \lambda_s^T \chi^s + \sigma(x^{(n+1)} - \chi^s)^2, \quad s = 1,\ldots,S \tag{23}$$

Step 2: Set $n = n + 1$, go to step 1 until $\|x^{(n)} - x^{(n-1)}\| \le \varepsilon_{\text{inner}}$ or $n > N$.

In practical applications, there exists an extreme case of the BCD method, which is also called alternating direction,[18] where the inner loop is terminated after just a single BCD iteration. Again, the subproblem (22) and (23) are both MIQP problems with convex quadratic objective function and can be solved using solver such as CPLEX. For the case that the objective function is quadratic, the convergence of the BCD method to a stationary point of the ALR problem (2) is ensured.[25]

Hybrid Solution Strategy

The solution methods presented in the previous section can be incorporated into the augmented Lagrangian optimization framework and feasible solution of the original problem can be finally obtained (as pointed out in Theorem 2). In the follows, a new method for the solution of the relaxation problem is proposed, which will be further incorporated into a two phase hybrid solution framework.

First, the ALR problem (3) can be rewritten as follows:

$$\min_{x\in\Omega_0}\left\{\min_{(\chi^s,y^s)\in\Omega_s,\forall s} \sum_s p_s q y^s + \sum_s \lambda_s^T (x - \chi^s) + \sigma \sum_s (x - \chi^s)^2\right\} \tag{24}$$

Using the following definition:

$$f_s(x) = \min_{(\chi^s,y^s)\in\Omega_s} p_s q y^s + \lambda_s^T (x - \chi^s) + \sigma(x - \chi^s)^2 \quad s = 1,\ldots,S \tag{25}$$

the problem (24) can be modified as follows:

$$z^* = \min_{x\in\Omega_0} \sum_s f_s(x) \tag{26}$$

Thus, the solution of the original relaxation problem (3) is transformed to the solution of problem (26), which includes implicit terms $f_s(x)$ in the objective. Since problem (25) is a MIQP problem, the optimal value function $f_s(x)$ is in general a piecewise quadratic nonconvex function.

We consider the problem (26) as a costly mixed integer nonlinear optimization problem, where the costly component is only the objective function since the evaluation of the objective function involves the solution of a set of mixed integer quadratic optimization subproblems (25). Problems (25) has convex quadratic objective function and all those sub-problems can be solved in parallel using MIQP solver such as CPLEX, thus the computation efficiency can be greatly improved.

Since we have incorporated copy variables for the first stage decision variables and applied ALR, solving problem (26) has the advantage that feasibility of subproblem (25) is always ensured since x is only involved in the objective function of (25). Notice that although a direct reformulation of original model (1) can also result in a similar decomposition, the feasibility of the subproblem in that case cannot be ensured since the variable x will appear in the constraints of the decomposed subproblem.

Based on the above methods for the solution of the ALR problem, the following hybrid solution framework for the original SIP problem (1) is proposed:

Hybrid Optimization Framework

(first phase)

Step 0: Initialization. Set $k = 1$, initial multipliers $\lambda_i^{(1)} \in [\lambda_{\min}, \lambda_{\max}]$, $i = 1,\ldots m$, penalty parameter $\sigma^{(1)} > 0$, tolerance $\varepsilon_{\text{outer}} > 0$, bounds for the multipliers $\lambda \in [\lambda_{\min}, \lambda_{\max}]$, constant $\gamma > 1$ and $0 < \tau < 1$.

Step 1: Apply traditional strategy (e.g., BCD) to solve the augmented Lagrangian relaxation problem (3) and obtain solution x^k.

Step 2: Update the multipliers and penalty parameters as follows:

$$\lambda_i^{(k+1)} = \min\left\{\max\left(\lambda_{\min}, \lambda_i^{(k)} + \sigma^{(k)}(x^{(k)} - \chi^{s(k)})_i\right), \lambda_{\max}\right\},$$

if $\|x^{(k)} - \chi^{s(k)}\|_\infty \le \tau\|x^{(k-1)} - \chi^{s(k-1)}\|_\infty$, set $\sigma^{(k+1)} = \sigma^{(k)}$, otherwise set $\sigma^{(k+1)} = \gamma\sigma^{(k)}$.

Set $err = \|x^{(k)} - \chi^{s(k)}\|$, $k = k + 1$.

If $err \le \varepsilon_{\text{outer}}$, go to step 1, otherwise, go to step 3.

(second phase)

Step 3: Apply global optimization solver to solve the reformulated augmented Lagrangian relaxation problem (26).

