

# Investigating the Recoverable Robust Single Machine Scheduling Problem Under Interval Uncertainty

Matthew Bold<sup>\*1</sup> and Marc Goerigk<sup>2</sup>

<sup>1</sup>STOR-i Centre for Doctoral Training, Lancaster University, Lancaster, UK

<sup>2</sup>Network and Data Science Management, University of Siegen, Siegen, Germany

## Abstract

We investigate the recoverable robust single machine scheduling problem under interval uncertainty. In this setting, jobs have first-stage processing times  $\mathbf{p}$  and second-stage processing times  $\mathbf{q}$  and we aim to find a first-stage and second-stage schedule with a minimum combined sum of completion times, such that at least  $\Delta$  jobs share the same position in both schedules.

We provide positive complexity results for some important special cases of this problem, as well as derive a 2-approximation algorithm to the full problem. Computational experiments examine the performance of an exact mixed-integer programming formulation and the approximation algorithm, and demonstrate the strength of a proposed polynomial time greedy heuristic.

**Keywords:** scheduling; optimisation under uncertainty; recoverable robustness

## 1 Introduction

In this paper, we consider a recoverable robust version of a single machine scheduling problem in which  $n$  jobs must be scheduled on a single machine without preemption, such that the total flow time, i.e. the sum of job completion times, is minimised. Under the  $\alpha|\beta|\gamma$  scheduling notation introduced by Graham et al. (1979), the nominal problem is denoted as  $1||\sum C_i$ . In reality, job processing times are usually subject to some degree of uncertainty. When this is the case, it is important to account for this uncertainty already in the construction of scheduling solutions. Here, we consider the case where the job processing times are assumed to lie within specified intervals and examine the recoverable robust optimisation that arises from assuming a two-stage decision process.

The nominal problem can be stated as follows. Given a set of jobs  $N = \{1, \dots, n\}$  with processing times  $\mathbf{p} = (p_1, \dots, p_n)$ , find an ordering of the jobs  $j \in N$  such that the sum of completion times is minimised. In other words, we want to find a permutation  $\sigma$  of the set  $N$ ,

---

<sup>\*</sup>Corresponding author, email: m.bold1@lancaster.ac.uk

such that  $\sum_{i \in N} (n+1-i)p_{\sigma(i)}$  is minimised. This nominal problem is easy to solve using the shortest processing time (SPT) rule of sorting the jobs by non-decreasing processing times. This problem can also be modelled as the following assignment problem with non-general cost structure:

$$\min \sum_{i \in N} \sum_{j \in N} p_j (n+1-i) x_{ij} \quad (1)$$

$$\text{s.t. } \sum_{i \in N} x_{ij} = 1 \quad \forall j \in N \quad (2)$$

$$\sum_{j \in N} x_{ij} = 1 \quad \forall i \in N \quad (3)$$

$$x_{ij} \in \{0, 1\} \quad \forall i, j \in N, \quad (4)$$

where  $x_{ij} = 1$  if job  $j$  is scheduled in position  $i$ , and  $x_{ij} = 0$  otherwise.

The problem we consider in this paper considers a two-stage decision process. Suppose that in the first-stage, the jobs  $j \in N$  have processing times given by  $\mathbf{p} = (p_1, \dots, p_n)$ , and in the second-stage they have different processing times given by  $\mathbf{q} = (q_1, \dots, q_n)$ . We aim to find a first-stage schedule and second-stage schedule that minimises the combined cost of these two schedules, such that the position of at least  $\Delta$  jobs remain unchanged between the two schedules, i.e. only up to  $n - \Delta$  jobs change position. This problem can be written as

$$\min \sum_{i \in N} \sum_{j \in N} p_j (n+1-i) x_{ij} + \sum_{i \in N} \sum_{j \in N} q_j (n+1-i) y_{ij} \quad (\text{RecSMSP})$$

$$\text{s.t. } |X \cap Y| \geq \Delta$$

$$\mathbf{x}, \mathbf{y} \in \mathcal{X}$$

where  $X = \{(i, j) \in N \times N : x_{ij} = 1\}$  is the assignment corresponding to the first-stage schedule,  $Y = \{(i, j) \in N \times N : y_{ij} = 1\}$  is the assignment corresponding to the second-stage schedule, and  $\mathcal{X} = \{\mathbf{x} \in \{0, 1\}^{n \times n} : (2), (3)\}$  is the set of feasible schedules. Condition  $|X \cap Y| \geq \Delta$  is equivalent to demanding that

$$\sum_{i \in N} \sum_{j \in N} x_{ij} y_{ij} \geq \Delta.$$

(RecSMSP) can be modelled as a mixed-integer program (MIP) following the linearisation of this constraint. This is achieved with the introduction of an additional set of  $\mathbf{z}$  variables, and the following constraints:

$$z_{ij} \leq x_{ij} \quad \forall i, j \in N$$

$$z_{ij} \leq y_{ij} \quad \forall i, j \in N$$

$$\sum_{i \in N} \sum_{j \in N} z_{ij} \geq \Delta$$

$$z_{ij} \in \{0, 1\} \quad \forall i, j \in N$$

Observe how the  $\mathbf{z}$  variables represent the shared assignments of the first and second-stages, i.e.  $z_{ij} = 1$  if job  $j$  is assigned to position  $i$  in both the first and second-stage schedules.

This problem can be considered to be a recoverable robust optimisation problem with interval uncertainty (see Liebchen et al. (2009)). A recoverable robust optimisation problem consist of two-stages. A full first-stage solution must be determined under the problem uncertainty, before an adversary chooses a worst-case realisation of the uncertain data. Then, in response to this realisation, the first-stage solution can be recovered, i.e. amended in some limited way, to obtain a second-stage solution. For a general survey on robust discrete optimisation problems, we refer the reader to Kasperski and Zieliński (2016a).

We consider exactly this recoverable robust problem in the setting where the second-stage job processing times  $\mathbf{q}$ , are uncertain, but known to lie within an interval (or box) uncertainty set given by

$$\mathcal{U} = \{\mathbf{q} \in \mathbb{R}_+^n : q_j \in [\hat{q}_j - \bar{q}_j, \hat{q}_j + \bar{q}_j], j \in N\}.$$

That is, each job  $j \in N$  has a nominal second-stage processing time given by  $\hat{q}_j$ , but can deviate from this nominal value by up to  $\bar{q}_j$ . The worst-case scenario for this uncertainty set occurs when each job simultaneously achieves its worst-case processing time, i.e.  $\hat{q}_j + \bar{q}_j$  for each  $j \in N$ . Denoting the first-stage duration of  $j \in N$  as  $p_j$ , and its worst-case second-stage duration as  $q_j$ , we get (RecSMSP).

There exists a number of papers that consider various discrete optimisation problems in the setting of recoverable robust optimisation with interval uncertainty. The resulting problems they study therefore also contain the same intersection constraints that we consider here. Our work in this paper extends this framework to the context of single-machine scheduling. Büsing (2012) studies recoverable robust shortest path problems, whilst Hradovich et al. (2017a,b) investigate the recoverable robust spanning tree problem. Kasperski and Zieliński (2017) consider the recoverable robust selection problem under both discrete and interval uncertainty and, most recently, Goerigk et al. (2021) consider the recoverable robust travelling salesman problem. The single-machine scheduling problem we consider is an assignment problem with a specific cost structure, and therefore Fischer et al. (2020) is the paper that is most closely related to our work. In this paper, the authors examine the complexity of the recoverable robust assignment problem with interval uncertainty. Amongst other results, they show that this problem is W[1]-hard with respect to  $\Delta$  and  $n - \Delta$ . Even though they use only 0-1 costs for this reduction, note that this hardness result does not extend to the scheduling cost structure that we consider here.

As a brief comment on other papers that consider robust assignment problems, Deĭneko et al. (2006) studies the problem in the context of discrete scenarios and Pereira and Averbakh (2011) considers interval uncertainty in combination with the regret criterion.

The single machine scheduling problem with the objective of minimising the (weighted) sum of completion times under uncertain job processing times is well-studied, particularly for the case of discrete uncertain scenarios. Daniels and Kouvelis (1995), Kouvelis and Yu (1997), Yang and Yu (2002) and Aloulou and Della Croce (2008) show that even for just two discrete scenarios, robust versions of this problem are NP-hard, whilst Mastrolilli et al. (2013) show that no polynomial-time approximation algorithm exists. Zhao et al. (2010) propose a cutting plane algorithm to solve the problem. More recently, Kasperski and Zieliński (2016b) considered the problem for the ordered weighted averaging (OWA) criterion, of which classical robustness is a special case, and Kasperski and Zieliński (2019) considered the problem for the value at risk (VaR) and conditional value at risk (CVaR) criteria.

The case of interval uncertainty has also received a lot of attention. Daniels and Kouvelis (1995) characterised a set of dominance relations between jobs in an optimal schedule in the

case of interval uncertainty. Lebedev and Averbakh (2006) showed that the problem is NP-hard for regret robustness, whilst Montemanni (2007) presented a compact MIP that was shown to be able to solve instances involving up to 45 jobs. Kasperski and Zieliński (2008) showed that the regret problem is 2-approximable when the corresponding nominal problem is solvable in polynomial time. For a survey of robust single machine scheduling for both discrete and interval uncertainty, see Kasperski and Zielinski (2014).

For robust single machine scheduling in the context of budgeted uncertainty, Lu et al. (2014) presented an MIP and heuristic to solve the problem, before Bougeret et al. (2019) examined its complexity. Most recently, Bold and Goerigk (2020) considered the recoverable robust problem under a different similarity measure to the one considered in this paper.

