

**stichting  
mathematisch  
centrum**



---

AFDELING MATHEMATISCHE BESLISKUNDE  
(DEPARTMENT OF OPERATIONS RESEARCH)

BW 142/81

JUNI

M.A.H. DEMPSTER, M.L. FISHER, L. JANSEN,  
B.J. LAGEWEG, J.K. LENSTRA, A.H.G. RINNOOY KAN

ANALYSIS OF HEURISTICS FOR STOCHASTIC PROGRAMMING:  
RESULTS FOR HIERARCHICAL SCHEDULING PROBLEMS

Preprint

---

**kruislaan 413 1098 SJ amsterdam**

*Printed at the Mathematical Centre, 413 Kruislaan, Amsterdam.*

*The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).*

ANALYSIS OF HEURISTICS FOR STOCHASTIC PROGRAMMING:  
RESULTS FOR HIERARCHICAL SCHEDULING PROBLEMS

M.A.H. DEMPSTER (*Balliol College, Oxford/IIASA, Laxenburg*)

M.L. FISHER (*The Wharton School, University of Pennsylvania, Philadelphia*)

L. JANSEN (*Algemene Bank Nederland, Amsterdam*)

B.J. LAGEWEG (*Mathematisch Centrum, Amsterdam*)

J.K. LENSTRA (*Mathematisch Centrum, Amsterdam*)

A.H.G. RINNOOY KAN (*Erasmus University, Rotterdam*)

ABSTRACT

Certain multi-stage decision problems that arise frequently in operations management planning and control allow a natural formulation as multi-stage stochastic programs. In job shop scheduling, for example, the first stage could correspond to the acquisition of resources subject to probabilistic information about the jobs to be processed, and the second stage to the actual allocation of the resources to the jobs given deterministic information about their processing requirements. For two simple versions of this hierarchical scheduling problem, we describe heuristic solution methods and show that their performance is asymptotically optimal both in expectation and in probability.

KEY WORDS & PHRASES: *multi-stage stochastic programming, hierarchical scheduling, heuristic, analytical evaluation, asymptotic optimality, convergence in expectation, convergence in probability*

NOTE: This report has been submitted for publication in a journal.



## 1. INTRODUCTION

Certain multi-stage decision problems that arise frequently in operations management planning and control allow a natural formulation as *multi-stage stochastic programming* problems. In the context of *job shop scheduling*, for example, at least two decision stages can usually be recognized. At the *aggregate* level, one has to decide upon the acquisition of resources; precise information about what will be required of them, however, is either unavailable because it results from unknown future developments, or intentionally suppressed to facilitate the decision making. Subsequently, at the *detailed* level, one has to decide upon the actual allocation of the resources over time, when all the relevant information is at hand.

Problems of this type occur in other settings as well, such as the design of *distribution* and *vehicle routing* systems [3]. They always involve a sequence of decisions over time, at an increasing level of detail and with increasing information becoming available. Quite often, however, probabilistic information about the effect of the decisions can be obtained in advance. Each decision can then be made to correspond to a stage in the stochastic programming model, the objective now being to set the decision variables at each stage in such a way that the joint outcome of the decision in question and all its successors is *optimal in expectation*.

As described in a companion paper [3], problems of this type are more and more frequently being solved by *hierarchical planning systems*. Such a system consists of suitably linked mathematical programming models, each of which corresponds to a particular decision stage. These models have so far always been *deterministic* in nature. By opting for a *stochastic* programming model instead, we not only arrive at a more accurate problem representation but also obtain a framework in which the quality of hierarchical planning systems can be evaluated *analytically* rather than by *ad hoc* methods such as Monte Carlo simulation.

The increased sophistication of stochastic programming models comes at the expense of an increased computational effort required to solve them to optimality. The decision problems occurring at each stage are already quite hard when all data are known with certainty. Insofar as these problems are of a combinatorial nature, they often belong to the notorious class of *NP-*

hard problems [8;13]. To estimate the effect of an optimal solution to such problems when certain data are only known in terms of probability distributions is a formidable computational task.

Any hierarchical planning system designed to solve the multi-stage stochastic program in its entirety is therefore likely to be of a *heuristic* nature, yielding an approximate solution rather than a guaranteed optimum. We shall see, however, that such *stochastic programming heuristics* can be subjected to the same type of *worst case performance analysis* that has been so successful in the area of deterministic combinatorial optimization [4;7;8]. The main difference with the deterministic case is that the heuristic and the optimum solution values which must be compared are both random variables.

We will present such analyses for two special cases of the job shop scheduling problem described above. In both cases, the aggregate (first stage) problem is to decide on the number and types of machines to be acquired. When this decision has to be made, the cost and speed of each type of machine and the number of jobs to be processed are known with certainty, but the job processing times are only known stochastically; in fact, we assume that they are independent, identically distributed random variables. After the aggregate decision has been made, a particular realization of the processing times becomes known. The detailed (second stage) problem is then to decide on the schedule of the jobs so as to minimize the maximum of their completion times. The overall objective is to minimize the machine cost plus the maximum job completion time.

Section 2 deals with the case in which all available machines are *identical* both in cost and in speed; all that has to be decided at the first stage is how many of them are to be acquired. Section 3 considers the more general case of *uniform* machines, each of which has its own cost and speed; the first stage decision involves the selection of an optimal subset of the available machines. In both cases, we will describe a simple heuristic and show that it sets the decision variables at both stages in such a way that, as the number of jobs goes to infinity, the error in the heuristic solution value relative to the optimum solution value approaches zero, in expectation as well as in probability.

These results are the first examples to our knowledge of a worst case analysis of heuristics for stochastic programming. They are sufficiently

encouraging to justify our hope that they can be extended to a wider class of problems. Section 4 briefly describes extensions of the model to the case in which the second stage objective is to minimize the maximum job tardiness with respect to a common due date, to a multi-period model involving recourse decisions, and to the case that the number of jobs is a random variable as well. Possible applications of our approach to other multi-stage decision problems on which we are currently working, such as the design of hierarchical distribution systems, are mentioned too. We expect that these results will lead to the formulation of general conditions under which asymptotically optimal stochastic programming heuristics exist.

The Appendix contains some results from probability theory that are used in the paper. A tilde ( $\sim$ ) under a variable will indicate that it is a random variable.

