Zuse Institute Berlin

Takustr. 7
14195 Berlin
Germany

GUILLAUME SAGNOL, CHRISTOPH BARNER, RALF BORNDÖRFER,
MICKAËL GRIMA, MATHEES SEELING, CLAUDIA SPIES,
KLAUS WERNECKE

Robust Allocation of Operating Rooms: a Cutting Plane Approach to handle Lognormal Case Durations

A first version of this report appeared in March 2016.
The work of the first author is carried out in the framework of MATHEON supported by Einstein Foundation Berlin.

ZIB Report 16-18 (July 2018)

Zuse Institute Berlin
Takustr. 7
14195 Berlin
Germany

Telephone: +49 30-84185-0
Telefax: +49 30-84185-125

E-mail: bibliothek@zib.de
URL: http://www.zib.de

ZIB-Report (Print) ISSN 1438-0064
ZIB-Report (Internet) ISSN 2192-7782

Robust Allocation of Operating Rooms: a Cutting Plane Approach to handle Lognormal Case Durations and Emergency Arrivals

Guillaume Sagnol[12], Christoph Barner[2], Ralf Borndörfer[1], Mickaël Grima[3], Mathees Seeling[2], Claudia Spies[2], Klaus Wernecke[2]

Keywords: Parallel machines scheduling; Extensible bin packing; OR in medicine; Robust optimization; Lognormal durations; Hilbert's projective metric

Abstract

The problem of allocating operating rooms (OR) to surgical cases is a challenging task, involving both combinatorial aspects and uncertainty handling. We formulate this problem as a parallel machines scheduling problem, in which job durations follow a lognormal distribution, and a fixed assignment of jobs to machines must be computed. We propose a cutting-plane approach to solve the robust counterpart of this optimization problem. To this end, we develop an algorithm based on fixed-point iterations that identifies worst-case scenarios and generates cut inequalities. The main result of this article uses Hilbert's projective geometry to prove the convergence of this procedure under mild conditions. We also propose two exact solution methods for a similar problem, but with a polyhedral uncertainty set, for which only approximation approaches were known. Our model can be extended to balance the load over several planning periods in a rolling horizon. We present extensive numerical experiments for instances based on real data from a major hospital in Berlin. In particular, we find that: (i) our approach performs well compared to a previous model that ignored the distribution of case durations; (ii) compared to an alternative stochastic programming approach, robust optimization yields solutions that are more robust against uncertainty, at a small price in terms of average cost; (iii) the *longest expected processing time first* (LEPT) heuristic performs well and efficiently protects against extreme scenarios, but only if a good prediction model for the durations is available. Finally, we draw a number of managerial implications from these observations.

1 Introduction

The operating theater (OT) is one of the most expensive hospital resources. Recent studies indicate that in certain hospitals, surgical interventions concentrate up to 70% of all patient admissions, and as much as 40% of the total expenses [18]. The management of the OT

[1]Zuse Institute Berlin
[2]Charité Universitätsmedizin, Berlin
[3]Technische Univerisität München

is a very complex task, which involves several hierarchical decision levels and combinatorial aspects for many different types of resources (operating rooms, surgeons, nurses, anesthesiologists, etc.), all this in an uncertain environment (surgical durations, emergency cases, availability of recovery beds). For this reason, there has been a considerable effort to develop optimization procedures to improve the management of resources in the operating theater; we refer the reader to [27] for a comprehensive review of the operations research literature on OT management.

This paper focuses on the problem of allocating operating rooms to elective patients, typically on the day prior to operation. More precisely, the goal is to assign operating rooms (OR) to a list of *patient blocks*, that is, groups of elective patients to be operated one after another by the same surgical team. This is a crucial planning step for the so-called *block-scheduling* system (cf. [27]), in which individual surgeons or surgery specialties have predefined slots of OR-time allocated in a periodic schedule (the *Master surgery schedule*, MSS), and cases must be booked within these slots.

The problem of managing the operating rooms is characterized by a very strong stochasticity; see e.g. [49]. In particular, it is well known that durations exhibit a close fit with the lognormal distribution, see e.g. [48, 31] and the references therein. This uncertainty frequently leads to operations that exceed the planned OR-time in one slot of the MSS. The excess of OR-time is known as *overtime*, and induces high costs for the hospital. The need to take into account the uncertainty, and to exploit distributional information on durations is thus extremely important, all the more so as lognormal variables are heavy-tailed. This suggests the use of robust optimization techniques, which aim at protecting against extreme scenarios. The problem we study is a variation of a robust optimization problem introduced in [19]. The only difference is in the model: we specifically take into account that durations are lognormal, which allows us to define a natural uncertainty set in terms of likely scenarios.

The efficiency of the OT can be measured by a combination of the number of under-utilized hours and the number of over-utilized hours in the operating rooms [20]. However, a few days before the day of surgery, the staff has already been scheduled, and so [20] claims that under-utilized time does not cause a loss of revenue for the surgical suite. This is consistent with [36], where it is shown that on a short-term perspective, the goal is solely to minimize the overtime in the OR. Also, we consider a fixed cost for *opening an OR*, as proposed in the model of [19]. However, under-utilization of the OR can still have indirect costs on a rather short-term perspective (even with a fixed staffing). This happens, e.g., when the load of surgeries is not balanced uniformly among several operating days. In this case, it might be well-suited to postpone some patients to a later day. Our model can easily be adapted to the situation in which the decision maker may cancel some jobs, or on the contrary when he may accept more jobs than initially planned, by using a rolling horizon with deferral costs. This situation naturally occurs in settings where the OR allocation problem is to be solved for a sequence of several planning periods; in this case, the ability to postpone (or bring forward) a job to a later (or earlier) planning period can help to balance the overtime over the whole planning horizon.

Since the end of the 90's, many papers have demonstrated the benefits of robust opti-

mization (RO) to handle uncertainty. In many cases, the method offers tractable mathematical formulations which are much easier to solve than their stochastic programming counterparts [7, 9]. Moreover, RO offers the possibility to tune the *budget of uncertainty* to choose the tradeoff between performance and robustness.

Another traditional selling point for these approaches is that no distributional information for the uncertain parameters is required. In the context of surgery scheduling however, we already mentioned that lognormality of the durations can be assumed, and we want to take advantage of this. We point out that most statistical studies on prediction models for the distribution of surgery durations, such as [31, 47], focus on the procedure time only. In our approach however, the relevant duration is the total duration of patient blocks, which consists of the sum of the procedure times, set-up times, and clean-up times of all patients in this block. In practice, most patient blocks contain between one and three surgical cases, and we think that the lognormal model is still a good model for the whole block duration. Indeed, it has been proposed to approximate the sum of lognormal distributions by the lognormal variable that matches its first two moments, which is the well-known Fenton-Willkinson approximation [23]. It is used routinely in financial engineering and other fields, such as signal processing, and provides a reasonable approximation for a range of lognormal parameters [14]. Another possibility would be to match three moments of the 3-parameters lognormal (which has an additional shift parameter), which has also been proposed in [48] to model surgery durations; it would be straightforward to extend our robustness model to the case of shifted lognormal distributions.

The fat-tail behaviour of the lognormal makes it likely that standard uncertainty models, such as the ellipsoidal uncertainty model of Ben-Tal and Nemirovsky [7] or the cardinality-constrained uncertainty model of Bertsimas and Sim [9], will offer a rather poor model of the real-life setting. To remedy this problem, we propose to use robust optimization with an uncertainty model which protects against all scenarios in a confidence region of the lognormal distribution. This turns robust optimization into a risk assessment technique (cf. Proposition 3.5), as other approaches relying on the conditional value-at-risk (CVaR), cf. [45], or the ordered weighted average (OWA); see [30].