The structure and contributions of this paper are as follows. Section 2 presents a number of positive results for the problem, including an efficient method for optimal scheduling when a set of jobs which must have the same first and second-stage positions is given. Section 3 then presents a 2-approximation algorithm and greedy heuristic. The performance of the exact MIP formulation as well as the heuristics we present are examined experimentally in Section 4, before concluding remarks are made in Section 5.

## 2 Problem properties

We begin our analysis of recoverable robust single machine scheduling problem by considering the following question: Given a set  $M \subseteq N$  of jobs that must be scheduled in the same position in both the first and the second stage, what is the best possible solution? That is, we consider the problem of calculating

$$\begin{aligned}
 f(M) = \min \quad & \sum_{i \in N} \sum_{j \in N} p_j(n+1-i)x_{ij} + \sum_{i \in N} \sum_{j \in N} q_j(n+1-i)y_{ij} && \text{(RecFix)} \\
 \text{s.t.} \quad & x_{ij} = y_{ij} && \forall i \in N, j \in M \\
 & \mathbf{x}, \mathbf{y} \in \mathcal{X} &&
 \end{aligned}$$

Note that a solution  $(\mathbf{x}, \mathbf{y})$  of (RecFix) may intersect on more jobs than just those in the set  $M$ . Also note that if  $|M| \geq \Delta$ , then the solution to this problem is feasible to (RecSMSP). Therefore, the optimal objective value of (RecSMSP) is equal to  $\min\{f(M) : |M| \geq \Delta\}$ . In Algorithm 1, we show how to calculate the value  $f(M)$  in polynomial time.

The following example demonstrates the implementation of this ordering rule. Consider the data shown in Table 1 and suppose that  $M = \{3, 4\}$ , i.e. jobs 3 and 4 must share the same position in the first and second-stage schedules. Then  $N \setminus M = \{1, 2, 5\}$ , i.e., jobs 1, 2 and 5 can be assigned different positions in the two schedules. The sorted first and second-stage processing times of jobs in  $N \setminus M$  are  $\mathbf{a} = (2, 3, 5)$  and  $\mathbf{b} = (1, 4, 6)$  respectively, and the joint processing times for the jobs in  $M$  are  $\mathbf{c} = (14, 6)$ . Then  $\mathbf{d} = \mathbf{a} + \mathbf{b} = (3, 7, 11)$  and  $\mathbf{e} = (3, 6, 7, 11, 14)$  as the merged and sorted vector processing times. This corresponds to the first-stage schedule  $\sigma_1 = (5, 4, 2, 1, 3)$ , and the second-stage schedule  $\sigma_2 = (2, 4, 1, 5, 3)$ . Note that, as required, jobs 3 and 4 are placed in the same position in both the first and second-stage schedules.

We now show that Algorithm 1 does indeed give an optimal solution to problem (RecFix).

**Theorem 1.** *Algorithm 1 calculates  $f(M)$  for any  $M \subseteq N$ .*

---

**Algorithm 1** Evaluation method for  $f(M)$ 


---

- 1: **procedure** EVAL( $\mathbf{p}, \mathbf{q}, M$ )
  - 2:     Set  $\mathbf{a} = (a_j)_{j \in N \setminus M}$  to be the vector of values  $p_j$ ,  $j \in N \setminus M$ , sorted by non-decreasing values
  - 3:     Set  $\mathbf{b} = (b_j)_{j \in N \setminus M}$  to be the vector of values  $q_j$ ,  $j \in N \setminus M$ , sorted by non-decreasing values
  - 4:     Set  $\mathbf{c} = (c_j)_{j \in M}$  to be the vector of values  $p_j + q_j$ ,  $j \in M$
  - 5:     Set  $\mathbf{d} = \mathbf{a} + \mathbf{b} = (a_1 + b_1, \dots, a_{N \setminus M} + b_{N \setminus M})$
  - 6:     Let  $\mathbf{e} = (e_j)_{j \in N}$  be the vector found by concatenating vectors  $\mathbf{c}$  and  $\mathbf{d}$  and sorting by non-decreasing values
  - 7:     **return**  $\sum_{i \in N} (n + 1 - i)e_i$
  - 8: **end procedure**
- 

$j$	1	2	<b>3</b>	<b>4</b>	5
$p_j$	5	3	<b>5</b>	<b>1</b>	2
$q_j$	4	1	<b>9</b>	<b>5</b>	6
$p_j + q_j$	9	4	<b>14</b>	<b>6</b>	8

Table 1: Example problem data. Columns of  $M = \{3, 4\}$  are in bold.

*Proof.* Let any  $M \subseteq N$  be given. We denote these jobs as  $M = \{j_1, \dots, j_m\}$  with  $m = |M|$ . Let us first assume that we already know the set of slots  $K = \{i_1, \dots, i_m\} \subseteq N$  into which the jobs will be scheduled. We will then show how to find such a set. We further denote by  $N \setminus M = \{j'_1, \dots, j'_{n-m}\}$  and  $N \setminus K = \{i'_1, \dots, i'_{n-m}\}$  the sets of jobs and slots that are not part of  $M$  and  $K$ , respectively. Without loss of generality, we assume that  $i_1 < i_2 < \dots < i_m$  and  $i'_1 < i'_2 < \dots < i'_{n-m}$ . The resulting problem thus decomposes into the two subproblems

$$\min_{\mathbf{z} \in \mathcal{X}} \sum_{k=1}^m \sum_{\ell=1}^m (n + 1 - i_k)(p_{j_\ell} + q_{j_\ell})z_{k\ell} + \min_{\mathbf{x}, \mathbf{y} \in \mathcal{X}} \sum_{k=1}^{n-m} \sum_{\ell=1}^{n-m} (n + 1 - i'_k)(p_{j'_\ell}x_{k\ell} + q_{j'_\ell}y_{k\ell})$$

where with slight abuse of notation,  $\mathcal{X}$  denotes the set of assignments with suitable dimension. Note that independent of the choice of  $K$ , we can find optimal solutions  $\mathbf{z}^*$ ,  $\mathbf{x}^*$  and  $\mathbf{y}^*$ . Sorting jobs  $j \in M$  by non-decreasing joint processing times  $p_j + q_j$  gives  $\mathbf{z}^*$ . We denote the resulting vector of joint processing times as  $\mathbf{c} = (c_j)_{j \in M}$ . Sorting jobs  $j \in N \setminus M$  by non-decreasing  $p_j$  gives  $\mathbf{x}^*$ , and by non-decreasing  $q_j$  gives  $\mathbf{y}^*$ . We denote the resulting vector of sums of the sorted processing times as  $\mathbf{d} = (d_j)_{j \in N \setminus M}$ .

Let us fix these assignments accordingly and consider how to find set  $K$ . This is equivalent to assigning the positions  $i_1, \dots, i_m$  and  $i'_1, \dots, i'_{n-m}$  to slots in  $N$ . Hence, the value  $f(M)$  is equal to

$$\min_{\tau \in \mathcal{X}} \sum_{i \in N} \sum_{\ell=1}^m (n + 1 - i\tau_{i\ell})c_\ell + \sum_{i \in N} \sum_{\ell=1}^{n-m} (n + 1 - i\tau_{i, \ell+m})d_\ell$$

An optimal solution to this problem can be found by assigning the positions according to non-decreasing processing times of their associated jobs, i.e., to sort the concatenated vector of processing times consisting of  $\mathbf{c}$  and  $\mathbf{d}$ . Algorithm 1 is exactly this solution procedure.  $\square$

Therefore, if it is known which set of jobs  $M$  with cardinality  $\Delta$  must share a position in

the first and second-stage schedule, the remaining problem can be solved in  $O(n \log n)$  time. This immediately gives the following result.

**Corollary 2.** *For a constant value of  $\Delta$ , (RecSMSP) can be solved in polynomial time  $O(n^{\Delta+1} \log n)$ .*

*Proof.* This can be seen by simply enumerating the  $\binom{n}{\Delta} \sim O(n^\Delta)$  possible ways of choosing a set of  $\Delta$  jobs to share first and second-stage assignments. For each of the  $O(n^\Delta)$  ways of fixing  $\Delta$  jobs, the remaining problem can be solved using Algorithm 1 in  $O(n \log n)$  time. Hence, the overall complexity is given by  $O(n^{\Delta+1} \log n)$ , i.e. polynomial for fixed  $\Delta$ .  $\square$

As observed in Fischer et al. (2020), it is straightforward to see that problem (RecSMSP) can be solved in polynomial time for constant value of  $\Delta$  by enumerating all possibilities for the intersection set  $|X \cap Y|$ . Note that whilst this approach would require us to check  $\binom{n^2}{\Delta}$  many candidates for general cost functions, this number is reduced to  $\binom{n}{\Delta}$  in our case.

We now consider a special case of (RecSMSP) where there are a constant number  $k$  of possible job processing times, i.e.  $p_j, q_j \in \{d_1, \dots, d_k\}$ , where  $d_1, \dots, d_k \in \mathbb{R}$  for constant  $k$ . Note that this results in at most  $k^2$  possible combinations of first and second-stage costs  $p_j$  and  $q_j$ , that is  $k^2$  possible job types. Furthermore,  $f(M) = f(M')$  for any two choices  $M, M' \subseteq N$  that contain the same number of each job type. We can conclude the following result.