Step 4: Update the multipliers and penalty parameters as follows:

$$\lambda_i^{(k+1)} = \min\left\{\max\left(\lambda_{\min}, \lambda_i^{(k)} + \sigma^{(k)}(x^{(k)} - \chi^{s(k)})_i\right), \lambda_{\max}\right\},$$

if $\|x^{(k)} - \chi^{s(k)}\|_\infty \le \tau\|x^{(k-1)}\|_\infty$, set $\sigma^{(k+1)} = \sigma^{(k)}$, otherwise set $\sigma^{(k+1)} = \gamma\sigma^{(k)}$.

Set $err = \|x^{(k)} - \chi^{s(k)}\|$, $k = k + 1$.

If $err \le \varepsilon_{\text{outer}}$, return the solution of step 3 as final solution for problem (1) and stop, otherwise, go to step 3.

The above algorithm consists of two phases: in the first phase, the ALR problem is solved using the traditional

Table 1. Optimal Solution of the Numerical Example

S	ξ_i	p_s	f^*	x^*
9	{5,10,15}	1/9	−59.33	(0, 2)
36	{5,7,9,11,13,15}	1/36	−61.22	(0, 4)
121	{5,6,7,8,9,10,11,12,13,14,15}	1/121	−62.29	(0, 4)

S: number of scenarios; ξ_i: possible value of the uncertain parameter; p_s: probability of each scenario; f^*: optimal objective value; x^*: optimal solution of (x_1, x_2).

decomposition method and a feasible solution for original problem (1) can be obtained; the second phase of above hybrid solution framework will improve the solution toward the optimal solution of the original problem (1). This is proved[23] and can be described with the following theorem 3. Notice that since the 2nd phase global optimization takes relative more time than the 1st phase, the objective of 1st phase is to provide good initial multipliers' values to avoid spending too much time on the second phase.

Theorem 3. For problem (6), if in each outer loop iteration of the augmented Lagrangian optimization algorithm, an ε_k-global minimization of the relaxation problems is found, where $\varepsilon_k \to \varepsilon$, then the convergence to ε-global minimum of the original problem is ensured.

In this article, solver *glcSolve* from TOMLAB package[26] is applied to solve the ALR problem in the second phase of the above algorithm. Solver *glcSolve* is specially developed for costly global mixed integer nonlinear programming problem. It is an extended implementation of a sampling algorithm DIRECT.[27] This global optimization algorithm begins by scaling the design box to an n-dimensional unit hypercube. The search is initiated by evaluating the objective function at the center point of the hypercube and then dividing the potentially optimal hyper-rectangles by sampling the longest coordinate directions of the hyper-rectangle. The sampling is done such that each sampled point becomes the center of its own n-dimensional rectangle or box. This division continues until termination (i.e., when a pre-specified iteration limit is reached) or when convergence is achieved.

A numerical example is used next to illustrate the basic steps of the proposed solution approach.

Numerical example

Consider the following numerical example which corresponds to the deterministic equivalent of a simple two stage SIP problem with discrete distribution of uncertainty on the right hand side of the constraints:

$$\min -1.5x_1 - 4x_2 + \sum_s p_s(-16y_1^s - 19y_2^s - 23y_3^s - 28y_4^s)$$

$$\text{s.t. } 0 \le x_1, x_2 \le 5, x_1 \in Z, \quad y \in \{0,1\}$$
$$2y_1^s + 3y_2^s + 4y_3^s + 5y_4^s \le \xi_1^s - x_1 \ \ \forall s$$
$$6y_1^s + y_2^s + 3y_3^s + 2y_4^s \le \xi_2^s - x_2 \ \ \forall s$$

In this problem, first stage variable x_1 is integer variable and x_2 is continuous. Uncertain parameter ξ_1 and ξ_2 are independent, they can take the same set of possible value as shown in Table 1. The optimal objective and first stage solution for the above problem under three cases with different uncertainty scenarios are shown in Table 1.

The proposed hybrid solution framework is used to solve the above problem. We list the details of the solution procedure for the 9-scenario case problem in Table 2 where BCD method is used in the first phase, and *glcSolve* is applied in the second phase. It can be observed that as the iterations proceed, the feasibility of the nonanticipativity constraints are gradually satisfied and finally a feasible solution of the original problem is generated at the end of the first phase since the norm of the nonanticipativity constraints' value satisfies the outer loop tolerance 0.01 which is very close to zero; whereas in the second phase, the solution from the first phase (including the Lagrangian multipliers, penalty parameter, the first stage variables) provide good initial estimation to be used in the second phase which then requires a few more steps to converge to the optimal solution.

Table 3 illustrates the computational results using a PC with 2.3 GHz CPU and 2Gb RAM (same for all the other computations in this article) for all the different number of uncertainty scenarios. In the first phase, different decomposition strategies for the ALR problem are also tested.

From the results shown in Table 3, it is observed that a feasible solution for the original problem is reached for all the methods during the first phase, and in the second phase, the optimal solution of the original problem is obtained. Comparing the three different methods, APP method takes relatively lessl time in the second phase. While the feasible solution generated from DQA method is in general better than the other two, APP method generates the feasible solution with the worst objective. However, for all the three solutions, the initial solution provided by the first phase leads to optimal solution by the second stage.