To the best of our knowledge, one of the first papers to consider the problem of allocating operating rooms to a list of surgical procedures is [41], who proposed a mixed integer programming (MIP) formulation to minimize the under- and overutilization of the ORs. This paper makes the assumption that all the procedures of a given practitioner are performed in the same OR. This is also the approach that we adopt here (patients to be operated by the same surgical team are grouped in a block), for two main reasons:

1. When a surgeon performs two procedures in two different ORs, there is a risk that the first procedure takes longer than expected, which induces waiting time in both ORs and generates overtime. In contrast, it is known that planning all the procedures of one surgeon in a single OR is a guarantee of stability, cf. [21], a feature desired by many OR planners, in particular at the Charité hospital in Berlin.

2. Mathematical formulations allowing a practitioner to change the OR within a day are much harder to solve, because we need to take synchronization issues into account. The

resulting problem is a stochastic RCPSP (Resource constrained project scheduling problem). While several MIP formulations are available for the deterministic RCPSP [33], in the stochastic setting precedence models must be used. These models suffer from relying on *big-M* constraints to avoid that two procedures performed by the same surgeon take place at the same time, which leads to weak relaxations and very long computing times.

We are aware that in some cases, in particular when the number of surgeons is a bottleneck for the planning, it might be better to let practitioners alternate between two rooms, so they can perform a surgery in room B while room A is being cleaned-up and prepared for the next patient. This is the approach used for example in [43, 44], where MIPs are proposed to solve a deterministic resource constrained scheduling problem. There are also stochastic programming approaches for the problem with room-changing surgeons: [35] scheduled one surgeon operating in two ORs, and [6] used the L-shaped method to solve a stochastic MIP model (with big-M's), for the case where surgeons may change room but have a predefined sequence of patients to operate in a given order.

The popularity of robust optimization techniques can also be observed in the literature on OT management. A non-exhaustive list of recent contributions using robust optimization follows: in [28], RO is used to allocate slack times in each OR to reduce the risk of overtime; an RO model is proposed in [1] to allocate patients to OR-blocks (in a block-scheduling system), by considering their individual due-dates; a closely related paper is [46], where a similar problem is handled by means of chance-constrained optimization, by assuming normal distribution of the surgical durations; a distributionally robust model is proposed in [37], to select elective admission ratios in order to balance bed occupancy.

We build on a robust optimization problem introduced by [19]. This paper presented an MIP model (called `MRORA`) to find an optimal allocation of the ORs, robust against all duration scenarios $\boldsymbol{d}$ for the patient blocks in the uncertainty set

$$\mathcal{D}_{\texttt{MRORA}} = \{\boldsymbol{d} \in \mathbb{R}^n : \ \forall i, \ell_i \leq d_i \leq u_i; \ \sum_i \frac{d_i - \ell_i}{u_i - \ell_i} \leq \tau\}, \tag{1}$$

where ℓ_i and u_i are lower and upper bounds for the duration of the ith patient block, and the parameter τ controls the *budget of uncertainty*. It was shown very recently [3] that the MIP model of [19] is inexact: this MIP only minimizes an upper bound of the worst case cost over $\mathcal{D}_{\texttt{MRORA}}$. In order to compare our approach (in which the uncertainty set consists of likely scenarios of the lognormal distribution) to the case of the popular budgeted uncertainty set $\mathcal{D}_{\texttt{MRORA}}$, we need to solve the robust allocation problem over $\mathcal{D}_{\texttt{MRORA}}$ *exactly*. Another original contribution of our paper is that we present two methods to solve the robust allocation problem over $\mathcal{D}_{\texttt{MRORA}}$.

Mathematically, the problem introduced in [19] is a parallel machine scheduling problem. In recent years, robust machine scheduling problems attracted a lot of attention. For example, [30] proposed approximation and pseudo-polynomial time algorithms to optimize a risk measure (ordered weighted averaging aggregation, OWA) of the makespan, with a run-time proportional to 2^k, where k is the number of considered scenarios in the uncertainty set.

Many papers focus on the problem of minimizing the sum of completion times: [11] proves the NP-hardness of the robust counterpart of this problem, when the uncertainty set is a budget polytope *à la* Bertsimas and Sim [9]. A scenario-based robust counterpart for the problem of minimizing the sum of completion times is studied in [51]: An MIP model and a 2-approximation algorithm are proposed. A variant of this problem with sequence-dependent setup times is also studied in [29], with the help of metaheuristics. The problem of maximizing the probability that the sum of completion times does not exceed a prescribed threshold was handled with a branch-and-bound procedure in [2], under normality assumptions on job durations.

Contrarily to the aforementioned approaches, the relevant criterion for the allocation of operating rooms is the total tardiness of the machines, which has not been investigated much from a robust optimization perspective. This criterion is however equivalent to the criterion of *extensible bin packing*, for which performance bounds of simple heuristics have been derived in the deterministic setting [17] and in the online setting [5].

This paper is organized as follows: In Section 2 we present a generalization of the robust OR allocation problem introduced in [19] for an arbitrary uncertainty set $\mathcal{D}$, and a cutting-plane approach to solve it. This solution procedure relies on the ability to solve the *separation problem* for the set $\mathcal{D}$, that is, the problem that generates new cuts by identifying the worst case scenario for a given allocation. Section 3 is concerned with the solution of this separation problem. We show how to solve the separation problem for the set $\mathcal{D}_{\text{MRORA}}$ in Section 3.1, and for a confidence region $\mathcal{D}_r$ of the lognormal distribution in Section 3.2. In particular, our main result is stated in Proposition 3.3; we show with the help of *Hilbert's projective metric* that fixed-point iterations can be used to solve very efficiently the separation problem, although this is a non-convex optimization problem. Then, we propose two alternative solution approaches in Section 4: the first one is a reformulation approach to solve exactly the problem introduced in [19] for robust optimization over a polyhedral uncertainty set (such as $\mathcal{D}_{\text{MRORA}}$). The resulting MIP is not compact in general, but it is if we restrict our attention to instances with a bounded number of operating rooms. The second approach is a sample average approach (SAA) to approximate the stochastic programming counterpart of the OR allocation problem. An important extension of our model is presented in Section 5: it allows one to use a rolling horizon approach, in which we also select the patients that will be operated for the next operating day. We present numerical results for the application to OR management in Section 6, in which we present an extensive comparison of different optimization and heuristic approaches. These experiments suggest a number of managerial implications which we present in Section 7. In particular, the *longest expected processing time first* (LEPT) heuristic performs well and efficiently protects against extreme scenarios, but only if a good prediction model for the durations is used. Hospitals should hence invest in the development of accurate data-driven predictors for the case durations, rather than relying on surgeons' estimates only for the allocation of patient blocks to ORs, a claim also supported, e.g., by [24].

Throughout this article, we adopt the terminology of the parallel machines scheduling literature, because we believe that the problem studied here could have other fields of appli-

cation. Hence, *patient blocks* are called *jobs* and *operating rooms* are called *machines*.

2 Problem Formulation

Throughout this article, plain italics denote scalars and lowercase boldface symbols denote vectors. In particular, the vector with elements v_i is denoted by $\boldsymbol{v}$. The symbol $\mathrm{Diag}(\boldsymbol{v})$ represents the diagonal matrix with elements v_i on the diagonal, and $\mathbf{0}_n$ is the zero vector of dimension n. We use the notation $\mathbb{R}_+$ for the set of nonnegative real numbers, and $\|\cdot\|_p$ denotes the usual ℓ_p-norm. The expected value of a random variable X is denoted by $\mathbb{E}[X]$, and $\mathbb{P}[E]$ stands for the probability of the event E.