**Corollary 3.** *Let  $p_j, q_j \in \{d_1, \dots, d_k\}$  for a constant value of  $k$ . Then, (RecSMSP) can be solved in polynomial time  $O(n^{k^2})$*

*Proof.* We consider the number of ways there are to choose  $\Delta$  jobs from  $k^2$  different job types. Observe that for all but the final job type, there are  $\Delta + 1$  ways of choosing up to  $\Delta$  jobs of that type. Having chosen the number of jobs from the first  $k^2 - 1$  job types, the number of jobs of the final type is simply equal to the number of jobs remaining. Hence there are  $O((\Delta + 1)^{k^2 - 1})$  ways to choose the  $\Delta$  jobs. For each of these choices, we can use the ordering rule presented in Algorithm 1 to solve the resulting problem. Note that sorting  $n$  bounded values is possible in  $O(n)$  time. Hence (RecSMSP) with  $k$  possible job processing times is solvable in  $O((\Delta + 1)^{k^2 - 1} \cdot n) \sim O(n^{k^2})$  time.  $\square$

### 3 A 2-approximation algorithm

In this section, we present an approximate solution to (RecSMSP) and prove that this solution has a guaranteed worst-case approximation ratio of 2.

**Theorem 4.** *A solution to the problem*

$$\begin{aligned} f(N) = \min \quad & \sum_{i \in N} \sum_{j \in N} p_j(n + 1 - i)x_{ij} + \sum_{i \in N} \sum_{j \in N} q_j(n + 1 - i)y_{ij} & (\text{UB-SMSP}) \\ & |X \cap Y| = n \\ & \mathbf{x}, \mathbf{y} \in \mathcal{X} \end{aligned}$$

*provides a 2-approximation to (RecSMSP).*

Observe that (UB-SMSP) is equivalent to (RecSMSP) with complete intersection of the first and second-stage solutions. It can thus be simplified to

$$\min_{\mathbf{x} \in \mathcal{X}} \sum_{i \in N} \sum_{j \in N} (p_j + q_j)(n + 1 - i)x_{ij}$$

Note that (UB-SMSP) can be solved in  $O(n \log n)$  time by simply ordering the jobs  $j \in N$  according to non-decreasing  $p_j + q_j$  values.

In the following, we provide a proof for Theorem 4. To this end, we consider a lower bound to (RecSMSP), which is provided by a solution to the problem

$$\left( \min_{\mathbf{x} \in \mathcal{X}} \sum_{i \in N} \sum_{j \in N} p_j(n + 1 - i)x_{ij} \right) + \left( \min_{\mathbf{y} \in \mathcal{X}} \sum_{i \in N} \sum_{j \in N} q_j(n + 1 - i)y_{ij} \right), \quad (\text{LB-SMSP})$$

that is, (RecSMSP) without the intersection constraint. Suppose  $\sigma_p$  and  $\sigma_q$  are orderings of jobs  $j \in N$  by non-decreasing first-stage costs  $p_j$  and non-decreasing second-stage costs  $q_j$ , respectively. Then (LB-SMSP) has objective value equal to

$$\sum_{j \in N} (p_{\sigma_p(j)} + q_{\sigma_q(j)})(n + 1 - j).$$

We compare the upper bound provided by (UB-SMSP) with the lower bound provided by (LB-SMSP) and refer to the values of these upper and lower bounds as  $UB$  and  $LB$ , respectively.

By sorting items by  $\sigma_p$  and  $\sigma_q$  and considering the different positions of specific items between these two sortings, we get a different perspective on what constitutes an instance of (RecSMSP). Specifically, we can consider sorted processing times  $p_1 \leq p_2 \leq \dots \leq p_n$  and  $q_1 \leq q_2 \leq \dots \leq q_n$ , and a permutation  $\pi$ , such that the processing times  $p_j$  and  $q_{\pi(j)}$  belong to the same job. Figure 1 provides an example of a (RecSMSP) considered in this way.

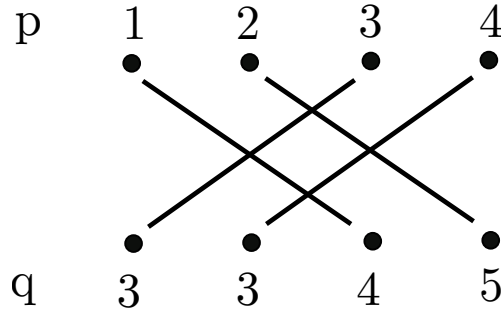


Figure 1: An example instance of (RecSMSP) involving four jobs. First-stage processing times are shown in the top row and second-stage processing times are shown in the bottom row. An edge between first-stage node and second-stage node indicates that these processing times belong to a single job.

Using this description of an instance, we begin by first proving the following structural lemma, which shows that a worst-case instance is of a form where the smallest value of  $\mathbf{p}$  is matched with the largest value of  $\mathbf{q}$ , the second smallest value of  $\mathbf{p}$  is matched with the second-largest value of  $\mathbf{q}$ , and so on.

**Lemma 5.** *Let any instance of (RecSMSP) be given, consisting of sorted vectors  $\mathbf{p}$ ,  $\mathbf{q}$  and a permutation  $\pi$ . Then the ratio  $\frac{UB}{LB}$  for this instance is less or equal to the ratio for the instance where  $\pi$  is replaced by  $\pi'$ , where  $\pi' = (n, n-1, \dots, 1)$ , i.e.,  $\pi'$  is a sorting of indices from largest to smallest. We refer to such an instance as being ‘fully-crossed’.*

*Proof.* Matching a first-stage cost to a second-stage cost represents a job having those respective processing times. Here, we match first and second-stage costs with the objective of maximising the ratio  $\frac{UB}{LB}$ . To this end, observe that  $LB$  is the same for all possible matchings  $\pi$  as the intersection constraint is ignored, and hence finding a matching to maximise the ratio  $\frac{UB}{LB}$  is equivalent to finding a matching that maximises  $UB$ .

Let us consider the dual problem of  $UB$ , given by

$$\begin{aligned} \max \quad & \sum_{i \in N} u_i + v_i \\ \text{s.t.} \quad & u_j + v_i \leq (n+1-i)(p_j + q_{\pi(j)}) \quad \forall i, j \in N \end{aligned}$$

Consider any two jobs, and suppose that first and second-stage costs are labelled so that  $p_1 \leq p_2$  and  $q_1 \leq q_2$ . We compare the objective values between the case when  $p_1$  is matched to  $q_1$  and  $p_2$  is matched to  $q_2$ , i.e. the two jobs are uncrossed (see Figure 2a), and the case when  $p_1$  is matched to  $q_2$  and  $p_2$  is matched to  $q_1$ , i.e. the two jobs are crossed (see Figure 2b).

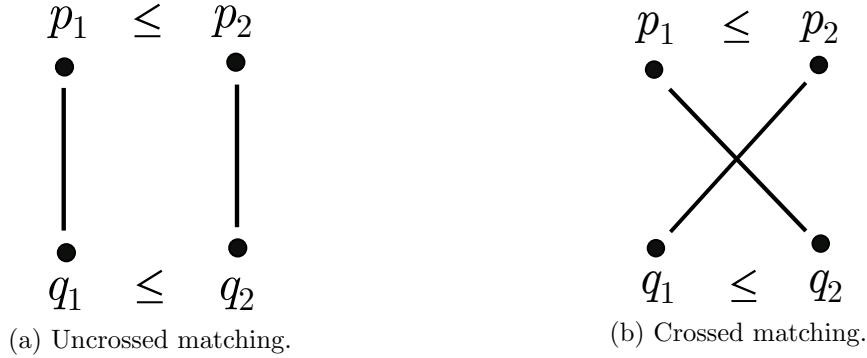


Figure 2: Uncrossed and crossed instances.

Let  $(\mathbf{u}^*, \mathbf{v}^*)$  be an optimal solution to the dual of (UB-SMSP) for the instance in which the two jobs are uncrossed, and let us denote its objective value as  $UB_{\text{uncrossed}}$ . Note that

$$\begin{aligned} u_1^* &= \min_{i \in N} \{(n+1-i)(p_1 + q_1) - v_i^*\} \\ u_2^* &= \min_{i \in N} \{(n+1-i)(p_2 + q_2) - v_i^*\}. \end{aligned}$$

Now consider the dual of (UB-SMSP) for the instance in which the two jobs are crossed. We denote its objective value as  $UB_{\text{crossed}}$  and construct a feasible solution  $(\mathbf{u}, \mathbf{v})$  by setting  $\mathbf{v} = \mathbf{v}^*$ ,

$$\begin{aligned} u_1 &= \min_{i \in N} \{(n+1-i)(p_1 + q_2) - v_i^*\} \\ u_2 &= \min_{i \in N} \{(n+1-i)(p_2 + q_1) - v_i^*\} \end{aligned}$$



and  $u_j = u_j^*$  for all other jobs  $j$ . Since this is a feasible solution, we have  $UB_{\text{crossed}} \geq \sum_{i \in N} u_i + v_i$ .

Now for  $\lambda \in [0, 1]$ , we consider the function  $f_i(\lambda) = (n+1-i)(\lambda(p_1+q_1)+(1-\lambda)(p_2+q_2))-v_i^*$  and define  $g(\lambda) = \min_{i \in N} f_i(\lambda) + \min_{i \in N} f_i(1-\lambda)$ . The minimum of concave functions is concave, and the sum of concave functions is concave, and therefore function  $g$  is concave. Furthermore, it is symmetric with respect to  $\lambda = 0.5$ . Hence,  $g$  is minimised for  $\lambda = 0$  and  $\lambda = 1$ , and thus for all  $\lambda \in [0, 1]$ ,  $g(\lambda) \geq g(0) = g(1)$ .