## 2. IDENTICAL MACHINES

The two-stage stochastic programming model studied in this section is the following. At the first stage, one has to decide on the number  $m$  of *identical parallel machines* that are to be acquired, while knowing the cost  $c$  of a single machine, the number  $n$  of *jobs*  $J_1, \dots, J_n$  that are to be processed, and the *probability distribution* of the vector  $\underline{p} = (p_1, \dots, p_n)$  of their *processing times*. At the second stage, after  $m$  has been determined, a *realization*  $\underline{p}$  of  $\underline{p}$  becomes known, and one has to decide on a *schedule* in which each machine processes at most one job at a time, each job  $J_j$  is processed during an uninterrupted time period of length  $p_j$ , and no job is processed prior to time 0.

The second stage objective is to minimize the maximum job completion time, given  $m$  and  $\underline{p}$ ; let  $C^*(m, \underline{p})$  denote the minimum value to be computed. The overall objective is to minimize, over all realizations of  $\underline{p}$ , the expected value of the total cost  $\tilde{z}^*(m) = cm + C^*(m, \underline{p})$ , i.e., to determine  $m^*$  such that

$$E\tilde{z}^*(m^*) = \min_m \{cm + EC^*(m, \underline{p})\}.$$

Note that, in precise terms,  $c$  is the ratio of machine cost to delay cost per time unit.

It is not obvious how to solve this problem to optimality. Computing  $C^*(m, \underline{p})$  for given  $m$  and  $\underline{p}$  is already an NP-hard problem [13], so determining  $EC^*(m, \underline{p})$  as a function of  $m$  for an arbitrary given probability distribution of  $\underline{p}$  seems virtually impossible. In designing a heuristic for the first stage problem, we apply an idea that is fundamental to all hierarchical planning systems: we suppress the combinatorial fine structure of the second stage problem and replace  $C^*(m, \underline{p})$  by a *lower bound*  $P/m$ , where  $P = \sum_{j=1}^n p_j$ . The resulting problem is then to minimize the expected value of the lower bound on  $\tilde{z}^*(m)$  given by  $\tilde{z}^{LB}(m) = cm + P/m$ , i.e., the heuristic sets  $m = m^{LB}$  where

$$E\tilde{z}^{LB}(m^{LB}) = \min_m \left\{ cm + \frac{EP}{m} \right\}.$$

The derivative of the objective function is zeroed by  $m = \sqrt{EP/c}$ . Since  $m$  must be a positive integer,  $m^{LB}$  is determined by minimizing  $cm + EP/m$  subject to  $m \in \{\lfloor \sqrt{EP/c} \rfloor, \lceil \sqrt{EP/c} \rceil\} \cap \mathbb{N}$ .

The heuristic for the second stage problem schedules the jobs  $J_1, \dots, J_n$



Problem instance:  $m^{LB} = 3$ ,  $n = 7$ ,  $p = (1,2,3,4,5,6,7)$ .

List:  $(1,2,3,4,5,6,7)$ .

Schedule:

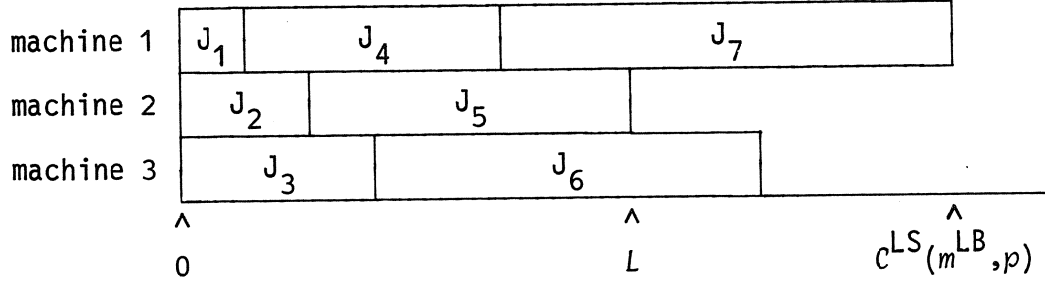


Figure 1. Illustration of the second stage heuristic.

with processing times  $p_1, \dots, p_n$  on  $m^{LB}$  machines according to *list scheduling*: the jobs are placed in arbitrary order and at each step the next job on the list is assigned to the earliest available machine [4;7;10] (cf. Figure 1). Let  $C^{LS}(m, p)$  denote the earliest time by which all jobs are completed under this rule, for given  $m$  and  $p$ , and let  $\tilde{z}^{LS}(m) = cm + C^{LS}(m, \tilde{p})$ . The two-stage stochastic programming heuristic produces a solution of total cost  $\tilde{z}^{LS}(m^{LB})$ .

(Since our results concern an *arbitrary* list scheduling rule, we may as well consider a stochastic version of the second stage model, in which  $p_j$  becomes known only at the completion of  $J_j$  ( $j = 1, \dots, n$ ), and solve it by the following heuristic: at time 0, a job is assigned to each of the  $m$  machines in list order; as soon as the first job is completed on a certain machine, the next job on the list is assigned to that machine, and so on until  $C^{LS}(m, \tilde{p})$  is realized.)

The *expected worst case performance* of our heuristic is characterized by Theorem 1.

$$\text{THEOREM 1. } \frac{\tilde{z}^{LS}(m^{LB})}{\tilde{z}^*(m^*)} \leq 1 + \frac{E p_{\max}}{2\sqrt{cEP}}.$$

*Proof.* Consider a schedule produced by the list scheduling rule on  $m^{LB}$  machines for a realization  $p$  of  $\tilde{p}$ . Let  $L$  denote the latest time that all machines are occupied and let job  $J_k$  be completed last. It follows from the nature of list scheduling that

$$C^{LS}_{(m^{LB}, p)} \leq L + p_k \leq \frac{P}{m^{LB}} + p_{\max}$$

(cf. Figure 1) and hence

$$z^{LS}_{(m^{LB})} \leq z^{LB}_{(m^{LB})} + p_{\max}.$$

Taking expectations, we find that

$$\begin{aligned} E z^{LS}_{(m^{LB})} &\leq E z^{LB}_{(m^{LB})} + E p_{\max} \\ &\leq E z^{LB}_{(m^*)} + E p_{\max} \quad (\text{by the definition of } m^{LB}) \\ &\leq E z^*_{(m^*)} + E p_{\max} \quad (\text{since } P/m^* \leq C^*(m^*, p)). \end{aligned}$$