We denote by $\mathcal{J}$ and $\mathcal{M}$ the sets of *jobs* and *machines*, of respective cardinality n and p. The binary variable z_m indicates whether machine $m \in \mathcal{M}$ is activated, and the binary variable x_{jm} tells whether job j is allocated to machine m. Each job must be allocated to one activated machine, so the set of all feasible solutions reads

$$\mathcal{X} := \left\{ (\boldsymbol{x}, \boldsymbol{z}) \in \{0,1\}^{n \times p} \times \{0,1\}^p : \begin{array}{rl} \forall j \in \mathcal{J}, & \sum_{m \in \mathcal{M}} x_{jm} = 1; \\ \forall j,m \in \mathcal{J} \times \mathcal{M}, & x_{jm} \leq z_m \end{array} \right\}.$$

Denote by T_m the time available on machine m (if it is activated), c_m^f the fixed cost for activating machine m and c_m^o the cost of overtime per unit of time on machine m. If the duration of job j is $d_j > 0$, the total cost of an allocation $(\boldsymbol{x}, \boldsymbol{z}) \in \mathcal{X}$ can be measured as

$$F(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{d}) := \sum_{m \in \mathcal{M}} c_m^f z_m + c_m^o \left(\sum_{j \in \mathcal{J}} x_{jm} d_j - T_m \right)^+,$$

where $(u)^+ := \max(u, 0)$ denotes the nonnegative part of $u \in \mathbb{R}$.

We consider the problem of finding the allocation $(\boldsymbol{x}, \boldsymbol{z})$ minimizing the costs, while protecting ourselves against a set of likely scenarios $\mathcal{D}$. This leads to the following robust optimization problem:

$$\min_{(\boldsymbol{x},\boldsymbol{z}) \in \mathcal{X}} \max_{\boldsymbol{d} \in \mathcal{D}} F(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{d}). \tag{2}$$

We propose to use a cutting plane approach to solve Problem (2). Given a finite set of scenarios $\hat{\mathcal{D}} = \{\boldsymbol{d}^{(i)} : i \in \mathcal{S}\} \subseteq \mathcal{D}$, we first observe that the *restricted master problem* $\min_{(\boldsymbol{x},\boldsymbol{z}) \in \mathcal{X}} \max_{\boldsymbol{d} \in \hat{\mathcal{D}}} F(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{d})$ can be formulated as a mixed integer linear program:

$$\begin{aligned}
\min_{\boldsymbol{x},\boldsymbol{z},\Delta,\boldsymbol{\delta}} \quad & \sum_{m \in \mathcal{M}} c_m^f z_m + \Delta && && \text{(3a)} \\
\text{s.t.} \quad & \delta_{im} \geq \sum_{j \in \mathcal{J}} x_{jm} d_j^{(i)} - z_m T_m, && \forall i \in \mathcal{S}, \forall m \in \mathcal{M}, && \text{(3b)} \\
& \delta_{im} \geq 0, && \forall i \in \mathcal{S}, \forall m \in \mathcal{M}, && \text{(3c)} \\
& \Delta \geq \sum_{m \in \mathcal{M}} c_m^o \delta_{im}, && \forall i \in \mathcal{S}, && \text{(3d)} \\
& (\boldsymbol{x}, \boldsymbol{z}) \in \mathcal{X} && && \text{(3e)}
\end{aligned}$$

The objective function (3a) minimizes the fixed cost $\sum_m c_m^f z_m$ and the robust overtime cost Δ, equations (3b) and (3c) define the overtime δ_{im} for machine m and scenario $\boldsymbol{d}^{(i)}$, and (3d) makes sure that Δ is the worst case overtime cost over all scenarios in $\hat{\mathcal{D}}$. Finally, (3e) ensures that $(\boldsymbol{x}, \boldsymbol{z})$ is a valid allocation.

We also point out that when several machines have the same values for c_m^f, c_m^o and T_m, it is possible to strengthen the above formulation by using symmetry-breaking constraints; see [19].

Now, we introduce the separation problem, which, given a current solution $(\boldsymbol{x}, \boldsymbol{z})$ of the restricted master problem (3), finds the worst scenario within the uncertainty set $\mathcal{D}$,

$$\max_{\boldsymbol{d} \in \mathcal{D}} F(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{d}). \tag{4}$$

The cutting plane algorithm to solve Problem (2) can be summarized as follows. We assume that a reference scenario (or *nominal scenario*) $\hat{\boldsymbol{d}} \in \mathcal{D}$ is given. For example, this can be the mean scenario $\hat{d}_i = \mathbb{E}[d_i]$, but this is not necessary. Start with $\mathcal{D}^{(1)} = \{\hat{\boldsymbol{d}}\}$. At iteration $k \in \mathbb{N}$, solve Problem (3) for $\hat{\mathcal{D}} = \mathcal{D}^{(k)}$ and set $(\boldsymbol{x}^{(k)}, \boldsymbol{z}^{(k)})$ to the optimal solution. Then, solve Problem (4) with $(\boldsymbol{x}, \boldsymbol{z}) = (\boldsymbol{x}^{(k)}, \boldsymbol{z}^{(k)})$, insert the worst case scenario $\boldsymbol{d}^{(k)}$ in the restricted uncertainty set, $\mathcal{D}^{(k+1)} = \mathcal{D}^{(k)} \cup \{\boldsymbol{d}^{(k)}\}$, and iterate.

It is straightforward that at each iteration, the optimal value of Problem (4) is an upper bound for the value of (2), while the optimal value of (3) provides a lower bound. Clearly, this process finishes after a finite number of steps, because $\mathcal{X}$ is finite, and $\boldsymbol{x}^{(j)} = \boldsymbol{x}^{(k)}$ with $j < k$ would indicate that the process has converged at iteration k; cf. [38]. This cutting-plane approach is summarized in Algorithm 1, where we use an additional tolerance parameter $\varepsilon > 0$ to speed-up the convergence.

We also point out that we can refine this approach, by using *lazy constraints* that add scenarios while the branch-and-bound tree of Problem (3) is being explored, see [8] for more details.

As mentioned in the introduction, this work is motivated by an application to surgery scheduling, where each job typically follows a lognormal distribution. In the next section, we show how to solve Problem (4) efficiently for adequate uncertainty sets.

3 Solving the separation problem

3.1 Separation problem over the MRORA uncertainty set

We recall the definition of the MRORA uncertainty set:

$$\mathcal{D}_{\text{MRORA}} = \{\boldsymbol{d} \in \mathbb{R}^n : \ \forall j, \ell_j \leq d_j \leq u_j; \ \sum_{j \in \mathcal{J}} \frac{d_j - \ell_j}{u_j - \ell_j} \leq \tau\}, \tag{5}$$

where ℓ_j and u_j are lower and upper bounds for the duration of job j, and τ is a parameter to control the *budget of uncertainty*. As stated in the introduction, [19] proposed a compact MIP

Algorithm 1 `(ROBUST_CUTS)`

Input: Instance defined by: $\mathcal{J}, \mathcal{M}, \mathcal{D} \subseteq \mathbb{R}_+^{\mathcal{J}}$,
$\forall m \in \mathcal{M}, \quad c_m^o, c_m^f, T_m \in \mathbb{R}_+$,
nominal scenario $\hat{\boldsymbol{d}} \in \mathcal{D}$,
tolerance parameter $\varepsilon > 0$;

Output: $\varepsilon-$approximate solution $(\boldsymbol{x}^*, \boldsymbol{z}^*)$ of Problem (2).