Optimization of the Capacity Planning Problem

Problem reformulation

In this section, the augmented Lagrangian algorithm is applied to solve the capacity expansion planning problem.

To apply the proposed hybrid solution strategy, in the first phase of the algorithm, traditional decomposition strategy (e.g., BCD) will be applied to solve the relaxation problem

Table 2. Solution Procedure of Hybrid Method for 9-scenario Case

	k	$\|\|x-X\|\|$	$\lambda(x-X)+\sigma\|\|x-X\|\|^2$	f^{ALR}	x^*	CPU sec
1st phase	1	1.429	2.043	−51.68	(3, 2.72)	84
	2	0.913	0.627	−50.69	(3, 2.70)	
	3	0.484	0.637	−49.83	(3, 2.64)	
	4	0.219	0.197	−50.78	(3, 2.85)	
	5	0.075	0.038	−51.25	(3, 3.03)	
	6	0.151	0.232	−51.35	(3, 3.10)	
	7	0.126	0.089	−50.47	(3, 3.05)	
	8	0.129	0.402	−50.80	(3, 3.09)	
	9	0.004	0.005	−50.39	(3, 3.00)	
2nd phase	10	0.071	−0.118	−59.23	(0, 1.94)	225
	11	0.036	−0.041	−59.28	(0, 1.98)	
	12	0.008	−0.0001	−59.33	(0, 2.00)	

k: outer loop iteration index; $\|\|x-X\|\|$: norm value of the nonanticipativity constraints function vector; $\lambda(x-X)+\sigma\|\|x-X\|\|^2$: value of the penalty term in the objective of the ALR problem; f^{ALR}: objective value of the ALR problem; x^*: solution of first stage variables.

Table 3. Solution of the Numerical Example

		DQA			APP			BCD		
	S	9	36	121	9	36	121	9	36	121
1st phase	K	12	11	15	10	10	13	9	8	11
	t	63	261	701	73	275	853	84	277	829
	f^*	−51.8	−54.2	−54.5	−50.2	−43.7	−38.7	−50.39	−53.3	−53.3
	x^*	(2,2)	(2,2)	(2,2)	(1,2)	(1,2)	(1,2)	(3,3)	(3,4)	(3,3)
2nd phase	K	3	5	3	2	4	3	3	2	3
	t	157	468	853	130	364	698	225	558	916
	f^*	−59.3	−61.2	−62.3	−59.4	−61.2	−62.3	−59.3	−61.2	−62.3
	x^*	(0,2)	(0,4)	(0,4)	(0,2)	(0,4)	(0,4)	(0,2)	(0,4)	(0,4)

S: number of scenarios; K: number of outer loop iterations; t: computation time (sec); f^*: objective value of the ALR problem; x^*: solution of first stage variables; $\varepsilon_{\text{outer}} = 0.01$, $\varepsilon_{\text{inner}} = 0.01$, $\tau = 0.5$, $\gamma = 2$, $\tau = 0.5$.

(3). For example, for the BCD method the following subproblems are defined:

$$\min_{Y,E,Q,W,P,S} -p_s\left\{\sum_j\sum_l\sum_t \gamma_{jlts}S_{jlts} - \sum_i\sum_t(\alpha_{it}E_{its}+\beta_{it}Y_{its}) - \sum_i\sum_t \delta_{its}W_{its} - \sum_j\sum_l\sum_t \Gamma_{jlts}P_{jlts}\right\} + \sum_i \lambda^Y_{is}(\overline{Y1}_i - Y_{i,1,s}) + \sum_i \lambda^E_{is}(\overline{E1}_i - E_{i,1,s}) + \sum_i \lambda^Q_{is}(\overline{Q1}_i - Q_{i,1,s}) + \sigma\sum_i (\overline{Y1}_i - Y_{i,1,s})^2 + \sigma\sum_i (\overline{E1}_i - E_{i,1,s})^2 + \sigma\sum_i (\overline{Q1}_i - Q_{i,1,s})^2 \quad (27)$$

s.t. (1b)–(1i), (1m) and

$$\min_{Y1,E1,Q1} \sum_s\sum_i \lambda^Y_{is}(Y1_i - \overline{Y_{i,1,s}}) + \sum_s\sum_i \lambda^E_{is}(E1_i - \overline{E_{i,1,s}}) + \sum_s\sum_i \lambda^Q_{is}(Q1_i - \overline{Q_{i,1,s}}) + \sigma\sum_s\sum_i (Y1_i - \overline{Y_{i,1,s}})^2 + \sigma\sum_s\sum_i (E1_i - \overline{E_{i,1,s}})^2 + \sigma\sum_s\sum_i (Q1_i - \overline{Q_{i,1,s}})^2 \quad (28)$$

s.t. (1n)–(1q)

Problem (27) represents a set of MIQP problems which can be solved in parallel where $\overline{Y1}$, $\overline{E1}$, $\overline{Q1}$ represent fixed values of first stage decision variables $Y1$, $E1$, $Q1$. Problem (28) represents a MIQP problem where $\overline{Y_{i,1,s}}$, $\overline{E_{i,1,s}}$, $\overline{Q_{i,1,s}}$ represent fixed value of $Y_{i,1,s}$, $E_{i,1,s}$, $Q_{i,1,s}$.

