Operations Research Letters 43 (2015) 219–225

Contents lists available at ScienceDirect

Operations Research Letters

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

Flexible system design: A perspective from service levels

Lifei Sheng [a,*], Huan Zheng [b], Ying Rong [b], Woonghee Tim Huh [a]

[a] Sauder School of Business, University of British Columbia, Canada
[b] Antai College of Economics and Management, Shanghai Jiaotong University, China

ARTICLE INFO

Article history:
Received 29 April 2014
Received in revised form
23 January 2015
Accepted 23 January 2015
Available online 18 February 2015

Keywords:
Flexibility
Service level
Capacity
Linear programming

ABSTRACT

We consider the capacity portfolio investment problem with flexible machines facing multiple products and demand uncertainty. For the problem of maximizing the service level, we approximate the objective with the largest inscribing sphere and provide a linear program formulation. We show, under certain conditions, that the optimal flexibility configuration consists only of dedicated machines and machines capable of producing only two types of products. Our work not only strengthens the sufficiency of limited flexibility in capacity investment but also suggests a novel way for approximating flexible system design problems in a computationally tractable form.

1. Introduction and literature review

This paper studies the problem of capacity portfolio investment under demand uncertainty. For example, when a firm manufactures several products, which flexible machines should be used and how much capacity should be invested?

The studies of process flexibility stem from the seminal work of Jordan and Graves [12], who have shown that limited flexibility can reap most of the benefits from the fully flexible machines. In the limited flexible system that they have considered, machines produce only 2 products each and form a "2-chain" structure (a long circle connecting all the machines and products). Subsequently, desirable properties of the 2-chain have been shown. Chou et al. [6] have shown that the 2-chain structure can achieve more than 90% benefit of full flexibility if demand follows a certain demand distribution. Simchi-Levi and Wei [15] have argued that the 2-chain is the best flexibility structure in a balanced system where all machines are 2-flexible and each product is produced from 2 machines. However, Desir et al. [11] have shown that 2-chain may not be optimal even with independent and identically distributed demand distributions if the design limits the number of possible "edges" to $2N$, where N is the total number of products. Wang and Zhang [17] have provided a closed-form distribution-free bound on the performance of general k-chain. Deng and Shen [10] have characterized the distribution-free performance of any given flexibility configuration. Chen et al. [5] have analyzed the problem of how to design an optimal sparse flexibility structure which uses the smallest number of flexibility links but achieves at least $1-\epsilon$ fraction of the sales under full flexibility. Chou et al. [7] have studied the worst case performance of generalized sparse flexibility structure. Other works that calibrate and design sparse flexibility structure with fixed capacity investment include [8,13,14,9].

In comparison, capacity investment involving both dedicated and flexible machines has been investigated by Van Mieghem [16]. His 2-product model does not readily generalize to an arbitrary number of products since the number of possible configurations is exponential in the number of products. Bassamboo et al. [2] have studied a similar problem using newsvendor networks, where only two "adjacent levels of flexibility" is needed in capacity investment when system is symmetric, and Bassamboo et al. [1] have extended to the case with a submodular set of products that a machine can process. A similar investigation in [3] with a parallel queuing system has shown that only dedicated machines and flexible machines capable of producing only 2 products is needed asymptotically.

In a flexible resource selection setting similar to [2], we consider the problem of maximizing the joint service level (the probability of having no stock-out) subject to a budget constraint. (In comparison, the objective of [2] is to minimize the penalty cost that is linear in shortage amounts.) Instead of working directly with this objective function, we use a lower bound, which we obtain by working with the notion of the largest inscribed sphere, and consequently we build up a linear program formulation for approximating the flexible system design problem.

* Corresponding author.
E-mail address: Fay.Sheng@sauder.ubc.ca (L. Sheng).

http://dx.doi.org/10.1016/j.orl.2015.01.009
0167-6377/

The main contributions of this work are summarized as below: (1) We study the flexible system design problem from a perspective of service levels. By approximating the objective with its largest inscribed sphere, we rewrite the problem into a linear program (LP) as a heuristic. To the best of our knowledge, we are the first one to apply this idea to the flexible system design problem. (2) We show that our LP heuristic has a good performance under uniform or normal demands. For general settings, where the demand support can be characterized or approximated by a simple polytope, our LP heuristic will still work well. In addition, while our objective is to maximize the joint service level, our LP heuristic solution also leads to a good performance of fill rate (the fraction of demand satisfied). (3) While we do not impose any *a priori* structure on flexibility (such as the chaining structure of [15]), we can show the following property for the approximate problem with the above-mentioned lower bound as the objective—that its optimal solution uses only dedicated machines and the machines capable of producing only 2 products, which strengthens the convention that limited flexibility, particularly limited to 2 products, is adequately sufficient in designing flexibility.

2. Problem description

Let the products be indexed by i, where $i = 1, \ldots, N$, and N be the total number of products. Let $D = (D_1, \ldots, D_N)$ denote the random demand vector. A machine can produce one or more of the products, and we say a machine has *level-k flexibility* if it is designed to manufacture k different products. We assume that the production rate of a machine is constant and independent of the product and machine type. Let $S \subseteq \{1, \ldots, N\}$ denote the capacity type, and let $x = (x_S)_{S\subseteq\{1,\ldots,N\}}$ be the decision variable associated with how much capacity would be invested on each machine. Let c_S denote the capacity installation cost for machine type S. Also, let B be the investment budget.