$L \leftarrow 0$
$U \leftarrow +\infty$
$\hat{\mathcal{D}} \leftarrow \{\hat{\boldsymbol{d}}\}$
while $U > (1+\varepsilon)L$ **do**
  $(\boldsymbol{x}, \boldsymbol{z}) \leftarrow$ optimal solution of the restricted master problem (3)
  $F^* \leftarrow$ optimal value of the restricted master problem (3)
  $L \leftarrow \max(L, F^*)$
  $\boldsymbol{d^*} \leftarrow$ optimal solution of the separation problem (4)
  if $F(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{d^*}) < U$ **then**
    $U \leftarrow F(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{d^*})$
    $(\boldsymbol{x}^*, \boldsymbol{z}^*) \leftarrow (\boldsymbol{x}, \boldsymbol{z})$
  end if
  $\hat{\mathcal{D}} \leftarrow \hat{\mathcal{D}} \cup \{\boldsymbol{d^*}\}$
end while
return $(\boldsymbol{x^*}, \boldsymbol{z^*})$

formulation for the robust OR allocation problem with respect to this uncertainty set, but this formulation was recently shown to be inexact [3]. The problem can still be solved exactly by using the cutting plane procedure presented in the previous section. To do this, we need to solve the separation problem (4) for the uncertainty set $\mathcal{D}_{\text{MRORA}}$. We show that it is possible to solve this problem efficiently in this section. Alternatively, a reformulation approach is presented in Section 4.1, where an equivalent MIP model is presented for Problem (2), when $\mathcal{D}$ is polyhedral.

For a given solution $(\boldsymbol{x}, \boldsymbol{z}) \in \mathcal{X}$, the separation problem (4) over $\mathcal{D}_{\text{MRORA}}$ takes the form:

$$\max_{\boldsymbol{d} \in \mathcal{D}_{\text{MRORA}}} \quad \sum_{m \in \mathcal{M}} c_m^f z_m + c_m^o \left(\sum_{j \in \mathcal{J}} x_{jm} d_j - T_m \right)^+ . \tag{6}$$

Now, we will show that the objective function can be linearized, at the cost of an enumeration over all 2^p subsets of machines. For instances with a reasonable number of machines (say, $p \leq 20$), this is not critical, since each of the 2^p subproblems are very easy to solve. Observe that for all $u \in \mathbb{R}$, we have $u^+ = \max_{\epsilon \in \{0,1\}} \epsilon u$. We can use this to reformulate each term with a $(\cdot)^+$ as a maximum over $\epsilon_m \in \{0, 1\}$:

$$F(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{d}) = \sum_{m \in \mathcal{M}} c_m^f z_m + c_m^o \max_{\epsilon_m \in \{0,1\}} \epsilon_m \big(\sum_{j \in \mathcal{J}} x_{jm} d_j - T_m \big).$$

Then, we can switch the order of the maximization over $\boldsymbol{\epsilon} \in \{0,1\}^p$ and over $\boldsymbol{d} \in \mathcal{D}_{\text{MRORA}}$, so the separation problem (6) is equivalent to

$$\max_{\boldsymbol{\epsilon} \in \{0,1\}^p} \quad \max_{d \in \mathcal{D}_{\text{MRORA}}} \quad \sum_{m \in \mathcal{M}} c_m^f z_m + c_m^o \epsilon_m \left(\sum_{j \in \mathcal{J}} x_{jm} d_j - T_m \right). \tag{7}$$

To solve this problem, we can solve the inner maximization problem for the 2^p values of the vector $\boldsymbol{\epsilon} \in \{0,1\}^p$. After a change of variables $d_j = \ell_j + (u_j - \ell_j)\, r_j$, for a fixed vector $\boldsymbol{\epsilon}$, the inner problem becomes

$$\begin{aligned} \sum_{m \in \mathcal{M}} c_m^f z_m - c_m^o \epsilon_m T_m + \quad \max_{\boldsymbol{r}} \quad & \sum_{j \in \mathcal{J}} \sum_{m \in \mathcal{M}} x_{jm} c_m^o \epsilon_m (\ell_j + (u_j - \ell_j)\, r_j) \\ & 0 \leq r_j \leq 1 \qquad (\forall j \in \mathcal{J}) \\ & \sum_j r_j \leq \tau \end{aligned} \tag{8}$$

This is a particular linear programming (LP) problem, which can be solved analytically. Indeed, we recognize a fractional knapsack problem, for which the greedy algorithm is well-known to be optimal [16]. More precisely, denote by $\boldsymbol{\delta}$ the vector with components $\delta_j = \sum_{m \in \mathcal{M}} x_{jm} c_m^o \epsilon_m (u_j - \ell_j)$, and assume that the components of $\boldsymbol{\delta}$ are sorted as $\delta_{j_1} \geq \delta_{j_2} \geq \cdots \geq \delta_{j_n}$. Then, a solution to Problem (8) is obtained by setting $r_j = 1$ (i.e., $d_j = u_j$) for the jobs $j_1, \ldots, j_{\lfloor \tau \rfloor}$, $r_j = \tau - \lfloor \tau \rfloor$ (i.e., $d_j = \ell_j + (\tau - \lfloor \tau \rfloor)(u_j - \ell_j)$) for the job $j = j_{\lceil \tau \rceil}$, and $r_j = 0$ ($d_j = \ell_j$) for the remaining jobs.

3.2 Separation problem over a lognormal confidence region

If we assume that $\log d_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$, it is natural to consider an uncertainty set of the form $\mathcal{D}_r := \{\boldsymbol{d} \in \mathbb{R}^n_+ : \log(\boldsymbol{d}) \in \mathcal{E}_r\}$, where $\mathcal{E}_r := \left\{\boldsymbol{y} \in \mathbb{R}^n : \sum_{j=1}^n \sigma_j^{-2}(y_j - \mu_j)^2 \leq r^2\right\}$ for some $r > 0$. Note that the set $\mathcal{D}_r$ defined above is simply a log-transformation of some confidence ellipsoid of the multivariate normal law $\mathcal{N}(\boldsymbol{\mu}, \mathrm{Diag}(\boldsymbol{\sigma})^2)$. Here, the parameter r allows to tune the budget of uncertainty, larger values of r leading to larger uncertainty sets. We shall discuss the choice of r in Section 3.3.

By using the same technique as in the previous section, Problem (4) may be reformulated as

$$\max_{\boldsymbol{\epsilon} \in \{0,1\}^p} \max_{\boldsymbol{d} \in \mathcal{D}_r} \sum_{m \in \mathcal{M}} c_m^f z_m + \epsilon_m c_m^o (\sum_{j \in \mathcal{J}} x_{jm} d_j - T_m), \tag{9}$$

which reduces to solving the inner maximization problem for the 2^p values of the vector $\boldsymbol{\epsilon} \in \{0,1\}^p$. Now, we make the change of variables $y_j = \log d_j$. For a fixed $\boldsymbol{\epsilon}$, the value of the inner maximization problem equals

$$\sum_{m \in \mathcal{M}} c_m^f z_m - \epsilon_m c_m^o T_m + \max_{\boldsymbol{y} \in \mathcal{E}_r} \sum_{j \in \mathcal{J}} v_j e^{y_j}, \tag{10}$$

where we have set $v_j := \sum_{m \in \mathcal{M}} \epsilon_m c_m^o x_{jm} \geq 0$. If we put aside the trivial case $\boldsymbol{v} = \boldsymbol{0}$, the necessary Karush-Kuhn-Tucker (KKT) conditions for the maximization problem in (10) can be stated as follows:

Lemma 3.1. *Let $\boldsymbol{v} \in \mathbb{R}^n_+, \boldsymbol{v} \neq \boldsymbol{0}$, and let the vector $\boldsymbol{y}$ be an optimal solution of the problem $\max_{\boldsymbol{y} \in \mathcal{E}_r} \sum_{j \in \mathcal{J}} v_j e^{y_j}$, then there exists a Lagrange multiplier $\lambda > 0$ such that*

$$\begin{cases} \forall j \in \mathcal{J}, \lambda(y_j - \mu_j)\sigma_j^{-1} = \sigma_j v_j e^{y_j} \\ \sum_{j \in \mathcal{J}} (y_j - \mu_j)^2 \sigma_j^{-2} = r^2. \end{cases} \tag{11}$$

Proof. The optimization problem of the lemma has a single constraint, which is clearly active at the optimum (i.e., $\boldsymbol{y}$ lies on the boundary of $\mathcal{E}_r$). Hence, the optimum cannot be at $\boldsymbol{y} = \boldsymbol{\mu}$, which means that the gradient of the constraint does not vanish at the optimum. So the constraint is qualified, and $\boldsymbol{y}$ must satisfy the Karush-Kuhn-Tucker (KKT) optimality conditions (see, e.g., [42]), which reads: $\exists \lambda \in \mathbb{R}$:

$$\forall j \in \mathcal{J}, \lambda(y_j - \mu_j)\sigma_j^{-2} = v_j e^{y_j} \qquad \text{[stationarity]} \tag{12a}$$

$$\sum_{j \in \mathcal{J}} (y_j - \mu_j)^2 \sigma_j^{-2} \leq r^2 \qquad \text{[primal feasibility]} \tag{12b}$$

$$\lambda \geq 0 \qquad \text{[dual feasibility]} \tag{12c}$$

$$\lambda(r^2 - \sum_{j \in \mathcal{J}} (y_j - \mu_j)^2 \sigma_j^{-2}) = 0 \qquad \text{[comp. slackness]} \tag{12d}$$

This can be further simplified to the condition (11) of the lemma, by observing that λ cannot be equal to 0 (take a j such that $v_j > 0$, then Equation (12a) implies $y_j > \mu_j$ and $\lambda > 0$). □

We can find the value of λ by substituting $(y_j - \mu_j)\sigma_j^{-1} = \lambda^{-1}\sigma_j v_j e^{y_j}$ in the second equation: $\lambda = r^{-1}(\sum_j \sigma_j^2 v_j^2 e^{2y_j})^{1/2}$. Substituting back in the first equation, we find that for all $j \in \mathcal{J}$,

$$(y_j - \mu_j)(r\sigma_j)^{-1} = \sigma_j v_j e^{y_j}(\sum_j \sigma_j^2 v_j^2 e^{2y_j})^{-1/2}.$$

In other words, the vector $\boldsymbol{w} := \mathrm{Diag}(r\boldsymbol{\sigma})^{-1}(\boldsymbol{y} - \boldsymbol{\mu})$ is a fixed point of the map $g : \boldsymbol{w} \mapsto f(\boldsymbol{w})/\|f(\boldsymbol{w})\|$ which maps the unit sphere of $\mathbb{R}^n$ onto itself, where

$$f(\boldsymbol{w}) := \boldsymbol{\sigma} \circ \boldsymbol{v} \circ \exp(\boldsymbol{\mu} + r\boldsymbol{\sigma} \circ \boldsymbol{w}),$$

the exponential is elementwise, and $\circ$ denotes the Hadamard (elementwise) product: $(\boldsymbol{a} \circ \boldsymbol{b})_i = a_i b_i$.

The next results give a sufficient condition – almost always satisfied in practice, cf. discussion in Section 3.4 – which guarantees that fixed point iterations of g converge, and we can use the fixed point to find a global optimum of (10). To do this, we prove the following result, which relies on *Hilbert's projective metric* d_H on the cone $K := \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x} > \boldsymbol{0}\}$. It is defined by $\forall \boldsymbol{x}, \boldsymbol{y} \in K,\ d_H(\boldsymbol{x}, \boldsymbol{y}) := \log \max_i \frac{x_i}{y_i} + \log \max_j \frac{y_j}{x_j}$; see [13]. Note that d_H is actually a metric over the space of rays of the cone K. However, d_H defines a metric over the subsets $K^{(1)} := \{\boldsymbol{x} \in K : \|x\|_1 = 1\}$ or $K^{(2)} := \{\boldsymbol{x} \in K : \|x\|_2 = 1\}$.

We list hereafter a few important properties of d_H, which are proved e.g. in [39]:

(i) $\forall \boldsymbol{x}, \boldsymbol{y} \in K,\ d_H(\boldsymbol{x}, \boldsymbol{y}) = 0$ implies $\boldsymbol{x} = \alpha \boldsymbol{y}$ for some $\alpha > 0$

(ii) $\forall \boldsymbol{x}, \boldsymbol{y} \in K,\ d_H(\boldsymbol{x}, \boldsymbol{y}) = d_H(\boldsymbol{y}, \boldsymbol{x})$

(iii) $\forall \boldsymbol{x}, \boldsymbol{y} \in K,\ \forall \lambda > 0,\ d_H(\boldsymbol{x}, \lambda \boldsymbol{y}) = d_H(\boldsymbol{x}, \boldsymbol{y})$

(iv) $\forall \boldsymbol{x}, \boldsymbol{y} \in K,\ \forall \boldsymbol{u} \in K,\ d_H(\boldsymbol{u} \circ \boldsymbol{x}, \boldsymbol{u} \circ \boldsymbol{y}) = d_H(\boldsymbol{x}, \boldsymbol{y})$

(v) $(K^{(2)}, d_H)$ is a complete metric space.

The next result gives the Lipschitz constant of the (elementwise) exponential function over $K^{(2)}$. It will be useful to ensure the convergence of fixed point iterations of g:

Theorem 3.2. *The function $h : \boldsymbol{x} \mapsto \exp(\boldsymbol{x})$ is contractant for Hilbert's projective metric over $K^{(2)}$, with a global Lipschitz constant equal to $\frac{1}{\sqrt{2}}$:*

$$\forall \boldsymbol{x}, \boldsymbol{y} \in K^{(2)},\ d_H\big(h(\boldsymbol{x}), h(\boldsymbol{y})\big) \leq \frac{1}{\sqrt{2}}\, d_H(\boldsymbol{x}, \boldsymbol{y}).$$

The proof of this result is included in the appendix. It relies on an intermediate result (Theorem A.1), which gives a formula to compute local Lipschitz constants for a function defined over $K^{(2)}$. We are now ready to prove the following proposition, which gives a simple condition ensuring the convergence of the fixed point iterations.

Proposition 3.3. *Assume that for all $j \in \mathcal{J}$, $r\sigma_j < \sqrt{2}$. Then, there exists a point $\boldsymbol{w}^* \in K^{(2)}$ such that the fixed point iterations $g(g(\cdots g(\boldsymbol{w}_0)))$ converge to $\boldsymbol{w}^*$ for all $\boldsymbol{w}_0 \in \mathbb{R}^n$. Moreover, $\boldsymbol{y}^* := \boldsymbol{\mu} + r\,\mathrm{Diag}(\boldsymbol{\sigma})\boldsymbol{w}^*$ is a global optimum of Problem (10).*

Algorithm 2 (LOGNORMAL_ADVERSE)

Input: Instance defined by: $\forall m \in \mathcal{M}, \quad c_m^o, c_m^f, T_m \in \mathbb{R}_+$,
Parameters for the uncertainty set $\mathcal{D}_r$: $\boldsymbol{\mu} \in \mathbb{R}^n, \boldsymbol{\sigma} \in \mathbb{R}^n_+, r > 0$,
Solution of the RMP $(\boldsymbol{x}, \boldsymbol{z})$,
tolerance parameter $\nu > 0$.