To apply the second phase algorithm of the hybrid solution strategy, the relaxed capacity planning problem (2) can be formulated as the following problem:

$$\min_{Y1,E1,Q1} \sum_s f_s(Y1,E1,Q1) \quad (29)$$

s.t. (1n)–(1q)

where:

$$f_s(Y1,E1,Q1) = \min_{Y,E,Q,W,P,S} -p_s\left\{\sum_j\sum_l\sum_t \gamma_{jlts}S_{jlts} - \sum_i\sum_t(\alpha_{it}E_{its}+\beta_{it}Y_{its}) - \sum_i\sum_t \delta_{its}W_{its} - \sum_j\sum_l\sum_t \Gamma_{jlts}P_{jlts}\right\} + \sum_i \lambda^Y_{is}(Y1_i - Y_{i,1,s}) + \sum_i \lambda^E_{is}(E1_i - E_{i,1,s}) + \sum_i \lambda^Q_{is}(Q1_i - Q_{i,1,s}) + \sigma\sum_i (Y1_i - Y_{i,1,s})^2 + \sigma\sum_i (E1_i - E_{i,1,s})^2 + \sigma\sum_i (Q1_i - Q_{i,1,s})^2 \quad (30)$$

s.t. (1b)–(1i), (1m)

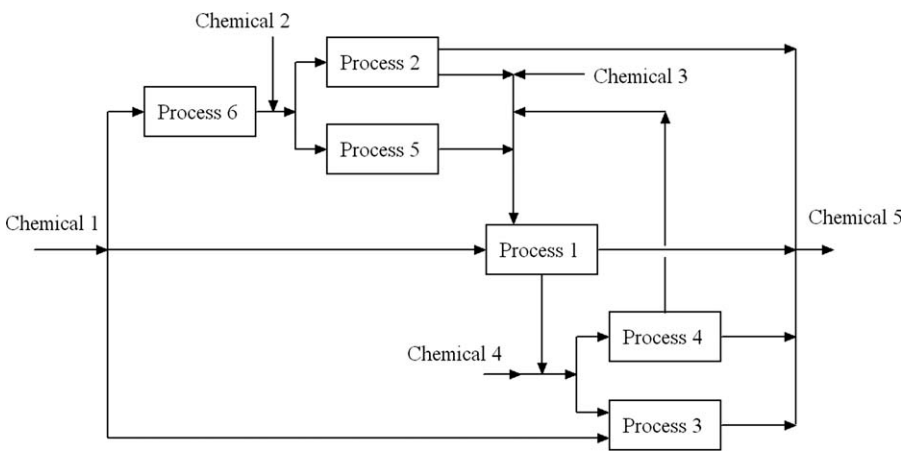

Figure 2. Process network of example 1.

Table 4. Possible Discrete Value of the Uncertain Parameters for Example 1

Upper Bound of Demand of in 1st Period	Upper Bound of Demand in 2nd Period
{60,62,64,66,68,70,72, 74,76,78,80,82,84, 86,88}	{10,12,14,16,18,20,22,24,26, 28,30,32,34,36,38}

Problem (29) is a mixed integer nonlinear optimization problem with nonlinear objective function. Notice that redundant constraints (1o)–(1q) are utilized in this problem to reduce the solution space. Problem (30) represents a set of MIQP problems which can be solved in parallel when the functions f_s ($Y1$, $E1$, $Q1$) are evaluated. Finally, notice that the subproblems (27), (28), and (30) are MIQP problems with convex quadratic objective function and they can be solved using solver such as CPLEX.

In the following subsection, two chemical process capacity planning examples are studied. The examples have been studied by Liu and Sahinidis.[28] Detail process data can be found in the appendices A and B.

Case studies

Example 1

In the example process network as shown in Figure 2, one product (chemical 5) is produced; raw materials (chemicals 1–4) can be bought from market or produced as intermediate product. There are no processes initially installed. Three time periods are considered each of 2, 3, 3 years length, respectively. The limit on the number of expansions is 2 for each process. The capital limits for each time period are $\$558 \times 10^5$, $\$708 \times 10^5$, $\$823 \times 10^5$, respectively. The upper bound of capacity expansions is 100 kton/year for each process in each time period.