On the other hand,

$$E z^*_{(m^*)} \geq E z^{LB}_{(\sqrt{EP/c})} = 2\sqrt{cEP}.$$

Combining the last two inequalities establishes the theorem.  $\square$

Before we discuss the asymptotic implications of Theorem 1, we note that there exist second stage heuristics with a better worst case performance than arbitrary list scheduling [7;10]. Moreover, certain simple variations on our heuristic have error bounds different from the one above. For example, if  $m^{LB}$  is determined by minimizing  $cm + E \max\{p_{\max}, P/m\}$ , then we can use the result from [9] that  $C^{LS}_{(m, p)} \leq (2-1/m) \max\{p_{\max}, P/m\}$  to obtain a heuristic whose expected total cost is less than twice the optimal expected value. However, this bound may be considerably higher than the bound given by Theorem 1, particularly in the realistic situation of uniformly bounded processing times. In any case, Theorem 1 will turn out to be sufficiently strong for establishing asymptotic optimality.

To investigate the performance of our heuristic as the number of jobs tends to infinity, we require the following assumption about the probability distribution of the processing times.

**ASSUMPTION A.** The processing times  $p_j$  ( $j = 1, \dots, n$ ) are independent, identically distributed random variables with finite second moment.

Under Assumption A,  $E\tilde{p} = n\mu$ , where  $\mu = E\tilde{p}_j$ . In the Appendix we prove the following lemma.

LEMMA 1. Under Assumption A,

- (a)  $\lim_{n \rightarrow \infty} E\tilde{p}_{\max}/\sqrt{n} = 0$ ;
- (b)  $\text{plim}_{n \rightarrow \infty} \tilde{p}_{\max}/\sqrt{n} = 0$ .  $\square$

We recall that a sequence of random variables  $\tilde{x}_n$  converges to a random variable  $\tilde{x}$  in probability (notation:  $\text{plim}_{n \rightarrow \infty} \tilde{x}_n = \tilde{x}$ ) if

$$\lim_{n \rightarrow \infty} P\{|\tilde{x}_n - \tilde{x}| > \varepsilon\} = 0 \quad \text{for all } \varepsilon > 0,$$

and in distribution if

$$\lim_{n \rightarrow \infty} F_{\tilde{x}_n}(x) = F_{\tilde{x}}(x) \quad \text{at all points of continuity of } F_{\tilde{x}},$$

where  $F_{\tilde{y}}$  denotes the distribution function of a random variable  $\tilde{y}$ . The first concept of stochastic convergence implies the second; in the present context, the two concepts coincide since the limiting random variable  $\tilde{x}$  will always be a constant [2, p.93, Ex.4].

The following result is an immediate consequence of Theorem 1 and Lemma 1(a).

THEOREM 2. Under Assumption A,

$$\lim_{n \rightarrow \infty} \frac{E\tilde{z}_{(m^{LB})}^{LS}}{E\tilde{z}_{(m^*)}^*} = 1. \quad \square$$

Theorem 2 states that the performance of our heuristic is *asymptotically optimal in expectation*. We will now show that it is also *asymptotically optimal in probability*. To this end, we will consider the probability limit

$$\text{plim}_{n \rightarrow \infty} \frac{\tilde{z}_{(m^{LB})}^{LS}}{\tilde{z}_{(m^*)}^*} = \text{plim}_{n \rightarrow \infty} \frac{cm^{LB} + C^{LS}(m^{LB}, p)}{cm^* + C^*(m^*, p)}.$$

The reason that this distributional analysis is not completely straightforward is that both  $m^{LB}$  and  $m^*$  have been chosen to minimize an expected value.

The following two lemmas state that at both stages the approximation on which the heuristic is based is asymptotically accurate in probability.

LEMMA 2. If  $m = O(\sqrt{n})$ , then under Assumption A

$$\text{plim}_{n \rightarrow \infty} \frac{C^*(m, \tilde{p})}{n\mu/m} = \text{plim}_{n \rightarrow \infty} \frac{C^{LS}(m, \tilde{p})}{n\mu/m} = 1.$$

*Proof.* For every realization  $p$  of  $\tilde{p}$  we have that

$$\frac{P}{m} \leq C^*(m, p) \leq C^{LS}(m, p) \leq \frac{P}{m} + p_{\max}.$$

Dividing this by  $n\mu/m$  yields

$$\frac{P-n\mu}{n\mu} + 1 \leq \frac{C^*(m, p)}{n\mu/m} \leq \frac{C^{LS}(m, p)}{n\mu/m} \leq \frac{P-n\mu}{n\mu} + 1 + \frac{mp_{\max}}{n\mu}. \quad (1)$$

The observation that, under Assumption A,  $\text{plim}_{n \rightarrow \infty} (P-n\mu)/n\mu = 0$ , and Lemma 1(b) imply the lemma.  $\square$

LEMMA 3. Under Assumption A,

$$\lim_{n \rightarrow \infty} \frac{m^*}{\sqrt{n\mu/c}} = \lim_{n \rightarrow \infty} \frac{m^{LB}}{\sqrt{n\mu/c}} = 1.$$

*Proof.* Since  $m^{LB} \in \{\lfloor \sqrt{n\mu/c} \rfloor, \lceil \sqrt{n\mu/c} \rceil\}$ , we trivially have that

$$\lim_{n \rightarrow \infty} \frac{m^{LB}}{\sqrt{n\mu/c}} = 1. \quad (2)$$

Suppose next that there exist an  $\varepsilon > 0$  and an  $n_0$  such that for all  $n > n_0$

$$\frac{m^*}{\sqrt{n\mu/c}} > 1 + \varepsilon. \quad (3)$$

We will show that this implies

$$E\tilde{Z}^*(m^*) > E\tilde{Z}^*(m^{LB}), \quad (4)$$

which contradicts the optimality of  $m^*$ .