We therefore conclude that

$$\begin{aligned} UB_{\text{crossed}} - UB_{\text{uncrossed}} &\geq \sum_{i \in N} u_i + v_i - \sum_{i \in N} u_i^* + v_i^* \\ &= u_1 + u_2 - u_1^* - u_2^* \\ &= g(\lambda) - g(0) \geq 0 \end{aligned}$$

where  $\lambda = (p_2 - p_1)/(p_2 + q_2 - p_1 - q_1)$ . Hence, by changing the matching  $\pi$  as described, the value of the upper bound is not decreased. Repeating this process, we find that there is always a fully-crossed worst-case instance as claimed.  $\square$

We are now in a position to prove Theorem 4.

*Proof of Theorem 4.* We suppose that the first and second-stage costs are ordered so that  $p_1 \leq p_2 \leq \dots \leq p_n$  and  $q_1 \leq q_2 \leq \dots \leq q_n$ , and that the jobs are labelled in order of their first-stage costs. Making use of Lemma 5, we restrict our consideration to instances where the matching of first and second-stage costs is fully crossed, i.e. job  $j$  has the  $j$ -th largest first-stage cost  $p_j$ , and the  $(n-j)$ -th largest second-stage cost  $q_{n-j}$ .

We examine the problem of choosing values  $\mathbf{p}$  and  $\mathbf{q}$  to maximise  $\frac{UB}{LB}$ , i.e.

$$\max_{(\mathbf{p}, \mathbf{q}) \in \mathcal{PQ}} \min_{\pi \in \Pi} \frac{\sum_{i \in N} (n+1-i)(p_{\pi(i)} + q_{n-\pi(i)})}{\sum_{i \in N} (n+1-i)(p_i + q_i)}, \quad (5)$$

where  $\pi \in \Pi$  is a permutation of the positions  $i \in N$  such that  $\pi(i)$  is the job scheduled in position  $i$ , and  $\mathcal{PQ} = \{(\mathbf{p}, \mathbf{q}) \in \mathbb{R}_+^{n+n} : p_1 \leq p_2 \leq \dots \leq p_n, q_1 \leq q_2 \leq \dots \leq q_n\}$ .

Normalising the objective function we get

$$\max_{(\mathbf{p}, \mathbf{q}) \in \mathcal{PQ}} \min_{\pi \in \Pi} \sum_{i \in N} (n+1-i)(p_{\pi(i)} + q_{n-\pi(i)}),$$

where  $\overline{\mathcal{PQ}} = \{(\mathbf{p}, \mathbf{q}) \in \mathcal{PQ} : \sum_{i \in N} (n+1-i)(p_i + q_i) = 1\}$ . For fixed  $\mathbf{p}$  and  $\mathbf{q}$ , the inner minimisation problem over the set of permutations  $\pi \in \Pi$  can be written as the following assignment problem

$$\begin{aligned} \min \quad & \sum_{i \in N} \sum_{j \in N} (n+1-i)(p_j + q_{n-j})x_{ij} \\ \text{s.t.} \quad & \sum_{i \in N} x_{ij} = 1 \quad \forall j \in N \\ & \sum_{j \in N} x_{ij} = 1 \quad \forall i \in N \\ & x_{ij} \in \{0, 1\} \quad \forall i, j \in N, \end{aligned}$$



We now proceed to show that there is a feasible solution to this dual problem with an objective value of 2. Since a feasible solution to the dual problem provides an upper bound to the primal problem, finding a feasible dual solution with an objective value of 2 guarantees that the ratio  $\frac{UB}{LB}$  is bounded above by 2, and therefore proves that (UB-SMSP) does indeed provide a 2-approximation to (RecSMSP).

Setting  $\alpha_i = \beta_i = 0$  for all  $i \in N$ , the dual problem becomes

$$\begin{aligned}
& \min \gamma \\
& \text{s.t. } \sum_{i \in N} x_{ij} = 1 \quad \forall j \in N \\
& \quad \sum_{j \in N} x_{ij} = 1 \quad \forall i \in N \\
& \quad (n+1-j)\gamma \geq \sum_{i \in N} (n+1-i)x_{ij} \quad \forall j \in N \quad (*) \\
& \quad (n+1-j)\gamma \geq \sum_{i \in N} (n+1-i)x_{i,n+1-j} \quad \forall j \in N \quad (**) \\
& \quad x_{ij} \geq 0 \quad \forall i, j \in N.
\end{aligned}$$

Observe that constraint (\*\*) can be rewritten as

$$j\gamma \geq \sum_{i \in N} (n+1-i)x_{ij} \quad \forall j \in N.$$

Suppose that  $\gamma = 2$ . Then constraints (\*) and (\*\*) enforce that

$$\begin{aligned}
2(n+1-j) &\geq n+1 - \sum_{i \in N} ix_{ij} \quad \forall j \in N \\
2j &\geq n+1 - \sum_{i \in N} ix_{ij} \quad \forall j \in N,
\end{aligned}$$

that is

$$\sum_{i \in N} ix_{ij} \geq \max\{2j - n - 1, n + 1 - 2j\} \quad \forall j \in N \quad (\dagger)$$

By considering the above constraint, we can determine an approach for generating a feasible dual solution with objective value  $\gamma = 2$ . This approach works as follows: take the positions  $i = n, n-1, \dots, 1$  in decreasing order, and assign jobs  $j \in N$  to them, alternating between the unassigned job with the largest index and the unassigned job with the smallest index.

$$\max\{2j - n - 1, n + 1 - 2j\} \quad \left| \begin{array}{cccccc} j & & & & & \\ \hline & 1 & 2 & 3 & 4 & 5 & 6 \\ & 5 & 3 & 1 & 1 & 3 & 5 \end{array} \right.$$

Table 2: Right-hand side of constraint (†) for each  $j \in N$  when  $n = 6$ .

We illustrate this approach with an example for  $n = 6$ . Table 2 shows constraint (†) for each  $j \in N$ .  $\sum_{i \in N} ix_{ij}$  is the position in the schedule that job  $j$  is assigned to, and hence these constraints enforce that jobs 1 and 6 are scheduled in the final two positions, jobs 2 and 5 are scheduled in the two positions before that, and jobs 3 and 4 are scheduled in the

first two positions. Thus there are two feasible positions for each job. Following the proposed approach for generating a feasible solution, we get that  $x_{6,6} = 1$ ,  $x_{5,1} = 1$ ,  $x_{4,5} = 1$ ,  $x_{3,2} = 1$ ,  $x_{2,4} = 1$ ,  $x_{1,3} = 1$ .

For any given instance, this approach can be used to generate a  $UB$  solution with objective value no worse than twice the  $LB$  solution provided by (LB-SMSP).  $\square$

We next consider the value of  $\frac{UB}{LB}$  as a function of  $n$  and show that the 2-approximation provided by the solution (UB-SMSP) becomes tight as  $n \rightarrow \infty$ .

First, consider the case in which  $n$  is even. We construct a fully-crossed instance with  $p_i = q_i = 0$  for  $i \leq n/2$  and  $p_i = q_i = 1$  for  $i \geq n/2 + 1$  (which we refer to as a fully-crossed 0-1 instance). Such an instance for  $n = 4$  is shown in Figure 5a. The value of  $UB$  is given by  $\sum_{i \in N} (n + i - 1) = \frac{(n+1)n}{2}$ , whilst the value of  $LB$  is given by

$$\sum_{i \in N, i \geq \frac{n}{2}} 2(n + 1 - i) = \frac{2(n/2 + 1)n/2}{2} = \frac{(n/2 + 1)n}{2},$$

and hence when  $n$  is even, we have that

$$\frac{UB}{LB} = \frac{n + 1}{n/2 + 1} = \frac{2n + 2}{n + 2} \leq 2.$$

Now consider the case in which  $n$  is odd. We slightly change the construction of this instance by setting  $p_i = q_i = 0$  for  $i \leq (n - 1)/2$  and  $p_i = q_i = 1$  for  $i \geq (n + 1)/2 + 1$  and  $p_i = 0$ ,  $q_i = 1$  for  $i = (n + 1)/2$ . This instance, for  $n = 5$ , is shown in Figure 5b. Again, the value of  $UB$  is given by  $\sum_{i \in N} (n + i - 1) = \frac{(n+1)n}{2}$ . However, when  $n$  is odd, the value of  $LB$  is given by

$$\frac{n + 1}{2} + \sum_{i \in N, i \geq \frac{n-1}{2}} 2(n + 1 - i) = \frac{n + 1 + 2(\frac{n-1}{2} + 1)(\frac{n-1}{2})}{2} = \frac{(n + 1)^2}{4},$$

and therefore, when  $n$  is odd, we have that

$$\frac{UB}{LB} = \frac{2n(n + 1)}{(n + 1)^2} = \frac{2n}{n + 1} \leq 2.$$

Thus, as  $n \rightarrow \infty$ ,  $\frac{UB}{LB} \rightarrow 2$ .

We continue this analysis by looking at the range in which the true approximation ratio lies, as a function of  $\Delta/n$ . In Figure 3, the line in blue shows the upper bound of 2 as stated by Theorem 4. The line in orange shows a lower bound of the approximation ratio, calculated as the actual approximation ratio for a specific instance, namely, a fully-crossed 0-1 instance with  $n = 100$ . The true approximation ratio is known to lie between these upper and lower bounds, where we already know that for  $\Delta = 0$ , the ratio 2 is tight, and for  $\Delta = n$ , (UB-SMSP) provides an optimal solution, i.e. the approximation guarantee becomes 1.