Output: $\nu-$approximate solution $\boldsymbol{d} \in \mathcal{D}_r$ of Problem (4).

$OPT \leftarrow \sum_{m\in\mathcal{M}} c_m^f z_m$ ▷Initialization, for the case $\boldsymbol{\epsilon} = \mathbf{0}_p$
$\boldsymbol{d}^* \leftarrow \exp(\boldsymbol{\mu})$ ▷ in fact, any $\boldsymbol{d} \in \mathcal{D}$ is optimal when $\boldsymbol{\epsilon} = \mathbf{0}_p$
for all $\epsilon \in \{0,1\}^p, \epsilon \neq \mathbf{0}_p$ **do**
  $v_j \leftarrow \sum_{m\in\mathcal{M}} \epsilon_m c_m^o x_{jm} \quad (\forall j \in \mathcal{J})$
  $\boldsymbol{w}^{(0)} \leftarrow \mathbf{0}_n$
  $\Delta \leftarrow +\infty$
  $i \leftarrow 0$
  while $\Delta > \nu$ **do**
    $\boldsymbol{f} \leftarrow \boldsymbol{\sigma} \circ \boldsymbol{v} \circ \exp(\boldsymbol{\mu} + r\boldsymbol{\sigma} \circ \boldsymbol{w}^{(i)})$
    $\boldsymbol{w}^{(i+1)} \leftarrow \frac{\boldsymbol{f}}{\|\boldsymbol{f}\|_2}$
    $\Delta \leftarrow \|\boldsymbol{w}^{(i+1)} - \boldsymbol{w}^{(i)}\|_2$
    $i \leftarrow i + 1$
  end while
  $\boldsymbol{d} \leftarrow \exp(\boldsymbol{\mu} + r\boldsymbol{\sigma} \circ \boldsymbol{w}^{(i)})$
  $OPT_\epsilon \leftarrow \sum_{m\in\mathcal{M}} c_m^f z_m + \epsilon_m c_m^o (\sum_{j\in\mathcal{J}} x_{jm} d_j - T_m)$
  if $OPT_\epsilon > OPT$ **then**
    $OPT \leftarrow OPT_\epsilon$
    $\boldsymbol{d^*} \leftarrow \boldsymbol{d}$
  end if
end for
return $\boldsymbol{d^*}$

Proof. If $v_j = 0$ for some $j \in \mathcal{J}$, then it is clear that the fixed-point iterates $\boldsymbol{w}^{(k)} = g^k(\boldsymbol{w}_0)$ will satisfy $w_j^{(k)} = 0$ for all $k \geq 1$. So we assume without loss of generality that $v_j > 0$ for all $j \in \mathcal{J}$ for the rest of this proof.

Note that the existence of a fixed point of g is guaranteed by Brouwer's theorem (see [26]), and any fixed point must lie in $K^{(2)}$. By using the properties of Hilbert's projective metric, we find that

$$\begin{aligned} \forall \boldsymbol{x}, \boldsymbol{y} \in K, d_H\big(g(\boldsymbol{x}), g(\boldsymbol{y})\big) &= d_H\big(f(\boldsymbol{x}), f(\boldsymbol{y})\big) \\ &= d_H\big(\exp(r\boldsymbol{\sigma} \circ \boldsymbol{x}), \exp(r\boldsymbol{\sigma} \circ \boldsymbol{y})\big) \\ &\leq r\|\sigma\|_\infty d_H(e^{\boldsymbol{x}}, e^{\boldsymbol{y}}). \end{aligned}$$

In the above expression, the first equality follows from Properties (ii) and (iii), and the second equality follows from (iv). Therefore, Theorem 3.2 implies that g is contractant for Hilbert's metric over $K^{(2)}$ if $r\|\boldsymbol{\sigma}\|_\infty < \sqrt{2}$. Since $(K^{(2)}, d_H)$ is a complete metric space (property (v)),

Banach's fixed point theorem ensures the uniqueness of a fixed point $\boldsymbol{w}^*$ and the convergence of fixed point iterations when $r\|\boldsymbol{\sigma}\|_\infty < \sqrt{2}$; cf. [26]. In this case, $\boldsymbol{y}^* := \boldsymbol{\mu} + r\boldsymbol{\sigma} \circ \boldsymbol{w}^*$ is the unique solution of the necessary conditions (11), so $\boldsymbol{y}^*$ maximizes $\sum_j v_j e^{y_j}$ over $\mathcal{E}_r$. □

Our approach is summarized in Algorithm 2, which can be used to solve the separation problem at line 8 of Algorithm 1; the condition of Proposition 3.3 ensures that this procedure converges.

3.3 Choice of the parameter r

Care must be taken in setting the value of r defining $\mathcal{E}_r$, to avoid overconservatism. Indeed, the optimal solution of Problem (2) does not only protect against scenarios in $\mathcal{D}_r$, but also against all duration scenarios in $\bar{\mathcal{D}}_r = \{\boldsymbol{d} - \boldsymbol{u} : \boldsymbol{d} \in \mathcal{D}_r, \boldsymbol{u} \geq \boldsymbol{0}\}$. For the lognormal model $\log d_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$, we can give an analytical formula for the probability that a scenario lies in $\bar{\mathcal{D}}_r$:

Lemma 3.4. *If the job durations are lognormal and independent, $\log d_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$, then the probability $P_n(r) := \mathbb{P}[\boldsymbol{d} \in \bar{\mathcal{D}}_r]$ is given by*

$$P_n(r) := \Phi(r)^n - (\Phi(r) - \frac{1}{2})^n + \frac{1}{2^n} F_{\chi^2_n}(r^2), \tag{13}$$

where Φ is the standard normal cumulative distribution function (CDF), and $F_{\chi^2_n}$ is the CDF of the χ^2-distribution with n degrees of freedom.

Proof. The probability that $\boldsymbol{d}$ lies in $\bar{\mathcal{D}}_r$ is the same as the probability that the vector with coordinates $X_j = \sigma_j^{-1}(\log d_j - \mu_j)$, which has a standard multivariate normal distribution $X \sim \mathcal{N}(\boldsymbol{0}, I)$, lies in the set $S = \{\boldsymbol{y} - \boldsymbol{u} : \boldsymbol{y} \in \mathbb{R}^n, \|\boldsymbol{y}\|_2 \leq r,\ \boldsymbol{u} \geq \boldsymbol{0}\}$, which corresponds to the grey area depicted in Figure 1 for the case $n = 2$. Define $S_1 = \{\boldsymbol{y} \in \mathbb{R}^n :\ y_j \leq r, \forall j \in \mathcal{J}\}$, $S_2 = \{\boldsymbol{y} \in \mathbb{R}^n :\ 0 \leq y_j \leq r, \forall j \in \mathcal{J}\}$, $S_3 = \{\boldsymbol{y} \in \mathbb{R}^n_+ :\ \|y\|_2 \leq r\}$, and observe that

$$S = S_1 \setminus (S_2 \setminus S_3), \qquad \text{with } (S_2 \setminus S_3) \subseteq S_1 \text{ and } S_3 \subseteq S_2.$$

In Figure 1, the set S_1 corresponds to the quadrant below the point A, and $S_2 \setminus S_3$ is the region denoted by stripes, which is the difference between the "cube" S_2 with corners A and O and the part S_3 of the ball of radius r that is situated in the positive quadrant. Hence, from the inclusion-exclusion principle we have

$$P_n(r) = \mathbb{P}[X \in S_1] - \mathbb{P}[X \in S_2] + \mathbb{P}[X \in S_3].$$