If (3) would hold as an equality, then we would have that

$$\frac{E\tilde{Z}^{LB}(m^*)}{E\tilde{Z}^{LB}(\sqrt{n\mu/c})} = \frac{cm^* + n\mu/m^*}{2\sqrt{cn\mu}} = \frac{1}{2}(1 + \varepsilon + \frac{1}{1+\varepsilon}).$$

Inequality (3) and the unimodality of the function  $E\tilde{Z}^{LB}(m)$  therefore imply that, if  $\varepsilon < 1$ ,

$$E\tilde{Z}^*(m^*) \geq E\tilde{Z}^{LB}(m^*) > \frac{1}{2}(1 + \varepsilon + \frac{1}{1+\varepsilon})E\tilde{Z}^{LB}(\sqrt{n\mu/c}) > (1 + \frac{1}{4}\varepsilon^2)E\tilde{Z}^{LB}(\sqrt{n\mu/c}). \quad (5)$$

Consider the first, second and fourth term of the inequalities (1). Setting  $m = m^{LB}$ , taking expectations, and using  $m^{LB} = O(\sqrt{n})$  and Lemma 1(a), we obtain

$$\lim_{n \rightarrow \infty} \frac{EC^*(m^{LB}, p)}{n\mu/m^{LB}} = 1$$

and hence

$$\lim_{n \rightarrow \infty} \frac{EZ_{\tilde{z}}^*(m^{LB})}{EZ_{\tilde{z}}^{LB}(m^{LB})} = 1.$$

In view of (2), it follows that for  $n$  sufficiently large

$$(1 + \frac{1}{4}\epsilon^2)EZ_{\tilde{z}}^{LB}(\sqrt{n\mu}/c) > EZ_{\tilde{z}}^*(m^{LB}). \quad (6)$$

Inequalities (5) and (6) together give the contradiction (4).

The assumption that there exists an  $\epsilon > 0$  and an  $n_0$  such that for all  $n > n_0$

$$\frac{m^*}{\sqrt{n\mu}/c} < 1 - \epsilon$$

similarly yields a contradiction. We conclude that

$$\lim_{n \rightarrow \infty} \frac{m^*}{\sqrt{n\mu}/c} = 1,$$

as required.  $\square$

By combining Lemmas 2 and 3, we obtain the desired result.

**THEOREM 3.** *Under Assumption A,*

$$\text{plim}_{n \rightarrow \infty} \frac{z_{\tilde{z}}^{LS}(m^{LB})}{z_{\tilde{z}}^*(m^*)} = 1. \quad \square$$

The following result is also a straightforward consequence of Lemmas 2 and 3.

**THEOREM 4.** *Under Assumption A,*

$$\text{plim}_{n \rightarrow \infty} \frac{z_{\tilde{z}}^*(m^*)}{2\sqrt{cn\mu}} = \text{plim}_{n \rightarrow \infty} \frac{z_{\tilde{z}}^{LS}(m^{LB})}{2\sqrt{cn\mu}} = 1. \quad \square$$

Theorem 4 implies that the optimal solution value is asymptotically equal to  $2\sqrt{cn\mu}$  and that the heuristic solution value approaches the same limit. Similar strong results have been obtained for other probabilistic extensions of com-

binatorial optimization problems, mostly of a geometric nature, such as the Euclidean traveling salesman problem [1;15] and the planar K-median problem [5;12].

Under the additional assumption that the processing times have a finite *first* moment as well, it is possible to show that the performance of our heuristic is even *asymptotically optimal with probability 1*. We recall that a sequence of random variables  $\tilde{x}_n$  converges to a random variable  $\tilde{x}$  *almost surely* or *with probability 1* if

$$P\{\lim_{n \rightarrow \infty} \tilde{x}_n = \tilde{x}\} = 1.$$

To see why this is true, consider the last part of the proof of Lemma 2. The strong law of large numbers implies that, if  $\mu$  is finite, then  $P\{\lim_{n \rightarrow \infty} (\tilde{P} - n\mu)/n\mu = 0\} = 1$ . In the Appendix we prove that, under Assumption A,  $P\{\lim_{n \rightarrow \infty} \tilde{p}_{\max}/\sqrt{n} = 0\} = 1$ . It follows that the convergence results of Lemma 2 hold not only in probability but also almost surely, and hence the same is true with respect to Theorems 3 and 4.

### 3. UNIFORM MACHINES

The model studied in this section is an extension of the one considered in Section 2. The only difference is that, at the first stage, rather than deciding on a number of identical machines, one has to select a subset from a set  $M$  of *uniform machines*, while knowing the *cost*  $c_i$  and *speed*  $s_i$  of each machine  $i \in M$ . When, at the second stage, job  $J_j$  is assigned to machine  $i$ , it has to be processed during a period of length  $p_j/s_i$ .

We define  $q_i = c_i/s_i$  for all  $i \in M$ . Let  $c^L, c^U, s^L, s^U$  be such that  $c^L \leq c_i \leq c^U$  and  $s^L \leq s_i \leq s^U$  for all  $i \in M$ , and let  $q^L = c^L/s^U$ ,  $q^U = c^U/s^L$ . For a given subset  $M \subset M$ , let  $c(M) = \sum_{i \in M} c_i$ ,  $s(M) = \sum_{i \in M} s_i$ , and let  $C^*(M, p)$  denote the minimum value of the maximum job completion time, given a realization  $p$  of  $p$ .

The objective is to minimize the expected value of  $\tilde{w}^*(M) = c(M) + C^*(M, p)$ , i.e., to determine  $M^* \subset M$  such that

$$E\tilde{w}^*(M^*) = \min_M \{c(M) + EC^*(M, p)\}.$$

In designing a heuristic for the first stage problem, we proceed along similar lines as in the previous section: we replace the second stage objective  $C^*(M, p)$  by a lower bound  $P/s(M)$  and attempt to minimize the expected value of the lower bound on  $\tilde{w}^*(M)$  given by  $\tilde{w}^{LB}(M) = c(M) + P/s(M)$ . Let  $M^{LB} \subset M$  be such that

$$E\tilde{w}^{LB}(M^{LB}) = \min_M \left\{ c(M) + \frac{EP}{s(M)} \right\}. \quad (7)$$

Unfortunately, it is unlikely that  $M^{LB}$  can be determined in polynomial time.

LEMMA 4. The problem of minimizing  $E\tilde{w}^{LB}(M)$  over all  $M \subset M$  is NP-hard.

*Proof.* We will show that problem (7) is a generalization of the following known NP-complete problem [13]:

PARTITION: Given a set  $T = \{1, \dots, t\}$  and positive integers  $a_1, \dots, a_t, b$  with  $\sum_{i \in T} a_i = 2b$ , does there exist a subset  $T' \subset T$  such that  $\sum_{i \in T'} a_i = b$ ?