The uncertain parameters considered in this example are the upper bound of the demand of chemical 5 in the first and second time period. It is assumed that the two uncertain parameters follow independent uniform distribution in the ranges [60, 88] and [10, 38], respectively, and 15 discrete values are considered for each of them as shown in Table 4. For this case, the resulted number of uncertainty scenarios is 225 and each of them takes a probability of 1/225. This leads to the SIP model (1) with 4050 binary variables, 18935 continuous variables and 29701 constraints. As a comparison, the expected value problem (EVP) (which is obtained by replacing each random parameter with its expectation) has only 18 binary variables, 119 continuous variables and 133 constraints. The solution of the EVP model is $\$4341 \times 10^5$.

Table 5. Capacity Expansion Solution of Example 1

		Hybrid Method	
	Direct Solution	First Phase	Second Phase
Objective	-4067×10^5	-2764×10^5	-4067×10^5
CPU sec	74	12	6130
Solution	$Y1_1 = 1$, $E1_1 = 9462$	$Y1_3 = 1$, $E3_1 = 3354$	$Y1_1 = 1$, $E1_1 = 9462$

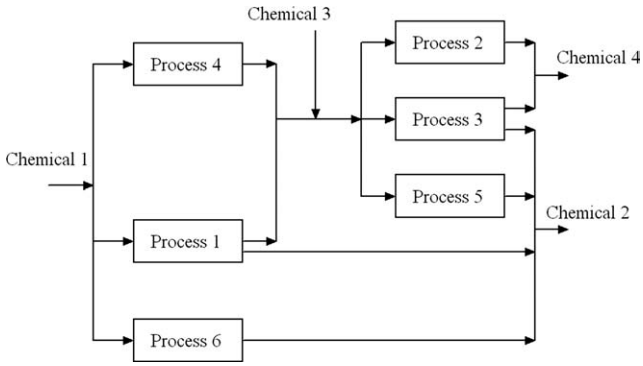

Figure 3. Process network of capacity expansion example.

Considering the two-stage stochastic optimization problem, its deterministic equivalent model is solved directly first, which results in a NPV of $\$4067 \times 10^5$, and the capacity planning solution is shown in Table 5. Note that this NPV is about 6% lower than the solution of the EVP problem, which indicates a trade-off between profit and feasibility in addressing the future uncertainties. To avoid solving the stochastic programming model (1) with a large number of binary variables, the augmented Lagrangian optimization method is applied next. Using the proposed hybrid method for the solution of the relaxation problem, the first phase of the augmented Lagrangian method converges in two iterations within only 12 CPU sec. The solution of the second phase takes longer (6130 CPU sec) and arrives at the optimal solution after two more iterations.

Example 2

In this example the process network shown in Figure 3 is considered. Two products (chemicals 2 and 4) are produced using two raw materials (chemicals 1 and 3). There are no processes initially installed. Three time periods are considered each of 2, 2, and 3 years, respectively. The number of capacity expansions is limited to two for each process. The capital limits for each time period are $\$892 \times 10^5$, $\$446 \times 10^5$, $\$975 \times 10^5$, respectively. The upper bound of capacity expansions is 100 kton/year for each process and each time period.

The uncertain parameters considered in this study are the availabilities of chemicals 1 and 3 and the demands of chemicals 2 and 4 of the first time period. It is assumed that the uncertainty is of up to 50% of original nominal values given in Table B2. It is also assumed that the four uncertain parameters are independent and each of them can take several values in different levels with the same probability.

First, we study the case where uncertainty is described as shown in Table 6 and a total of 81 scenarios are considered and each of them takes the probability of 1/81. The solution

Table 6. Possible Discrete Value of the Uncertain Parameters for Example 2

Upper Bound of Availability in 1st Period		Upper Bound of Demand in 1st Period	
Chemical 1	Chemical 3	Chemical 2	Chemical 4
{26, 53, 80}	{25, 50, 75}	{26, 52, 78}	{30, 61, 91}

Table 7. Solution of Example 2 with 81 Scenarios

	Direct Solution	Hybrid Method	
		First Phase	Second Phase
Objective	$-15{,}047 \times 10^5$	$-12{,}273 \times 10^5$	$-15{,}047 \times 10^5$
CPU sec	36	134	7010
Solution	$Y1_3 = 1$, $E1_3 = 28{,}546$	$Y1_3 = 1$, $E1_3 = 12{,}273$	$Y1_3 = 1$, $E1_3 = 28{,}546$

Table 8. Model Statistics for Example 2

Levels	Scenarios	Binary/Continuous/Constraints
1	1	18/105/136
2	16	288/1275/2161
3	81	1458/6345/10,936
4	256	4608/19,995/34,561
5	625	11,205/48,777/84,376

of the EVP model is $\$15{,}203 \times 10^5$. The solutions of the two-stage stochastic model using direct solution method and the proposed hybrid method are both presented in Table 7. It can be seen again that the EVP model results in an optimistic decision comparing to the two-stage stochastic model.