Finally, the result of the lemma follows from the analytical expressions of $\mathbb{P}[X \in S_i]$, $i = 1, 2, 3$. For S_1 and S_2, these probabilities are easily obtained because the X_j's are independent: $\mathbb{P}[X \in S_1] = \prod_{j \in \mathcal{J}} \mathbb{P}[X_j \leq r] = \Phi(r)^n$ and $\mathbb{P}[X \in S_2] = \prod_{j \in \mathcal{J}} \mathbb{P}[0 \leq X_j \leq r] = (\Phi(r) - \Phi(0))^n = (\Phi(r) - \frac{1}{2})^n$. For S_3, we use the symmetry of $\mathcal{N}(\boldsymbol{0}, I)$ around $\boldsymbol{0}$ and the definition of the χ^2 distribution:

$$\mathbb{P}[X \in S_3] = \frac{1}{2^n} \mathbb{P}\left[\sum_{i=1}^n X_i^2 \leq r^2\right] = \frac{1}{2^n} F_{\chi^2_n}(r^2).$$

□

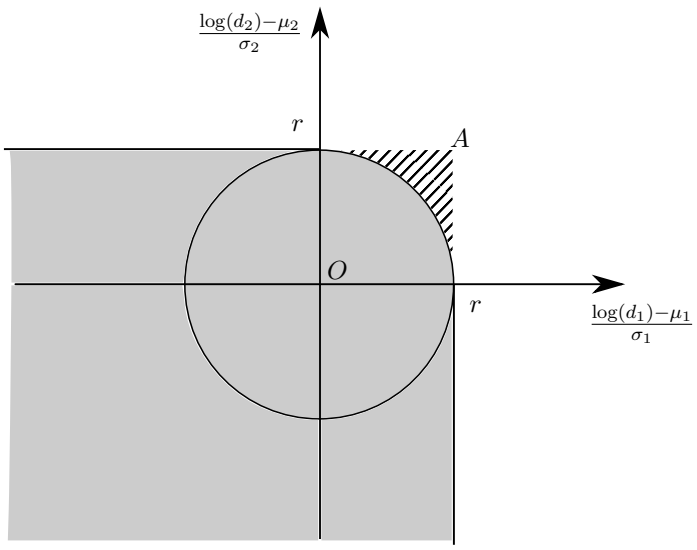


Figure 1: The probability $P_n(r)$ that $\boldsymbol{d} \in \bar{\mathcal{D}}_r$ is equal to the probability that a random vector $X \sim \mathcal{N}(0, \boldsymbol{I}) \in \mathbb{R}^n$ lies in the grey area. The formula (13) expresses this probability as the difference between the probability of the quadrant below A and the striped region.

For a confidence level α, we can hence choose r by solving the equation $P_n(r) = 1 - \alpha$. Then, Problem (2) minimizes an upper bound of the $(1-\alpha)-$quantile of $F(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{d})$:

Proposition 3.5. *Assume that the d_j's follow independent lognormal distributions, with $\log d_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$. Let $(\boldsymbol{x}^*, \boldsymbol{z}^*)$ be the optimal solution of the robust parallel machines scheduling problem (2) and let $F^* := \max_{\boldsymbol{d} \in \mathcal{D}_r} F(\boldsymbol{x}^*, \boldsymbol{z}^*; \boldsymbol{d})$ be its optimal value, where*

$$\mathcal{D}_r := \left\{ \boldsymbol{d} \in \mathbb{R}^n_+ : \sum_{j=1}^{n} \sigma_j^{-2} (\log d_j - \mu_j)^2 \leq r^2 \right\}$$

and $P_n(r) = 1 - \alpha$. Then,

$$\mathbb{P}\Big[F(\boldsymbol{x}^*, \boldsymbol{z}^*; \boldsymbol{d}) \leq F^*\Big] \geq 1 - \alpha.$$

Proof. We know that $F(\boldsymbol{x}^*, \boldsymbol{z}^*; \boldsymbol{d}) \leq F^*$ for all $\boldsymbol{d} \in \mathcal{D}_r$, by definition of F^*. But this is also true for all scenarios that are dominated by a scenario in $\mathcal{D}_r$, i.e., for $\boldsymbol{d} \in \bar{\mathcal{D}}_r := \mathcal{D}_r - \mathbb{R}^n_+$. It follows that $\mathbb{P}\Big[F(\boldsymbol{x}^*, \boldsymbol{z}^*; \boldsymbol{d}) \leq F^*\Big] \geq \mathbb{P}[\boldsymbol{d} \in \bar{\mathcal{D}}_r]$ and by construction $\mathbb{P}[\boldsymbol{d} \in \bar{\mathcal{D}}_r] = P_n(r) = 1 - \alpha$. □

3.4 Discussion on the assumptions of Proposition 3.3

Estimates of μ_j and σ_j usually come from an analysis of historical data. It seems reasonable to assume that one can obtain estimates $\sigma_j \leq 0.5$, because $\sigma_j = 0.5$ already allows huge deviations from the nominal scenario: 95%-confidence interval is $[0.37 m_j, 2.67 m_j]$, where $m_j := e^{\mu_j}$ is the median of d_j. In this situation, if we choose r by solving $P_n(r) = 1 - \alpha$, the

condition $r\|\boldsymbol{\sigma}\|_\infty < \sqrt{2}$ is satisfied for $n \leq 21$ jobs at the robustness level $\alpha = 0.05$, and for $n \leq 45$ at $\alpha = 0.1$.

4 Alternative approaches

4.1 Pseudo-compact MIP reformulation for polyhedral uncertainty

In this section, we present an alternative to the cutting plane approach of Section 3.1, which works when $\mathcal{D}$ is *any* nonempty polyhedral set of the form $\mathcal{D} := \{\boldsymbol{d} : A\boldsymbol{d} \leq \boldsymbol{b}\}$ (in particular, this includes the case $\mathcal{D} = \mathcal{D}_{\text{MRORA}}$). We will show that Problem (2) can be reformulated as a single MIP, but its size grows as 2^p (hence the term *pseudo-compact* in the title of this section). For the motivating OR allocation problem, at the scale of a department with $p \leq 6$ operating rooms, our computational results show that this approach is more efficient than the cutting planes, cf. Table 1.

By using the formulation (7) of the separation problem, the robust allocation problem can be reformulated as

$$\begin{aligned} \min_{(\boldsymbol{x},\boldsymbol{z})\in\mathcal{X},\Delta} \quad & \sum_{m\in\mathcal{M}} c^f_m z_m + \Delta \\ s.t. \quad & \forall \boldsymbol{\epsilon} \in \{0,1\}^p, \qquad \max_{\{\boldsymbol{d}:A\boldsymbol{d}\leq\boldsymbol{b}\}} \quad \sum_{m\in\mathcal{M}} c^o_m \epsilon_m \Big(\sum_{j\in\mathcal{J}} x_{jm} d_j - T_m\Big) \leq \Delta. \end{aligned} \tag{14}$$