Given any instance of PARTITION, we construct an instance of problem (7) by

defining  $M = T$ ,  $c_i = s_i = a_i$  ( $i \in T$ ) and  $EP = b^2$ . It is easily verified that, for a subset  $T \subset T$ ,  $\sum_{i \in T} a_i = b$  if and only if  $c(T) + EP/s(T) \leq 2b$ . Hence, PARTITION yields a positive answer if and only if (7) has a solution with value at most  $2b$ .  $\square$

Lemma 4 strongly suggests the inevitability of a heuristic approach to solve (7) in polynomial time. We propose to use a *greedy heuristic* that yields an approximate solution  $M^G$  to (7). Let  $M = \{1, 2, \dots, |M|\}$ , suppose that the machines are numbered according to nondecreasing ratios  $q_i$ , and define  $C_i = \sum_{h=1}^i c_h$ ,  $S_i = \sum_{h=1}^i s_h$ ,  $W_i = C_i + EP/S_i$  ( $i \in M$ ) and  $W_0 = \infty$ . The greedy heuristic selects a subset  $M^G = \{1, \dots, g\} \subset M$ , where  $g$  is the largest index such that  $W_{g-1} > W_g$ .

An absolute bound on the worst case performance of the first stage heuristic is established by Lemma 5.

LEMMA 5. *The greedy solution  $M^G$  satisfies*

- (a)  $W_g = \min_{i \in M} \{W_i\}$ ;
- (b)  $E_{\tilde{W}}^{LB}(M^G) \leq E_{\tilde{W}}^{LB}(M^{LB}) + c^U$ .

*Proof.* (a) The values  $W_i$  define a piecewise linear function  $W(S)$  on the interval  $[0, s(M)]$  as follows: if  $S = \lambda S_i + (1-\lambda)S_{i+1}$  for some  $i \geq 0$  and  $\lambda \in [0, 1]$ , then  $W(S) = \lambda W_i + (1-\lambda)W_{i+1}$ . This function is convex, since its slope over  $[S_i, S_{i+1}]$  increases with  $i$ :

$$\frac{W_{i+1} - W_i}{S_{i+1} - S_i} - \frac{W_i - W_{i-1}}{S_i - S_{i-1}} = q_{i+1} - q_i + \frac{(s_{i+1} + s_i)EP}{S_{i+1}S_iS_{i-1}} > 0.$$

Therefore,  $S_g$  is uniquely determined as the smallest value of  $S$  for which  $W(S)$  takes on its minimum.

(b) There exists an  $i \geq 0$  such that  $S_i \leq s(M^{LB}) \leq S_{i+1}$ . Since  $C_i$  is the minimum machine cost at which a total speed  $S_i$  can be achieved, we have  $C_i \leq c(M^{LB})$ . Hence,

$$\begin{aligned} E_{\tilde{W}}^{LB}(M^G) - E_{\tilde{W}}^{LB}(M^{LB}) &\leq \left(C_{i+1} + \frac{EP}{S_{i+1}}\right) - \left(c(M^{LB}) + \frac{EP}{s(M^{LB})}\right) \\ &\leq C_{i+1} - C_i = c_{i+1} \leq c^U. \quad \square \end{aligned}$$



The heuristic for the second stage problem again schedules the jobs  $J_1, \dots, J_n$  with processing requirements  $p_1, \dots, p_n$  on the machines in  $M^G$  according to *list scheduling*. Let  $C^{LS}(M, p)$  denote the earliest time by which all jobs are completed under this rule, for given  $M$  and  $p$ , and let  $\tilde{w}^{LS}(M) = c(M) + C^{LS}(M, \tilde{p})$ . The overall heuristic produces a solution with value  $\tilde{w}^{LS}(M^G)$ .

The *expected worst case performance* of our heuristic is characterized by Theorem 5.

$$\text{THEOREM 5. } \frac{\tilde{w}^{LS}(M^G)}{\tilde{w}^*(M^*)} \leq 1 + \frac{c^U + E\tilde{p}_{\max}/s^L}{2\sqrt{q^L E\tilde{p}}}$$

*Proof.* It is easily seen that for every realization  $p$  of  $\tilde{p}$

$$C^{LS}(M^G, p) \leq \frac{P}{s(M^G)} + \frac{p_{\max}}{s^L}. \quad (8)$$

Hence,

$$\begin{aligned} \tilde{w}^{LS}(M^G) &\leq \tilde{w}^{LB}(M^G) + E\tilde{p}_{\max}/s^L \\ &\leq \tilde{w}^{LB}(M^{LB}) + c^U + E\tilde{p}_{\max}/s^L \quad (\text{by Lemma 5}) \\ &\leq \tilde{w}^{LB}(M^*) + c^U + E\tilde{p}_{\max}/s^L \quad (\text{by the definition of } M^{LB}) \\ &\leq \tilde{w}^*(M^*) + c^U + E\tilde{p}_{\max}/s^L \quad (\text{since } P/s(M^*) \leq C^*(M^*, p)). \end{aligned}$$

On the other hand,

$$\tilde{w}^*(M^*) \geq \tilde{w}^{LB}(M^*) \geq q^L s(M^*) + \frac{E\tilde{p}}{s(M^*)} \geq 2\sqrt{q^L E\tilde{p}}.$$

These inequalities establish the theorem.  $\square$

For an asymptotic analysis, we need some assumptions about the set  $M$ . It is reasonable to assume that both costs and speeds have constant lower and upper bounds, i.e.,  $c^L, c^U, s^L, s^U$  are constants rather than problem instance dependent variables. This will imply that the number of selected machines grows as  $\sqrt{n}$ , as in the previous section. It is then also reasonable to assume that the number of available machines grows faster than  $\sqrt{n}$ , but remains polynomially bounded in  $n$  in order to allow an efficient implementation of the greedy heuristic. We therefore require the following.

ASSUMPTION B. The parameters  $c^L, c^U, s^L, s^U$  for which  $c^L \leq c_i \leq c^U$  and  $s^L \leq s_i \leq s^U$  ( $i \in M$ ) are fixed constants. Moreover, there exist constants  $D > 0$ ,  $D' > 0$ ,  $\varepsilon' \geq \varepsilon > 0$  such that  $Dn^{\frac{1}{2}+\varepsilon} \leq |M| \leq D'n^{\frac{1}{2}+\varepsilon'}$ .