Furthermore, cases of Example 2 with different uncertainty scenarios are solved to study the proposed solution method. Table 8 describes the size of the stochastic model under different cases. It can be observed that the stochastic problem becomes intractable as the number of uncertainty scenarios increases (with CPLEX 10 solver in GAMS and 2 h resource limits, no feasible solution was obtained).

From the results shown in Table 9, it is observed that the augmented Lagrangian algorithm has the capability in generating both the feasible and optimal solution for the SIP model of the capacity planning problem. For the first solution phase of the ALR problem, applying only BCD method obtains feasible solution with relative small computation time. In the second solution phase, the global optimal solution for the ALR problem is solved which requires much more time, however the final solution is improved and optimal solution is obtained. So the proposed solution strategy provides us the flexibility in generating both feasible and optimal solution for the capacity planning problem.

Summary and Discussion

In this article, the problem of capacity expansion planning in the face of uncertainty is studied. A two stage SIP based formulation is proposed, which treats the capacity expansion decisions in the first time period as first stage variables whereas the remaining capacity expansion decisions are treated as second stage variables. The resulted stochastic programming problem contains mixed integer first stage and also mixed integer second stage recourse problems. To address the computational complexity of solving the deterministic equivalent problem under discrete scenario description of uncertainty, an augmented Lagrangian optimization method is proposed.

In the proposed augmented Lagrangian optimization framework, a hybrid solution method is developed consisting of two phases: the first phase uses a traditional decomposition strategy (e.g., BCD) to solve the relaxation problem, whereas a global optimization solver is used in the second phase to solve the relaxation problem. Using the proposed solution approach a feasible solution of the original problem can be obtained after the first phase; and the optimal solution of the original problem can be reached in the second phase.

Major advantages of the proposed solution technique include: (1) the sub-problems in the proposed decomposition framework can be solved in parallel, they are mixed integer convex quadratic optimization problems and can be solved using standard MIQP solvers such as CPLEX; (2) the feasibility of the subproblems are always ensured through the incorporation of auxiliary copy variables and the corresponding nonanticipativity constraints. It is worth pointing out that in the proposed solution framework, the computational efficiency can be further improved through parallel computation of the subproblems in the solution procedure since there is no interaction between those subproblems.

Although a two stage stochastic programming model is used in this article, it may be more appropriate to apply multistage stochastic programming based formulation to capture the dynamic decision making characteristic as the uncertainty evolves. So our future work focuses on the application of the proposed approach to a multistage based stochastic integer formulation,[7] which in general takes similar structure but has more nonanticipativity constraints based on the scenario tree representation of the uncertainty. It is also worth to further explore the interaction between the two solution phases in the hybrid solution framework so as to fully utilize the first phase algorithm's computation efficiency and the second phase algorithm's ability to move solution towards global optimality.

Table 9. Computational Statistics for Example 2

			Hybrid Method			
	Direct Solution		First Phase		Second Phase	
S	t	f^*	t	f	t	f
16	3	$-15{,}021 \times 10^5$	14	$-10{,}211 \times 10^5$	2522	$-15{,}021 \times 10^5$
81	36	$-15{,}047 \times 10^5$	134	$-12{,}273 \times 10^5$	7010	$-15{,}047 \times 10^5$
256	4584	$-15{,}368 \times 10^5$	245	$-12{,}043 \times 10^5$	11,254	$-15{,}368 \times 10^5$
625	Intractable		654	$-11{,}542 \times 10^5$	20,366	$-15{,}236 \times 10^5$

S: number of scenarios; t: CPU time in sec; f^*: objective value of direct solution; f: objective value of relaxation problem.

Acknowledgments

The authors gratefully acknowledge financial support from the National Science Foundation under Grants CBET 0625515 and CBET 0966861.

Notation

Index

i = processes
j = chemicals
l = markets
t = time periods
s = scenarios

Variables

(First stage decisions)

$Y1_i$ = binary capacity expansion decision for process i in the first time period
$E1_i$ = amount of capacity expansion for process i to be installed in the first time period
$Q1_i$ = total capacity of process i in the first time period

(Second stage decisions)

Y_{its} = binary variable (value 1 indicates the expansion for process i in the beginning of period t under scenario s)
E_{its} = capacity expansion of process i to be installed in period t under scenario s
Q_{its} = total capacity of process i in period t under scenario s
W_{its} = operating level of process i in period t under scenario s
P_{jlts} = amount of chemical j purchased from market l in period t under scenario s
S_{jlts} = amount of chemical j sold to market l in period t under scenario s