Observe that an optimal solution to a fully-crossed 0-1 instance will always use the available  $\lfloor \frac{n-\Delta}{2} \rfloor$  swaps on the outer-most jobs. See Figure 4 for an illustration of this. Knowing this, it becomes straightforward to compute the objective function of an optimal solution to such

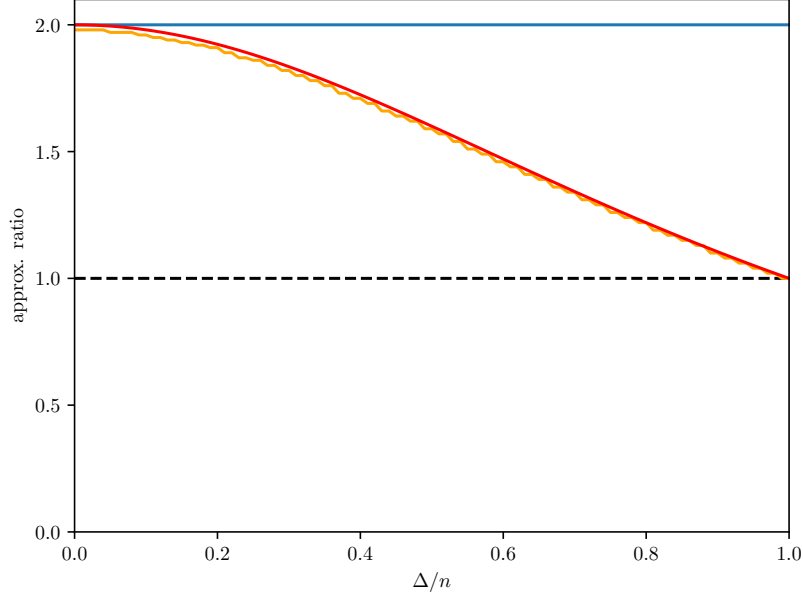


Figure 3: The upper bound on the approximation ratio of (UB-SMSP), as proved in Theorem 4, is shown in blue. A lower bound on the approximation ratio as a function of  $\Delta/n$ , computed using a fully-crossed 0-1 instance with  $n = 100$  is shown in orange. The lower bound as a function of  $\Delta/n$  for fully-crossed 0-1 instances as  $n \rightarrow \infty$  is shown in red.

an instance. For example, in the case where  $n - \Delta$  is even, the optimal objective value,  $v^*$ , is given by

$$\begin{aligned}
 v^* &= \underbrace{\frac{2\left(\frac{n-\Delta}{2} + 1\right)\frac{n-\Delta}{2}}{2}}_{\text{outer } (n-\Delta)/2 \text{ swaps}} + \underbrace{\frac{\left(\left(\frac{n-\Delta}{2} + \Delta\right) + \left(\frac{n-\Delta}{2} + 1\right)\right)\Delta}{2}}_{\text{inner } \Delta \text{ fixed jobs}} \\
 &= \frac{n^2 + \Delta^2 + 2n}{4}.
 \end{aligned}$$

Therefore, when  $n - \Delta$  is even, the true approximation ratio can be computed as

$$\begin{aligned}
 \frac{UB}{v^*} &= \frac{2n(n+1)}{n^2 + \Delta^2 + 2n} \\
 &= \frac{2n^2 + 2n}{(1 + \gamma^2)n^2 + 2n} \rightarrow \frac{2}{1 + \gamma^2}
 \end{aligned}$$

as  $n \rightarrow \infty$ , where  $\gamma = \Delta/n$ . Note that a very similar analysis arrives at the same result for the case where  $n - \Delta$  is odd. In Figure 3, we plot this limiting curve in red. Looking at the plot we see that for  $\Delta = 0$ , when the first and second solutions require no intersection, the upper and lower bounds converge as  $n \rightarrow \infty$ , confirming that the 2-approximation is tight. For  $\Delta > 0$ , it remains possible that still stronger guarantees can be found.

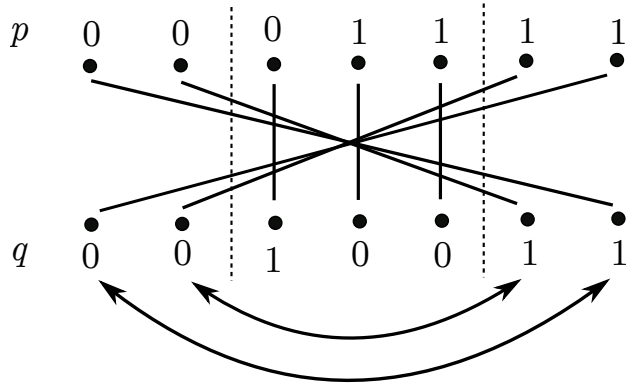


Figure 4: Optimal solution for a fully-crossed 0-1 instance where  $n = 7$  and  $\Delta = 3$ . The  $\lfloor \frac{n-\Delta}{2} \rfloor = 2$  available swaps are applied to the outer-most jobs. This structure is shared by the optimal solutions to all fully-crossed 0-1 instances.

(UB-SMSP) is found by forcing the first and second-stage schedules to be the same. Clearly, this approach is overly conservative, and in practice it is beaten by solutions in which only a subset of jobs share a position across the two stages. The following result shows that the 2-approximation guarantee still holds for such solutions.

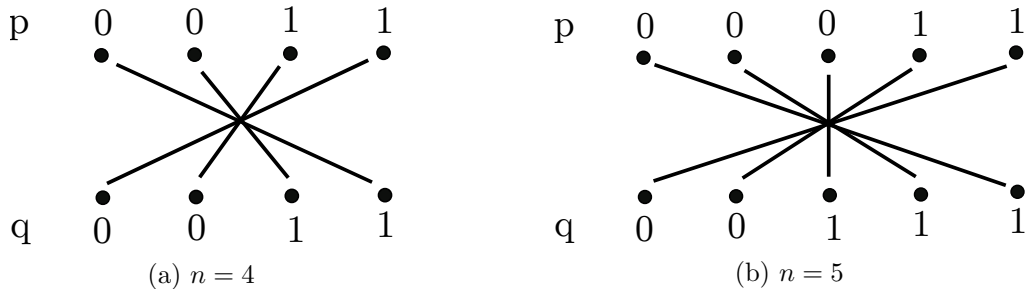


Figure 5: Examples showing the constructed instances for  $n = 4$  and  $n = 5$ .

**Lemma 6.** *Function  $f(M)$  is monotonically non-decreasing, i.e.,  $f(M) \leq f(M \cup \{i\})$  for any  $i \in N \setminus M$ .*

*Proof.* The optimisation problem (RecFix) has constraints for all  $j \in M$ . The problem of solving (RecFix) for  $M$  is therefore a relaxation of the problem of solving (RecFix) for  $M \cup \{i\}$ , and the claim follows.  $\square$

This means that any heuristic that gives a feasible pair of schedules based on Algorithm 1 is a 2-approximation as well.

**Corollary 7.** *For any  $M \subseteq N$  with  $|M| \geq \Delta$ , the solution generated by Algorithm 1 gives a 2-approximation for (RecSMSP) in polynomial time.*

Corollary 7 suggests a natural greedy heuristic for solving (RecSMSP). This is outlined in Algorithm 2. The procedure begins by setting  $M = \emptyset$  and iteratively adds to it the element  $i \in N \setminus M$  with the smallest objective value  $f(M \cup \{i\})$  to the set  $M$ , until  $|X \cap Y| \geq \Delta$ .

Due to Corollary 7, this greedy heuristic is ensured to be a 2-approximation. And due to Lemma 6, we also know that its objective value can be no worse than (UB-SMSP). Each iteration of the greedy heuristic takes  $O(n^2 \log n)$  time and there are  $O(\Delta)$  iterations, hence the heuristic runs in  $O(n^3 \log n)$ .

---

**Algorithm 2** Greedy heuristic for (RecSMSP)

---

```

1: procedure GREEDY( $\mathbf{p}, \mathbf{q}, \Delta$ )
2:   initialise  $M, X, Y = \emptyset$ 
3:   while  $|X \cap Y| < \Delta$  do
4:     Set  $v = \infty$ 
5:     for  $j' \in N \setminus M$  do
6:        $M' \leftarrow M \cup \{j'\}$ 
7:        $v', X, Y \leftarrow \text{EVAL}(\mathbf{p}, \mathbf{q}, M')$  ▷ Solve (RecFix) using EVAL
8:       if  $v' < v$  then
9:          $v \leftarrow v'$ 
10:         $j \leftarrow j'$ 
11:      end if
12:    end for
13:     $M \leftarrow M \cup \{j\}$ 
14:  end while
15:  return  $v, X, Y$ 
16: end procedure

```

---

## 4 Computational experiments

In this section, we compare results from solving the exact MIP formulation of (RecSMSP) with those of the 2-approximation provided by solving (UB-SMSP), as well from the greedy heuristic outlined at the end of the previous section.

Before presenting these results in detail, we outline the test instances used for these experiments and the computational hardware on which these experiments were performed. Section 4.1 investigates the performance of the MIP, focussing on its solution times and the gap to its linear relaxation. Following this, Section 4.2 looks at the 2-approximation and shows that in practice it considerably outperforms its theoretical worst-case performance. Finally, Section 4.3 presents results from the greedy heuristic.

For each value of  $n \in \{10, 20, 50, 100\}$ , 100 instances have been generated by randomly sampling  $p_i$  and  $q_i$ ,  $i \in N$ , from the set  $\{1, 2, \dots, 100\}$ . Experiments have been performed on these four instance sets for values of  $\Delta \in \{0, 1, 2, \dots, n\}$ . These problem instances, in addition to the complete results data, can be downloaded from <https://github.com/boldm1/recoverable-robust-interval-SMSP>.

All the experiments have been run on 4 cores of a 2.30GHz Intel Xeon CPU, limited to 16GB RAM. The exact model has been solved using Gurobi 9.0.1, with a time limit of 20 minutes. The optimality gap tolerance (MIPGap) was changed to  $1 \times 10^{-6}$ ; all other parameters were set to their default values.