The following result is an immediate consequence of Theorem 5 and Lemma 1(a).

THEOREM 6. Under Assumptions A and B,

$$\lim_{n \rightarrow \infty} \frac{E_{\tilde{W}}^{LS}(M^G)}{E_{\tilde{W}}^*(M^*)} = 1. \quad \square$$

As before, we now proceed to show that the performance of our heuristic is not only asymptotically optimal in expectation but also in probability. The strong property of the identical machine model expressed in Theorem 4 does not hold for the uniform machine model, and hence the proof is more complicated.

The following lemma states that the total speeds of the sets  $M^{LB}$ ,  $M^G$  and  $M^*$  grow as  $\sqrt{n}$ . We use the notation  $g(n) = \Theta(f(n))$  to indicate that there exist constants  $C > 0$  and  $C' > 0$  such that  $Cf(n) \leq |g(n)| \leq C'f(n)$  for  $n$  sufficiently large.

LEMMA 6. Under Assumptions A and B,

- (a)  $s(M^{LB}) = \Theta(\sqrt{n})$ ;
- (b)  $s(M^G) = \Theta(\sqrt{n})$ ;
- (c)  $s(M^*) = \Theta(\sqrt{n})$ .

*Proof.* (a) We observe that  $E_{\tilde{W}}^{LB}(M^{LB})$  is not greater than the minimum expected lower bound obtainable under the assumption that  $c_i = c^U$  and  $s_i = s^L$  for all  $i \in M$ . As we have seen in Section 2, the latter value tends to  $2\sqrt{q^U n\mu}$  as  $n \rightarrow \infty$ . Hence, we may choose any  $q^* > q^U$  to insure that for  $n$  sufficiently large

$$q^L s(M^{LB}) + \frac{n\mu}{s(M^{LB})} \leq E_{\tilde{W}}^{LB}(M^{LB}) \leq 2\sqrt{q^* n\mu},$$

which implies  $C\sqrt{n} \leq s(M^{LB}) \leq C'\sqrt{n}$  for

$$C = \frac{\sqrt{q^*} - \sqrt{q^* - q^L}}{q^L} \sqrt{\mu}, \quad C' = \frac{\sqrt{q^*} + \sqrt{q^* - q^L}}{q^L} \sqrt{\mu}.$$

This proves part (a) of the lemma.

(b) For  $n$  sufficiently large, we have that

$$\begin{aligned}
 q^L s(M^G) + \frac{n\mu}{s(M^G)} &\leq E_{\tilde{w}}^{LB}(M^G) \\
 &\leq E_{\tilde{w}}^{LB}(M^{LB}) + c^U \quad (\text{by Lemma 5}) \\
 &\leq 2\sqrt{q^* n\mu} + c^U = \theta(\sqrt{n}).
 \end{aligned}$$

Part (b) of the lemma now follows by a similar argument as used above.

(c) For  $n$  sufficiently large, we have that

$$\begin{aligned}
 q^L s(M^*) + \frac{n\mu}{s(M^*)} &\leq E_{\tilde{w}}^{LB}(M^*) \\
 &\leq E_{\tilde{w}}^*(M^*) \\
 &\leq E_{\tilde{w}}^{LS}(M^{LB}) \\
 &\leq q^U s(M^{LB}) + EC^{LS}(M^{LB}, p) \\
 &\leq q^U s(M^{LB}) + n\mu/s(M^{LB}) + E_{\tilde{p}_{\max}}/s^L \quad (\text{cf. (8)}) \\
 &\leq q^U C'\sqrt{n} + n\mu/C\sqrt{n} + o(\sqrt{n}) = \theta(\sqrt{n}) \quad (\text{by Lemma 1(a)}).
 \end{aligned}$$

Part (c) of the lemma follows immediately.  $\square$

The following lemma presents an asymptotic upper bound on the heuristic solution value and an asymptotic lower bound on the optimal solution value, which hold with probability arbitrarily close to 1.

**LEMMA 7.** *Under Assumptions A and B and for every  $\epsilon > 0$ , there exist constants  $C > 0$  and  $C' > 0$  such that for  $n$  sufficiently large*

- (a)  $P\{\tilde{w}^{LS}(M^G) \leq \tilde{w}^*(M^*) + C + p_{\max}/s^L\} \geq 1 - \epsilon;$
- (b)  $P\{\tilde{w}^*(M^*) \geq C'\sqrt{n}\} \geq 1 - \epsilon.$

*Proof.* (a) We observe that

$$\begin{aligned}
 w^{LS}(M^G) &\leq c(M^G) + P/s(M^G) + p_{\max}/s^L \quad (\text{cf. (8)}) \\
 &= E_{\tilde{w}}^{LB}(M^G) + (P-n\mu)/s(M^G) + p_{\max}/s^L
 \end{aligned}$$

$$\begin{aligned}
&\leq E \tilde{w}^{LB}(M^{LB}) + (P-n\mu)/s(M^G) + c^U + p_{\max}/s^L \quad (\text{by Lemma 5}) \\
&\leq w^*(M^*) + ((n\mu-P)/s(M^*) + (P-n\mu)/s(M^G)) + c^U + p_{\max}/s^L.
\end{aligned}$$

Let us examine the second term of this upper bound in more detail:

$$\left| \frac{n\mu-P}{s(M^*)} + \frac{P-n\mu}{s(M^G)} \right| \leq |P-n\mu| \left( \frac{1}{s(M^*)} + \frac{1}{s(M^G)} \right).$$

Under Assumption A,  $Ep_j^2 = \sigma^2$  is finite. By the Central Limit Theorem, we have that for every  $\varepsilon > 0$  and for  $n$  sufficiently large

$$P\left\{ \left| \frac{P-n\mu}{\sigma\sqrt{n}} \right| \leq \phi^{-1}(1-\frac{1}{2}\varepsilon) \right\} \geq 1 - \varepsilon,$$

where  $\phi$  denotes the standard normal distribution function. This observation and Lemma 6(b,c) imply that for every  $\varepsilon > 0$  there exists a constant  $C > 0$  such that for  $n$  sufficiently large

$$P\left\{ \left| \frac{n\mu-P}{s(M^*)} + \frac{P-n\mu}{s(M^G)} \right| \leq C \right\} \geq 1 - \varepsilon.$$

This proves part (a) of the lemma.