Parameters

p_s = probability for scenario s
Q_{i0} = initial capacity of process i
γ_{jlts}, Γ_{jlts} = sales, purchase prices of chemical j in market l during time period t under scenario s
α_{it}, β_{it} = variable, fixed term of investment cost for process i during period t
δ_{its} = unit operating cost for process i during time period t under scenario s
E_{it}^L, E_{it}^U = lower and upper bounds for capacity expansions of process i in period t
$NEXP(i)$ = maximum allowable number of expansions for process i
CI_t = capital investment limitation corresponding to period t
η_{ij} = production recipe of chemical j of process i
μ_{ij} = consumption recipe of chemical j of process i
a_{jlts}^L, a_{jlts}^U = lower and upper bounds for purchases of product j from market l in period t under scenario s
d_{jlts}^L, d_{jlts}^U = lower and upper bounds for sales of chemical j in market l during period t under scenario s

Literature Cited

1. Sahinidis NV, Grossmann IE, Fornari RE, Chathrathi M. Optimization model for long range planning in the chemical industry. *Comput Chem Eng*. 1989;13:1049–1063.
2. Sahinidis NV, Grossmann IE. Reformulation of multiperiod MILP models for planning and scheduling of chemical processes. *Comput Chem Eng*. 1991;15:255–272.
3. Ierapetritou MG, Pistikopoulos EN. Novel optimization approach of stochastic planning models. *Ind Eng Chem Res*. 1994;33:1930–1942.
4. Liu ML, Sahinidis NV. Long range planning in the process industries: a projection approach. *Comput Oper Res*. 1996;23:237–253.
5. Rajagopalan S, Singh MR, Morton TE. Capacity expansion and replacement in growing markets with uncertain technological breakthroughs. *Manage Sci*. 1998;44:12–30.
6. Chen ZL, Li S, Tirupati D. A scenario-based stochastic programming approach for technology and capacity planning. *Comput Oper Res*. 2002;29:781–806.
7. Ahmed S, Sahinidis NV. An approximation scheme for stochastic integer programs arising in capacity expansion. *Oper Res*. 2003;51:461–474.
8. Laporte G, Louveaux FV. The integer L-shaped method for stochastic integer programs with complete recourse. *Oper Res Lett*. 1993;13:133–142.
9. Caroe CC, Schultz R. Dual decomposition in stochastic integer programming. *Oper Res Lett*. 1999;24:37–45.
10. Loketangen A, Woodruff DL. Progressive hedging and tabu search applied to mixed integer (0, 1) multistage stochastic programming. *J Heurist*. 1996;2:111–128.
11. Sen S, Higle JL. The C 3 theorem and a D 2 algorithm for large scale stochastic mixed-integer programming: set convexification. *Math Program*. 2005;104:1–20.
12. Ruszczynski A, Ermoliev YM, Norkin VI. On optimal allocation of indivisibles under uncertainty. *Oper Res*. 1998;46:381–395.
13. Schultz R, Stougie L, Van Der Vlerk MH. Solving stochastic programs with integer recourse by enumeration: a framework using Grobner basis. *Math Program*. 1998;83:229–252.
14. Carpentier P, Cohen G, Culioli JC, Renaud A. Stochastic optimization of unit commitment: a new decomposition framework. *IEEE Trans Power Syst*. 1996;11:1067–1073.
15. Tosserams S, Etman LFP, Rooda JE. Augmented Lagrangian coordination for distributed optimal design in MDO. *Int J Numer Meth Eng*. 2008;73:1885–1910.
16. Ruszczynski A. On convergence of an augmented Lagrangian decomposition method for sparse convex optimization. *Math Oper Res*. 1995;20:634–656.
17. Cohen G. Auxiliary problem principle and decomposition of optimization problems. *J Optim Theory Appl*. 1980;32:277–305.
18. Bertsekas DP, Tsitsiklis JN, *Parallel and Distributed Computation*. Englewood Cliffs, New Jersey: Prentice-Hall, 1989.
19. Hestenes M. Multiplier and Gradient Methods. *Journal of Optimization Theory and Applications*. 1969;4:303–320.
20. Powell MJD. A method for nonlinear constraints in minimization problems. In: *Optimization*. Fletcher R, editor. New York: Academic Press, 1969.
21. Rockafellar RT. A dual approach to solving nonlinear programming problems by unconstrained optimization. *Math Program*. 1973;5:354–373.
22. Andreani R, Birgin EG, Martínez JM, Schuverdt ML. On augmented Lagrangian methods with general lower-level constraints. *SIAM J Optim*. 2008;18:1286–1309.
23. Birgin EG, Floudas CA, Martínez JM. Global minimization using an Augmented Lagrangian method with variable lower-level constraints. *Math Program Ser A*. 2010;125:139–162.
24. Qi L, Wei Z. On the constant positive linear dependence condition and its application to SQP methods. *SIAM J Optim*. 2000;10:963–981.
25. Luo ZQ, Tseng P. Error bounds and convergence analysis of feasible descent methods: a general approach. *Annals Oper Res*. 1993;46:157–178.
26. Holmström K, Göran A, Edvall MM, *User's guide for TOMLAB*. Seattle, WA: Tomlab Optimization Inc., 2009.
27. Jones DR. *DIRECT global optimization algorithm*. In: *Encyclopedia of Optimization*. Floudas CA, Pardalos PM, editors. The Netherlands: Kluwer Academic Publishing, 2001.
28. Liu ML, Sahinidis NV. Process planning in a fuzzy environment. *Eur J Oper Res*. 1997;100:142–169.