## 4.1 MIP

We begin by considering the performance of the MIP formulation for (RecSMSP). Unsurprisingly, its performance depends on the size of  $n$ . However, more interestingly, it also depends on the number of free assignments across the first and second-stages, given by  $n - \Delta$ , and in particular whether this value is even or odd. We first present results that show the limitations of the MIP and demonstrate this dependence on the value of  $n - \Delta$ , before we then investigate the causes of this dependence by examining illustrative examples.

Note that since  $n$  is even for each of the instance sets we consider, even and odd values of  $\Delta$  correspond to even and odd values of  $n - \Delta$ , respectively. Therefore, for ease of presentation, throughout this section we refer to the problem in terms of even and odd  $\Delta$ , rather than in terms of even and odd  $n - \Delta$ .

For  $n = 10$  and  $n = 20$ , the MIP was solved to optimality within the 20 minute time limit for all instances and for all values of  $\Delta$ . For  $n = 50$ , however, 5 instances for  $\Delta = 47$  and 47 instances for  $\Delta = 49$  were not solved to optimality within the time limit. And for  $n = 100$ , the number of instances that were not solved to optimality within the time limit for the values of  $\Delta$  are shown in the Table 3. Note how it is only for certain odd values of  $\Delta$  (and therefore odd values of  $n - \Delta$ ) that the MIP cannot solve all instances.

$\Delta$	85	87	89	91	93	95	97	99
# non-opt.	2	3	9	17	43	57	83	100

Table 3: Number of instances with  $n = 100$  for which the MIP could not be solved to optimality within 20 minutes, for different values of  $\Delta$ . All instances were solved to optimality for the values of  $\Delta$  not present in the table.

Figure 6 shows the average time to solve the MIP for even and odd values of  $\Delta$  as a percentage of  $n$ , for  $n \in \{10, 20, 50, 100\}$ . A log-scale has been used for clarity. These plots show that solution times scale with  $n$  as expected. However, for odd values of  $\Delta$  we see that solution times increase considerably for large  $\Delta$ , whereas when  $\Delta$  is even, it has little impact on solution times.

Clearly, the MIP cannot be easily solved for  $n \geq 50$  for certain large values of  $\Delta$ , and therefore the performance of the approximate solution approaches we propose are of genuine interest. Before we examine their performances over the following two sections, we investigate the marked difference in difficulty between instances with even and odd  $\Delta$ .

We first consider the tightness of the MIP formulation by looking at the average gap between its solutions and the solutions of its linear relaxation (i.e. the average LP gap) in Figure 7. Observe that the average LP gaps for instances where  $\Delta$  is even are orders of magnitudes smaller than for instances where  $\Delta$  is odd; the largest single LP gap for an even- $\Delta$  instance and an odd- $\Delta$  instance for each of the instance sets are given in Table 4.



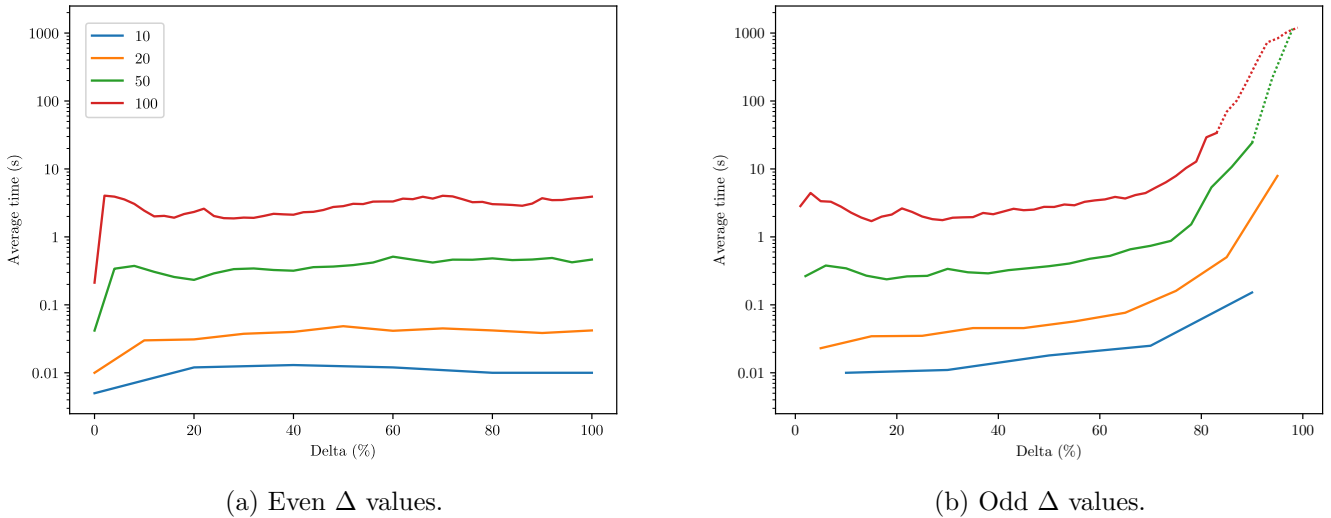


Figure 6: Average time to solve the MIP for even and odd values of  $\Delta$  as a percentage of  $n$ , for  $n \in \{10, 20, 50, 100\}$ . Line becomes dotted when not all instances were solved to optimality within the 1200s (20 minute) time limit.

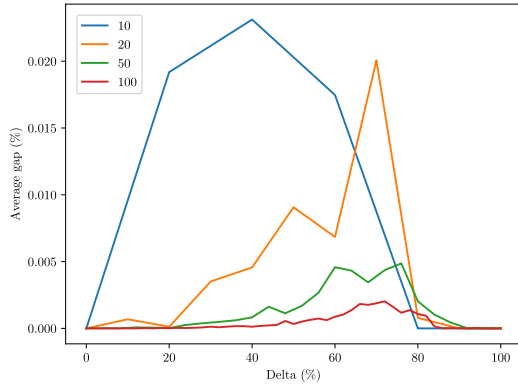
$n$	LP gap (%)			
	even- $\Delta$		odd- $\Delta$	
10	0.20	( $\Delta = 4$ )	5.66	( $\Delta = 9$ )
20	0.17	( $\Delta = 14$ )	2.89	( $\Delta = 19$ )
50	0.07	( $\Delta = 38$ )	1.77	( $\Delta = 49$ )
100	0.03	( $\Delta = 82$ )	0.89	( $\Delta = 99$ )

Table 4: Largest LP gap for a single even- $\Delta$  instance and odd- $\Delta$  instance for each of the instance sets.

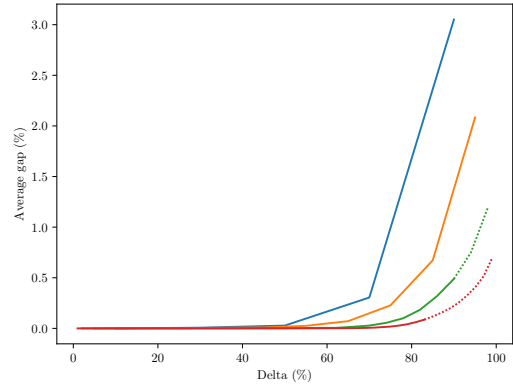
Figure 8 shows the percentage of instances for which the LP gap is non-zero, for even and odd values of  $\Delta$  as a percentage of  $n$ , for  $n \in \{10, 20, 50, 100\}$ . Looking at this figure, we see that the proportion of instances with non-zero LP gaps is broadly similar for each of the four instance sets over the values of  $\Delta$ . Therefore, to explain the results seen in Figure 7, it must be the case that the LP gaps become smaller as  $n$  increases. Figure 8 also shows that for larger values of  $\Delta$ , the frequency of non-zero LP gaps increases dramatically for instances with odd values of  $\Delta$ , but not for instances with even values of  $\Delta$ .

In an attempt to shed light on this distinction, we consider the example with  $n = 4$  jobs with binary processing times given in Figure 5a. Given  $\Delta$ , we are free to allow up to  $n - \Delta$  assignments to differ between the first and second-stage schedules. For clarity of presentation, in the following we assume that the first-stage schedule is a fixed horizontal assignment, and make changes to the assignments in the second-stage only.

Observe that for a given  $\Delta$ , the best integral solution is found by swapping the positions of the jobs scheduled first and last, then swapping the positions of the jobs scheduled second and second-last, and so on, until no more swaps can be made without exceeding the  $n - \Delta$  available free assignments. When  $n - \Delta$  is even, the integral solution is able to change the

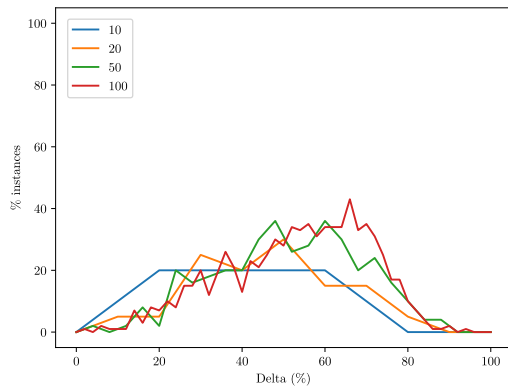


(a) Even  $\Delta$  values.