(b) We choose a constant  $\gamma \in (0,1)$  and note that

$$w^*(M^*) \geq w^{LB}(M^*) \geq q^L s(M^*) + \frac{P-\gamma n\mu}{s(M^*)} + \frac{\gamma n\mu}{s(M^*)}.$$

Since

$$P\left\{ \frac{P-\gamma n\mu}{s(M^*)} > 0 \right\} = P\{P-n\mu > (\gamma-1)n\mu\},$$

Assumption A and the Central Limit Theorem imply that this probability is arbitrarily close to 1 for  $n$  sufficiently large. Part (b) of the lemma now follows from the above in combination with Lemma 6(c).  $\square$

We finally obtain the desired result.

**THEOREM 7.** *Under Assumptions A and B,*

$$\text{plim}_{n \rightarrow \infty} \frac{\tilde{w}^{LS}(M^G)}{\tilde{w}^*(M^*)} = 1.$$

*Proof.* Lemmas 1(b) and 7 imply that  $\text{plim}_{n \rightarrow \infty} \tilde{w}^{LS}(M^G)/\tilde{w}^*(M^*) \leq 1$ . The obser-

vation that  $w^{LS}(M^G) \geq w^*(M^*)$  for every realization of  $p$  establishes the theorem.  $\square$

In contrast to Theorem 4, all we can say about the asymptotic form of  $\tilde{w}^*(M^*)$  and  $\tilde{w}^{LS}(M^G)$  is that for every  $\varepsilon > 0$  there exist constants  $C > 0$  and  $C' > 0$  such that for  $n$  sufficiently large

$$P\{C\sqrt{n} \leq \tilde{w}^*(M^*) \leq \tilde{w}^{LS}(M^G) \leq C'\sqrt{n}\} \geq 1 - \varepsilon.$$

## 4. CONCLUDING REMARKS

In this final section we review some extensions of the model investigated in Sections 2 and 3 and possible applications of our approach to other multi-stage decision problems.

A natural extension of the job shop scheduling problem is to assume that all jobs have a *common due date*  $d$  and that the second stage objective is to minimize the maximum job *tardiness*, i.e. the maximum amount by which any completion time exceeds the due date. If, in the identical machine model, one attempts to design a first stage heuristic by minimizing the expected lower bound

$$cm + E\max\{\frac{P}{m} - d, 0\},$$

then  $m^{LB}$  is given by the solution to the integral equation

$$m = \sqrt{\int_{dm}^{\infty} \text{PDF}(P)/c},$$

which specializes to  $m = \sqrt{EP/c}$  if  $d = 0$ .

There is an interesting difference between this first stage heuristic and those in Sections 2 and 3. The latter heuristics remain the same when the random processing times are simply replaced by their expected values. The above heuristic, however, makes use of distributional information on the processing times, due to the nonnegativity restriction on the lower bound for the second stage cost.

In the case that  $d = o(\sqrt{n})$ , it can be proved that choosing  $m^{LB}$  as in Section 2 yields a heuristic that is still asymptotically optimal in expectation and in probability, and this result extends immediately to the uniform machine model. This is not too surprising: if  $d$  grows more slowly than the expected schedule length, the probability that at least one job is tardy tends to 1, so that the asymptotic difference between maximum tardiness and maximum completion time can be ignored.

The case that  $d = \Theta(\sqrt{n})$  is therefore much more interesting, in particular since it is a first step towards a *multi-period* extension of the model. Here the objective would be to estimate the amount of resources required for a multi-period production plan under certain assumptions on the possible over-

flow of work from one period to the next one. In that context, it is natural to allow true recourse decisions such as the acquisition of additional resources at a higher price if the original estimate turns out to be inadequate.

Another, quite simple, extension would be to treat the number of jobs to be executed as a random variable as well.

The techniques developed in this paper to design and analyze asymptotically optimal stochastic programming heuristics can probably be applied to other multi-stage decision problems. It seems likely that such results can be obtained most easily when the optimal final stage objective value has a simple asymptotic form and when a heuristic exists that yields asymptotically the same value. We are currently working on two- and three-stage formulations of vehicle routing system design problems, for which such results do exist [14]. We expect that it will be possible to capture the conditions under which this approach is successful in a general framework, containing several special cases of practical interest.

The primary purpose of this paper has been to show that the quality of hierarchical planning systems can be analyzed in a more rigorous fashion than has been customary so far. They are, after all, nothing more than heuristics to solve multi-stage stochastic programs. Given the obvious intractability of such problems, precise statements about the performance of approximation algorithms for their solution are of immediate interest to researchers and practitioners.

## APPENDIX

In this appendix we collect together scattered results from the literature to establish the following diagram of logical implications concerning a sequence of independent, identically distributed, nonnegative random variables (here, processing times)  $(p_1, \dots, p_n)$  with common distribution function  $F_p$ , as  $n \rightarrow \infty$ .

$$\begin{array}{ccc}
 E p_{\tilde{j}}^2 < \infty & \Leftrightarrow (1) & \frac{p_{\max}}{\sqrt{n}} \xrightarrow{as} 0 \\
 & \Downarrow (2) & \Downarrow (4) \\
 \frac{E p_{\max}}{\sqrt{n}} \rightarrow 0 & \Rightarrow (3) & \frac{p_{\max}}{\sqrt{n}} \xrightarrow{P} 0
 \end{array}$$

Here, we assume an underlying probability space  $(\Omega, \mathcal{F}, P)$  [2,p.22] and use the following notation:

- $p_{\max} = \max\{p_1, \dots, p_n\}$ ;
- $\xrightarrow{as}$ : convergence *almost surely or with probability 1*;
- $\xrightarrow{P}$ : convergence *in probability*.