Appendix A: Data for Capacity Planning Example 1

Table A1. Fixed and Variable Investment Coefficients for Example 1

	β_{it} (10^5\$)			α_{it} (10^2\$/(ton year))		
Process	$t = 1$	$t = 2$	$t = 3$	$t = 1$	$t = 2$	$t = 3$
1	340	584	479	11.52	6.82	9.24
2	316	382	352	10.08	9.35	5.34
3	468	238	413	11.91	7.40	9.25
4	338	307	401	4.89	10.10	8.61
5	439	226	201	7.78	7.13	11.81
6	269	203	291	6.69	5.82	6.74

Table A2. Prices of Raw Materials and Products, Upper Bounds for Amiabilities/Demands for Example 1

	Price (10^2\$/ton)			Availability/Demand (kton/year)		
Chemical	$t = 1$	$t = 2$	$t = 3$	$t = 1$	$t = 2$	$t = 3$
	Raw materials:			Availabilities		
1	22.98	34.67	27.89	37	33	39
2	60.23	75.21	79.25	68	36	32
3	18.36	34.81	15.12	45	58	28
4	17.44	21.89	27.74	68	39	46
	Products:			Demands:		
5	58.03	55.14	48.58	91	40	94

Table A3. Operating Expenses Coefficients for Example 1

	δ_{it} (10^2\$/ton)		
Process	$t = 1$	$t = 2$	$t = 3$
1	0.5	0.5	0.3
2	0.3	0.3	0.4
3	0.3	0.4	0.3
4	0.3	0.5	0.3
5	0.2	0.5	0.6
6	0.5	0.5	0.3

Table A4. Mass Balance Coefficients for Example 1

	μ_{ij} (η_{ij})				
Process	$C = 1$	$C = 2$	$C = 3$	$C = 4$	$C = 5$
1	1.22		0.65	(1)	(0.89)
2		0.52	(1.06)		(1)
3				0.7	(1)
4	0.68		(1)	0.79	(0.65)
5		0.82	(1)		
6	1.07	(1)			

Appendix B: Data for Capacity Planning Example 2

Table B1. Fixed and Variable Investment Coefficients for Example 2

	β_{it} (10^5\$)			α_{it} (10^2\$/(ton year))		
Process	$t = 1$	$t = 2$	$t = 3$	$t = 1$	$t = 2$	$t = 3$
1	206	257	399	9.25	8.03	4.90
2	484	510	300	5.18	10.97	11.20
3	409	547	403	8.46	5.97	7.64
4	294	532	550	8.16	4.18	7.68
5	508	496	401	8.41	5.35	4.42
6	334	396	568	10.36	6.39	9.12

Table B2. Prices of Raw Materials and Products, Upper Bounds for Amiabilities/Demands for Example 2

	Price (10^2\$/ton)			Availability/Demand (kton/year)		
Chemical	$t = 1$	$t = 2$	$t = 3$	$t = 1$	$t = 2$	$t = 3$
	Raw materials:			Availabilities		
1	37.54	28.40	25.87	53	55	36
3	29.05	22.69	34.71	50	58	34
	Products:			Demands:		
2	51.36	23.77	40.36	52	73	60
4	42.04	45.64	55.48	61	72	85

Table B3. Operating Expenses Coefficients for Example 2

	δ_{it} (10^2\$/ton)		
Process	$t = 1$	$t = 2$	$t = 3$
1	0.2	0.3	0.2
2	0.3	0.3	0.3
3	0.6	0.5	0.5
4	0.3	0.4	0.3
5	0.6	0.5	0.4
6	0.6	0.6	0.4

Table B4. Mass Balance Coefficients for Example 2

	μ_{ij} (η_{ij})			
Process	$C = 1$	$C = 2$	$C = 3$	$C = 4$
1	1.11	(0.59)	(1)	
2			0.51	(1)
3		(1)	0.74	(0.87)
4	0.57		(1)	
5	0.61	(1)	0.58	
6	0.93	(1)		

Manuscript received Dec. 8, 2010, and revision received Feb. 16, 2011.