Our objective is to maximize the joint service level, the probability that all demands are satisfied under the given capacity decision x. Given the demand realization $(d_1, \ldots, d_N)$, it can be satisfied if there is a feasible allocation $\{u_{Si}\}$ satisfying the following inequalities:

$$\sum_{S\subseteq\{1,\ldots,N\}:S\cap\{i\}\neq\emptyset} u_{Si} \geq d_i \quad \forall i, \qquad \sum_{i\in\{1,\ldots,N\}:S\cap\{i\}\neq\emptyset} u_{Si} \leq x_S \quad \forall S,$$
$$\text{and} \quad u_{Si} \geq 0 \quad \forall S, i.$$

It can be shown, using the max-flow min-cut theorem, that $(d_1, \ldots, d_N)$ can be satisfied under capacity investment $x = (x_S)_{S\subseteq\{1,\ldots,N\}}$ if and only if $(d_1, \ldots, d_N) \in \Omega(x)$, where

$$\Omega(x) = \left\{ (d_1', \ldots, d_N') \,\middle|\, \sum_{j\in J} d_j' \leq \sum_{\{S:J\cap S\neq\emptyset\}} x_S \right.$$
$$\left. \text{for each } J \subseteq \{1, \ldots, N\} \right\}. \tag{1}$$

Thus, the problem of maximizing the joint service level under a budget constraint can be formulated as:

$$\max \left\{ P(D \in \Omega(x)) \,\middle|\, \sum_{S\subseteq\{1,\ldots,N\}} c_S x_S \leq B,\ x_S \geq 0 \right.$$
$$\left. \text{for each } S \subseteq \{1, \ldots, N\} \right\}. \tag{2}$$

Note that the number of decision variables is exponential in N, and thus may be difficult to solve when N is large. Thus, we first analyze a symmetric and uniformly distributed demand case (Section 3), and then discuss possible extensions (Section 4).

3. Lower bound model: symmetric uniform demand

Since the exact formulation in (2) can be difficult to analyze, we consider instead an approximate model where its objective is replaced with a lower bound. In this section, we are to establish a lower bound approximation model under symmetric and uniformly distributed demand. We will show that the lower bound model is equivalent to a linear program problem (LP) and therefore leads to attractive properties in its optimal solution. Further we will discuss the effectiveness of our approximation, compared to the exact problem (2).

We start by introducing a few assumptions.

Assumption 1 (*Symmetric Cost*)**.** All level-k flexibility has the same unit cost denoted by s_k, i.e., $c_S = s_k$ for each S satisfying $|S| = k$.

Without loss of generality, we normalize $s_1 = 1$. Since c_S is increasing in S, it follows that $1 = s_1 \leq s_2 \leq \cdots \leq s_N$.

Assumption 2 (*Uniform and Symmetric Demand*)**.** Demand D_i, $i = 1, \ldots, N$, is independently and uniformly distributed on $[0, 1]$.

Under Assumption 2, the joint service level is equivalent to the volume of the intersection between the feasible production region $\Omega(x)$ and demand support $[0, 1]^N$, but the objective function in (2) may not behave as well. For example, if $N = 2$, it can be shown that $P(D \in \Omega(x)) = (x_1 + x_{12})(x_2 + x_{12}) - (x_{12})^2/2$, given that $0 \leq x_i + x_{12} \leq 1$ $(i = 1, 2)$, which is neither convex nor concave in (x_1, x_2, x_{12}). (See Fig. 1(a) for an illustration.) To avoid the non-convexity of the objective function, we propose a lower bound approximation model. This model is based on a novel idea of an inscribed sphere, and it is easy to solve since it can be transformed to an LP of a reasonable size.

3.1. The lower bound model: description and analysis

Instead of computing the volume of the intersection between $\Omega(x)$ and $[0, 1]^N$, we compute the volume of the largest inscribed sphere within it. Clearly, the sphere's volume provides a lower bound to the joint service level. (See Fig. 1(b) for an illustration.)

Consider a sphere centered at $p = (p_1, \ldots, p_N)$ and radius r. The volume of this N-dimensional sphere is

$$\frac{\pi^{\frac{N}{2}}}{\Gamma\left(\frac{N}{2}+1\right)} r^N, \tag{3}$$

where Γ indicates the Gamma function $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt$. Note that the volume of the sphere is strictly increasing in r, thus the maximization of the volume is equivalent to maximizing r. If the sphere is inside the intersection between the feasible production region $\Omega(x)$ and the support of the uniform demand distribution, we must have

$$p \in \Omega(x), \tag{4}$$
$$r \leq dist(p, \partial_x\Omega(x)), \quad \text{and} \tag{5}$$
$$r \leq p_i \quad \text{and} \quad r \leq 1 - p_i \quad \forall i, \tag{6}$$

where the first condition states that the center p should be in $\Omega(x)$, and the second condition requires that the radius of the sphere should not exceed the distance from the center p to the boundary of $\Omega(x)$ denoted by $\partial_x\Omega(x)$. The third condition ensures that the sphere lies within the support of demand distribution, i.e. the unit cube $[0, 1]^N$.

From (1), the boundaries of $\Omega(x)$ are the hyperplanes obtained by fixing each inequality to equality. By referring to the formula of the Euclidean distance from a point to a hyperplane, one can show