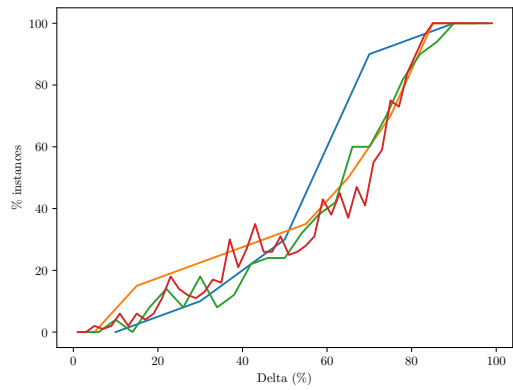


(b) Odd  $\Delta$  values.

Figure 7: Average LP gap for even and odd values of  $\Delta$  as a percentage of  $n$ , for  $n \in \{10, 20, 50, 100\}$ . Line becomes dotted when not all instances were solved to optimality within the 20 minute time limit.



(a) Even  $\Delta$  values.



(b) Odd  $\Delta$  values.

Figure 8: Percentage of instances for which the LP gap is non-zero, for even and odd values of  $\Delta$  as a percentage of  $n$ , for  $n \in \{10, 20, 50, 100\}$ .

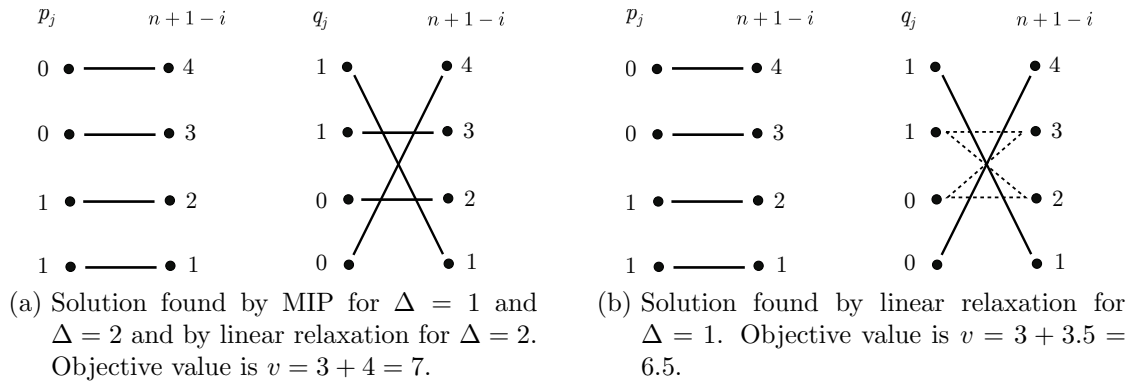


Figure 9: Solutions to instance shown in Figure 5a, found by MIP and its linear relaxation for  $\Delta = 1$  and  $\Delta = 2$ .

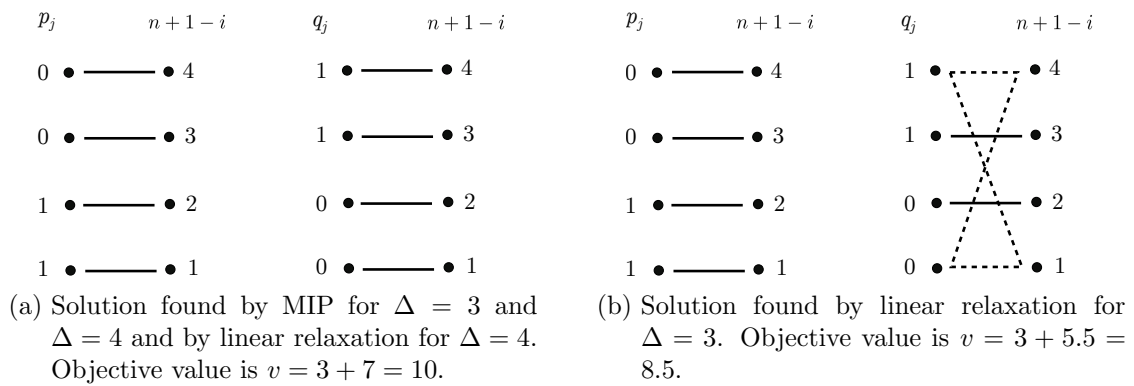


Figure 10: Solutions to instance shown in Figure 5a, found by MIP and its linear relaxation for  $\Delta = 3$  and  $\Delta = 4$ .

assignments of exactly  $n - \Delta$  jobs, and its linear relaxation finds the same solution and gains no advantage. When  $n - \Delta$  is odd however, the integral solution is only able to change the assignments of  $n - \Delta - 1$  jobs. In this case, the linear relaxation is able to make use of fractional assignments to beat the integral solution.

For example, when  $\Delta = 1$ , the MIP swaps the positions of only two jobs (Figure 9a). The linear relaxation gains an advantage over the integral solution since it can fractionally swap a further two jobs to make use of the last remaining free assignment (Figure 9b). In this instance the LP gap is given by  $\frac{7-6.5}{6.5} \approx 7.7\%$ .

Similarly, when  $\Delta = 3$ , the integral solution cannot feasibly swap the positions of any jobs (Figure 10a), whereas its linear relaxation is able to fractionally swap the positions of the first and last jobs (Figure 10b). The LP gap in this case is given by  $\frac{10-8.5}{8.5} \approx 17.6\%$ .

Experimental results seem to suggest that the maximum LP gap that can be achieved is 20%, which occurs for a fully-crossed instance with binary processing times when  $n = 2$  and  $\Delta = 1$  and when  $n = 3$  and  $\Delta = 2$ .

## 4.2 UB-SMSP

As the results in the previous section demonstrate, large instances of (RecSMSP) cannot be solved within a reasonable time limit by the MIP. Therefore, accurate heuristic approximations are valuable for solving this problem. In this section, we look at the quality of the solutions to (UB-SMSP) in practice, to compare them against their theoretical worst-case performance ratio of 2. Recall that (UB-SMSP) can be solved trivially by ordering the jobs  $j \in N$  by non-decreasing  $p_j + q_j$  values. In our experiments, (UB-SMSP) was solved in less than 0.01 seconds for every instance.

Figure 11 shows the average relative gap between the solution to (UB-SMSP) and the solution to (RecSMSP) found by solving the MIP, plotted for values of  $\Delta$  as a percentage of  $n$ , for  $n \in \{10, 20, 50, 100\}$ . A dotted line has been used to indicate the range of  $\Delta$  values for which the MIP was not solved to optimality for all instances within the 20 minute time limit for the odd values of  $\Delta$  within this range. The average gaps displayed for the odd values of  $\Delta$  in this range are therefore estimates of the true average gap.

For all problem sizes the average gap decreases as  $\Delta$  increases, reaching 0 when  $\Delta = n$ , where (RecSMSP) and (UB-SMSP) become equivalent. For  $n = 10$ , the largest gap between for any single instance was 21.7%; for  $n = 20$  the largest single gap was 18.2%, for  $n = 50$  it was 23.0%, and for  $n = 100$  it was 21.1%. For each value of  $n$ , this largest single gap was attained when  $\Delta = 0$ . Clearly, the worst-case gaps we see in practice are considerably smaller than the theoretical worst-case gap of 100%.

## 4.3 Greedy heuristic

Finally, we examine the performance of greedy heuristic presented in Algorithm 2 for solving (RecSMSP). Recall that, as a result of Corollary 7, the objective value of solutions found by the greedy heuristic can be no worse than the solutions to (UB-SMSP), and as these results show, in practice they are considerably better than this.

We first look at the average run time of the greedy heuristic, shown in Figure 12 for values of  $\Delta$  as a percentage of  $n$ , for each of the four instance sets. As expected, we see that the average run time grows with  $n$  and with  $\Delta$ . The longest run time of any single instance was 4.72 seconds, which occurred for an instance with  $n = 100$ ,  $\Delta = 95$ .

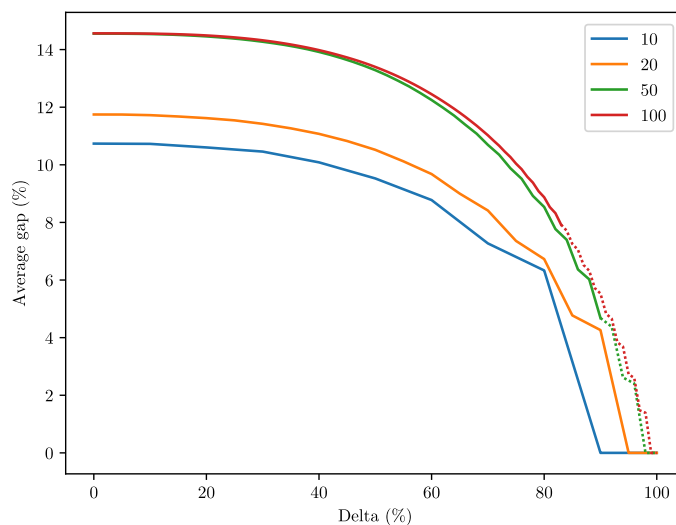


Figure 11: Average gap between solution to (UB-SMSP) and the solution to (RecSMSP) for values of  $\Delta$  as a percentage of  $n$ , for  $n \in \{10, 20, 50, 100\}$ . Line becomes dotted for the range of  $\Delta$  values for which not all instances were solved to optimality by the MIP within its 20 minute time limit for the odd values of  $\Delta$  within this range.

Figure 13 shows the average relative gap between the solution found by the greedy heuristic and the solution to the MIP for values of  $\Delta$  as a percentage of  $n$ , for  $n \in \{10, 20, 50, 100\}$ . A dotted line connects the data points at each value of  $\Delta$ , whilst a solid line connects the data points corresponding to odd values of  $\Delta$  only. The figure shows that the accuracy of the greedy heuristic varies depending on whether  $\Delta$  is odd or even, demonstrated by the characteristic zig-zagging of the dotted line.