The latter two notions have been defined in Section 2.

$$(1) \quad E p_{\tilde{j}}^2 < \infty$$

$$\Leftrightarrow \forall \varepsilon > 0: \sum_{n=1}^{\infty} P\{p_{\tilde{j}}^2 \geq \varepsilon n\} < \infty \quad [2, p.44, Ex.4]$$

$$\Leftrightarrow \forall \varepsilon > 0: P\{p_{\tilde{j}}^2 \geq \varepsilon n \text{ infinitely often}\} = 0 \quad [6, p.210, Lm.4.2.1]$$

$$\Leftrightarrow \forall \varepsilon > 0: P\{p_{\max}^2 \geq \varepsilon n \text{ infinitely often}\} = 0 \quad [6, p.212, Th.4.2.1]$$

$$\Leftrightarrow \forall \varepsilon > 0: P\left\{\frac{p_{\max}}{\sqrt{n}} \geq \varepsilon \text{ infinitely often}\right\} = 0$$

$$\Leftrightarrow \frac{p_{\max}}{\sqrt{n}} \xrightarrow{as} 0 \quad [2, p.73, Th.4.2.2].$$

$$(2) \quad E p_{\tilde{j}}^2 < \infty$$

$$\Rightarrow \lim_{y \rightarrow \infty} \int_y^{\infty} x^2 dF_p(x) = 0 \Rightarrow \lim_{y \rightarrow \infty} y^2 (1 - F_p(y)) = 0.$$



Hence, for any  $\varepsilon > 0$ , we have [11] that

$$\begin{aligned}
 & \frac{E p_{\tilde{\lambda}_{\max}}}{\sqrt{n}} \\
 &= \frac{1}{\sqrt{n}} \int_0^\infty x dF_{\tilde{p}_{\max}}(x) \\
 &= \frac{1}{\sqrt{n}} \int_0^\infty x dF_{\tilde{p}}^n(x) \quad (\text{using independence of the } \tilde{p}_j) \\
 &= \frac{-1}{\sqrt{n}} \int_0^\infty x d(1 - F_{\tilde{p}}^n(x)) \\
 &= \frac{1}{\sqrt{n}} \int_0^\infty (1 - F_{\tilde{p}}^n(x)) dx \quad (\text{integrating by parts}) \\
 &= \int_0^\infty (1 - F_{\tilde{p}}^n(x\sqrt{n})) dx \quad (\text{substituting } x\sqrt{n} \text{ for } x) \\
 &= \int_0^\varepsilon (1 - F_{\tilde{p}}^n(x\sqrt{n})) dx + \int_\varepsilon^\infty (1 - F_{\tilde{p}}^n(x\sqrt{n})) dx \\
 &\leq \varepsilon + \int_\varepsilon^\infty (1 - (1 - \frac{\varepsilon^2}{nx^2})^n) dx.
 \end{aligned}$$

Since  $1-y > e^{-y/(1-y)}$  for  $0 < y < 1$ , we have that

$$1 - (1 - \frac{\varepsilon^2}{nx^2})^n < 1 - e^{-\varepsilon^2/(x^2 - \varepsilon^2/n)} \leq 1 - e^{-\varepsilon^2/(x^2 - \varepsilon^2/2)},$$

which is an integrable function on  $(\varepsilon, \infty)$ . It follows that

$$\lim_{n \rightarrow \infty} \int_\varepsilon^\infty (1 - (1 - \frac{\varepsilon^2}{nx^2})^n) dx = \int_\varepsilon^\infty (1 - e^{-\varepsilon^2/x^2}) dx < \varepsilon.$$

As a consequence,

$$\lim_{n \rightarrow \infty} \frac{E p_{\tilde{\lambda}_{\max}}}{\sqrt{n}} = 0.$$

It can be proved [11] that, conversely,

$$\lim_{n \rightarrow \infty} \frac{E p_{\tilde{\lambda}_{\max}}}{\sqrt{n}} = 0 \Rightarrow \lim_{y \rightarrow \infty} y^2 (1 - F_{\tilde{p}}(y)) = 0.$$

(3) This is a standard consequence of the Chebychev inequality [2,p.67,Th.4.1.4].

(4) This is also a standard result [2,p.66,Th.4.1.2].

## ACKNOWLEDGMENTS

This research was partially supported by NSF Grant ENG-7826500 to the University of Pennsylvania, by NATO Special Research Grant 9.2.02 (SRG.7) and by NATO Research Grant 1575. The authors gratefully acknowledge useful discussions with L. de Haan and L. Stougie.

## REFERENCES

1. J. BEARDWOOD, H.J. HALTON, J.M. HAMMERSLEY (1959) The shortest path through many points. *Proc. Cambridge Phil. Soc.* 55,299-327.
2. K.L. CHUNG (1974) *A Course in Probability Theory*, Second Edition, Academic Press, New York.
3. M.A.H. DEMPSTER, M.L. FISHER, L. JANSEN, B.J. LAGEWEG, J.K. LENSTRA, A. H.G. RINNOOY KAN (1981) Analytical evaluation of hierarchical planning systems. *Oper. Res.*, to appear.
4. M.L. FISHER (1980) Worst-case analysis of heuristic algorithms. *Management Sci.* 26,1-17.
5. M.L. FISHER, D.S. HOCHBAUM (1980) Probabilistic analysis of the planar K-median problem. *Math. Oper. Res.* 5,27-34.
6. J. GALAMBOS (1978) *The Asymptotic Theory of Extreme Order Statistics*, Wiley, New York.
7. M.R. GAREY, R.L. GRAHAM, D.S. JOHNSON (1978) Performance guarantees for scheduling algorithms. *Oper. Res.* 26,3-21.
8. M.R. GAREY, D.S. JOHNSON (1979) *Computers and Intractability: a Guide to the Theory of NP-Completeness*, Freeman, San Francisco.
9. R.L. GRAHAM (1966) Bounds for certain multiprocessing anomalies. *Bell System Tech. J.* 45,1563-1581.
10. R.L. GRAHAM, E.L. LAWLER, J.K. LENSTRA, A.H.G. RINNOOY KAN (1979) Optimization and approximation in deterministic sequencing and scheduling: a survey. *Ann. Discrete Math.* 5,287-326.
11. L. DE HAAN (1980) Personal communication.
12. D. HOCHBAUM, J.M. STEELE (1981) Steinhaus' geometric location problem for random samples in the plane. *J. Appl. Probab.*, to appear.
13. R.M. KARP (1972) Reducibility among combinatorial problems. In: R.E. MILLER, J.W. THATCHER (eds.) (1972) *Complexity of Computer Computations*, Plenum, New York, 85-103.
14. A. MARCHETTI SPACCAMELA, A.H.G. RINNOOY KAN, L. STOUGIE (1981) Hierarchical vehicle routing. To appear.
15. J.M. STEELE (1980) Subadditive Euclidean functionals and non-linear growth in geometric probability. Department of Statistics, Stanford University.

ONTVANGEN 15 JULI 1901