Even as the average gap between the greedy heuristic and the MIP increases with  $\Delta$ , this gap is bounded above by the average gap between the solutions to (UB-SMSP) and the MIP, which sharply decreases with  $\Delta$  (see Figure 11). The greedy heuristic finds solutions that are an order of magnitude closer to optimal than the solutions to (UB-SMSP). For  $n = 10$  the maximum gap between the solution found by the greedy heuristic and the MIP solution for a single instance is 0.55%, for  $n = 20$  it is 0.59%, for  $n = 50$  it is 0.33%, and for  $n = 100$  it is 0.18%. Hence, solutions found by the greedy heuristic are nearly optimal.

## 5 Conclusions

This paper has presented a theoretical and computational investigation of the recoverable robust single machine scheduling problem under interval uncertainty with the objective of minimising the sum of completion times, (RecSMSP). Firstly, a number of positive complexity results have been produced for key subproblems. Specifically, we present a polynomial time sorting algorithm in Algorithm 1 which can optimally solve the case where a subset of jobs that must share a first and second-stage assignment is given. Using this result, we show that the special cases of (RecSMSP) with a constant value of  $\Delta$  and with a constant

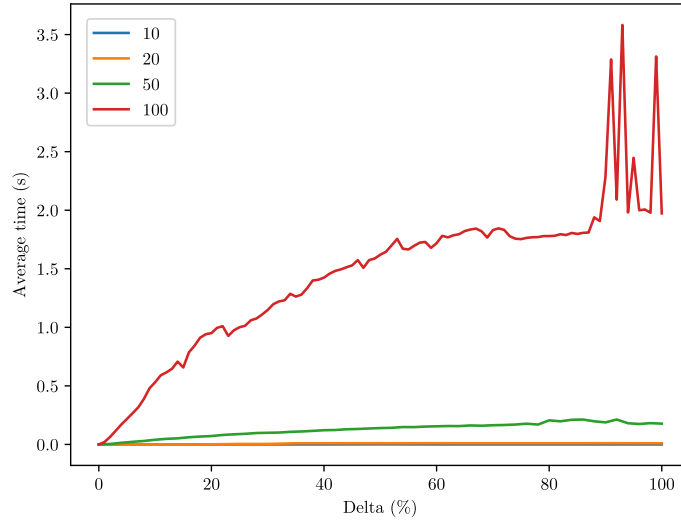


Figure 12: Average greedy heuristic run time for values of  $\Delta$  as a percentage of  $n$ , for  $n \in \{10, 20, 50, 100\}$ .

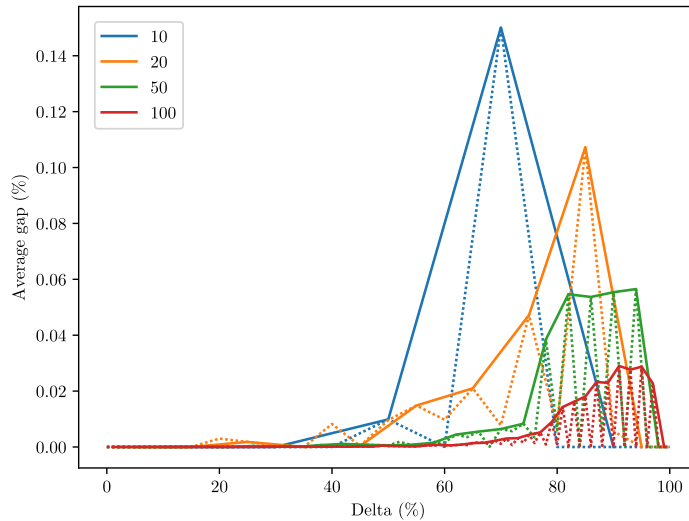


Figure 13: Average gap between greedy and MIP solution for values of  $\Delta$  as a percentage of  $n$ , for  $n \in \{10, 20, 50, 100\}$ , plotted as a dotted line. Odd- $\Delta$  data points are connected with a solid line.

number of possible processing times are solvable in polynomial time. We then introduce the problem (UB-SMSP) in which the first and second-stage assignments are forced to be identical, and prove that the solution to this problem provides a 2-approximation of (RecSMSP). This guarantee extends to the solutions found by a simple and fast greedy heuristic. A set of extensive computational experiments investigate the limitations of an exact mixed-integer programming approach to solving the problem, show that the solutions of (UB-SMSP) significantly outperform their worst-case guarantee in practice, and demonstrate the strength of the proposed greedy heuristic.

Note that the computational complexity of (RecSMSP) has yet to be characterised and remains unknown. In addition to investigating this, a promising direction for future research on this problem would be to consider the impact of alternative uncertainty sets on its complexity. Finally, another possibility for future research would be to look at this recoverable robust model for single machine scheduling problems with objective criteria different to the sum of completion times that we consider here.

## Acknowledgements

The authors are grateful for the support of the EPSRC-funded (EP/L015692/1) STOR-i Centre for Doctoral Training.

## References

- Aloulou, M. A. and Della Croce, F. (2008). Complexity of single machine scheduling problems under scenario-based uncertainty. *Operations Research Letters*, 36(3):338–342.
- Bold, M. and Goerigk, M. (2020). Recoverable robust single machine scheduling with budgeted uncertainty. *arXiv preprint arXiv:2011.06284*.
- Bougeret, M., Pessoa, A. A., and Poss, M. (2019). Robust scheduling with budgeted uncertainty. *Discrete Applied Mathematics*, 261:93–107.
- Büsing, C. (2012). Recoverable robust shortest path problems. *Networks*, 59(1):181–189.
- Daniels, R. L. and Kouvelis, P. (1995). Robust scheduling to hedge against processing time uncertainty in single-stage production. *Management Science*, 41(2):363–376.
- Deĭneko, V. G., Woeginger, G. J., et al. (2006). On the robust assignment problem under a fixed number of cost scenarios. *Operations Research Letters*, 34(2):175–179.
- Fischer, D., Hartmann, T. A., Lendl, S., and Woeginger, G. J. (2020). An investigation of the recoverable robust assignment problem. *arXiv preprint arXiv:2010.11456*.
- Goerigk, M., Lendl, S., and Wulf, L. (2021). On the recoverable traveling salesman problem. *arXiv preprint arXiv:2111.09691*.
- Graham, R. L., Lawler, E. L., Lenstra, J. K., and Kan, A. R. (1979). Optimization and approximation in deterministic sequencing and scheduling: a survey. In *Annals of Discrete Mathematics*, volume 5, pages 287–326. Elsevier.

- Hradovich, M., Kasperski, A., and Zieliński, P. (2017a). Recoverable robust spanning tree problem under interval uncertainty representations. *Journal of Combinatorial Optimization*, 34(2):554–573.
- Hradovich, M., Kasperski, A., and Zieliński, P. (2017b). The recoverable robust spanning tree problem with interval costs is polynomially solvable. *Optimization Letters*, 11(1):17–30.
- Kasperski, A. and Zieliński, P. (2008). A 2-approximation algorithm for interval data minmax regret sequencing problems with the total flow time criterion. *Operations Research Letters*, 36(3):343–344.
- Kasperski, A. and Zieliński, P. (2014). Minmax (regret) scheduling problems. *Sequencing and scheduling with inaccurate data*, pages 159–210.
- Kasperski, A. and Zieliński, P. (2016a). Robust discrete optimization under discrete and interval uncertainty: A survey. In *Robustness analysis in decision aiding, optimization, and analytics*, pages 113–143. Springer.
- Kasperski, A. and Zieliński, P. (2016b). Single machine scheduling problems with uncertain parameters and the OWA criterion. *Journal of Scheduling*, 19(2):177–190.
- Kasperski, A. and Zieliński, P. (2017). Robust recoverable and two-stage selection problems. *Discrete Applied Mathematics*, 233:52–64.
- Kasperski, A. and Zieliński, P. (2019). Risk-averse single machine scheduling: complexity and approximation. *Journal of Scheduling*, 22(5):567–580.
- Kouvelis, P. and Yu, G. (1997). *Robust discrete optimization and its applications*. Kluwer Academic Publishers Dordrecht, Netherlands.
- Lebedev, V. and Averbakh, I. (2006). Complexity of minimizing the total flow time with interval data and minmax regret criterion. *Discrete Applied Mathematics*, 154(15):2167–2177.
- Liebchen, C., Lübbecke, M., Möhring, R., and Stiller, S. (2009). The concept of recoverable robustness, linear programming recovery, and railway applications. In *Robust and online large-scale optimization*, pages 1–27. Springer.
- Lu, C.-C., Ying, K.-C., and Lin, S.-W. (2014). Robust single machine scheduling for minimizing total flow time in the presence of uncertain processing times. *Computers & Industrial Engineering*, 74:102–110.
- Mastrolilli, M., Mutsanas, N., and Svensson, O. (2013). Single machine scheduling with scenarios. *Theoretical Computer Science*, 477:57–66.
- Montemanni, R. (2007). A mixed integer programming formulation for the total flow time single machine robust scheduling problem with interval data. *Journal of Mathematical Modelling and Algorithms*, 6(2):287–296.
- Pereira, J. and Averbakh, I. (2011). Exact and heuristic algorithms for the interval data robust assignment problem. *Computers & Operations Research*, 38(8):1153–1163.



Yang, J. and Yu, G. (2002). On the robust single machine scheduling problem. *Journal of Combinatorial Optimization*, 6(1):17–33.

Zhao, H., Zhao, M., et al. (2010). A family of inequalities valid for the robust single machine scheduling polyhedron. *Computers & Operations Research*, 37(9):1610–1614.