This is an optimization problem in which the optimal value of an LP appears in some constraints. By using a standard technique of robust optimization, see e.g. [7], we can reformulate it as a set of equivalent linear constraints, by dualizing the maximization problems. Indeed, from the strong duality theorem of Linear Programming (see, e.g., [12]), we have

$$\begin{aligned} \max_{\boldsymbol{d}} \quad & \sum_{m\in\mathcal{M}}\sum_{j\in\mathcal{J}} c^o_m \epsilon_m x_{jm} d_j \quad = \quad \min_{\boldsymbol{y}\geq\boldsymbol{0}} \quad \boldsymbol{b}^T\boldsymbol{y} \\ s.t. \quad & A\boldsymbol{d} \leq \boldsymbol{b} \qquad\qquad\qquad s.t. \quad \boldsymbol{a}_j^T\boldsymbol{y} = \sum_{m\in\mathcal{M}} \epsilon_m c^o_m x_{jm} \qquad (\forall j \in \mathcal{J}), \end{aligned} \tag{15}$$

where $\boldsymbol{a}_j$ denotes the jth column of A (strong duality holds since $\mathcal{D}$ is nonempty, so the primal problem is feasible). Now, enumerate the $2^p - 1$ vectors of $\{0,1\}^p \setminus \{\boldsymbol{0}\}$ as $\boldsymbol{\epsilon}^{(1)}, \ldots, \boldsymbol{\epsilon}^{(2^p-1)}$ (note that the trivial case $\boldsymbol{\epsilon} = \boldsymbol{0}$ can be left aside, since it simply yields the constraint $\Delta \geq 0$). Plugging the equality (15) in each constraint of Problem (14) yields the following

Proposition 4.1. *Assume that the polyhedral uncertainty set $\mathcal{D} := \{\boldsymbol{d} : A\boldsymbol{d} \leq \boldsymbol{b}\}$ is nonempty.*

Then, the robust allocation problem (2) *is equivalent to the following MIP:*

$$\begin{aligned}
\min_{\boldsymbol{x},\boldsymbol{z},\Delta,\{\boldsymbol{y}_i\}} \quad & \sum_{m\in\mathcal{M}} c_m^f z_m + \Delta && && (16)\\
s.t. \quad & \boldsymbol{b}^T\boldsymbol{y}_i - \sum_{m\in\mathcal{M}} c_m^o \epsilon_m^{(i)} T_m \le \Delta && (\forall i = 1,\dots,2^p-1)\\
& \boldsymbol{a}_j^T\boldsymbol{y}_i = \sum_{m\in\mathcal{M}} \epsilon_m^{(i)} c_m^o x_{jm} && (\forall i = 1,\dots,2^p-1, \forall j\in\mathcal{J})\\
& \boldsymbol{y}_i \ge \boldsymbol{0} && (\forall i = 1,\dots,2^p-1)\\
& \Delta \ge 0, \quad (\boldsymbol{x},\boldsymbol{z})\in\mathcal{X}.
\end{aligned}$$

4.2 Stochastic Programming

Since robust optimization sometimes leads to overconservative solutions, which may be poor on average, it is also natural to compare the proposed robust optimization approach to stochastic programming, in which the expected value of the cost is minimized:

$$\min_{(\boldsymbol{x},\boldsymbol{z})\in\mathcal{X}} \mathbb{E}_{\boldsymbol{d}}[F(\boldsymbol{x},\boldsymbol{z};\boldsymbol{d})]. \tag{17}$$

The multicut L-shaped algorithm of Birge and Louveaux [10] was used in [19] to solve Problem (17). However, in our situation generating optimality cuts requires to evaluate $\mathbb{E}_{\boldsymbol{d}}[F(\boldsymbol{x}^*,\boldsymbol{z}^*;\boldsymbol{d})]$ and its subgradient, which is a costly process when durations d_j are log-normally distributed. To avoid this computational burden, we use the *sample average approximation* (SAA) method to approximate Problem (17). This approximation method has already been used in the context of OR management. This is the case, e.g., in [35], who used it to set the starting time of surgical cases, in the situation where there is a single surgeon operating in several ORs. Theoretical convergence results for the SAA method were studied in [32], and are illustrated for a resource allocation problem presenting some similarities with Problem (17).

We sample N_S duration scenarios $\boldsymbol{d}^{(1)},\dots,\boldsymbol{d}^{(N_S)}$ and solve the following MIP, which is a small variation of (3), except that there is a term accounting for the average overtime in the objective function, instead of the worst case overtime.

$$\begin{aligned}
\min_{\boldsymbol{x},\boldsymbol{z},\Delta,\boldsymbol{\delta}} \quad & \sum_{m\in\mathcal{M}} c_m^f z_m + \frac{1}{N_S}\sum_{i=1}^{N_S}\Delta_i && && (18a)\\
\text{s.t.} \quad & \delta_{im} \ge \sum_{j\in\mathcal{J}} x_{jm} d_j^{(i)} - z_m T_m, && \forall i\in\{1,\dots,N_S\}, \forall m\in\mathcal{M}, && (18b)\\
& \delta_{im} \ge 0, && \forall i\in\{1,\dots,N_S\}, \forall m\in\mathcal{M}, && (18c)\\
& \Delta_i \ge \sum_{m\in\mathcal{M}} c_m^o \delta_{im}, && \forall i\in\{1,\dots,N_S\}, && (18d)\\
& (\boldsymbol{x},\boldsymbol{z})\in\mathcal{X} && && (18e)
\end{aligned}$$

5 Extension: job selection and cancellation

We mentioned in the introduction that our model could be extended to handle the situation in which the decision maker can cancel planned jobs, or insert new ones. This can be modelled by assigning a penalty λ_j to each job that we decide to postpone to a later planning period. Of course, we can set a prohibitively high penalty to the jobs that must be selected in the considered time period. Since it is not mandatory anymore to select all jobs, the equalities $\forall j \in \mathcal{J},\ \sum_{m\in\mathcal{M}} x_{jm} = 1$ defining the feasibility set must be replaced by inequalities: $\forall j \in \mathcal{J},\ \sum_{m\in\mathcal{M}} x_{jm} \leq 1$. We must also add some terms in the objective function of the restricted master problem (3) to account for the penalties of non-selected jobs. So the objective function becomes:

$$\min_{\boldsymbol{x},\boldsymbol{z},\Delta,\boldsymbol{\delta}} \qquad \sum_{m\in\mathcal{M}} c^f_m z_m + \Delta + \sum_{j\in\mathcal{J}} \lambda_j (1 - \sum_{m\in\mathcal{M}} x_{jm}). \tag{3a'}$$

Finally, note that these modifications do not affect the separation problem, so we can still use Algorithm 2 to solve it.

In practice, it is possible to use a dynamic rule to update the deferral costs λ_j over a rolling horizon. We next present a brief sketch of this idea. We start with a pool of jobs $\mathcal{J}$ that can be allocated to the first time period. After each planning period, non-selected jobs are inserted in the pool of jobs for the next period, while their deferral costs λ_j are increased by some factor, so as to penalize long waiting times. It is also possible to set a due-date for job j, by setting λ_j to a very high value if j has not been selected until the due date.

6 Computational results

This section presents numerical results for the application to OR management that motivated this study. Our instances are based on real data from the department of general surgery of the Charité university hospital in Berlin. The data is presented in the next subsection. Then, we describe the different solution methods to be compared in Section 6.2, and we define some performance metrics in Section 6.3; the influence of the *expected workload* of the instance is discussed in Section 6.4; then, we study the effect of the opening costs c^f_m in Section 6.5. Our robust optimization approach is compared to the earlier robust optimization model of [19] and the new exact solution approaches for the robust counterpart over $\mathcal{D}_{\text{MRORA}}$ in Section 6.6; finally, we evaluate the different strategies with respect to the true durations in Section 6.7.

6.1 Data and instances

We use maximum likelihood estimators to fit the parameters of a lognormal model for the durations of 20,849 surgical procedures performed in the years 2011–2015, and for the time required to prepare and clean-up the OR before and after each operation. Our model is similar to [47], and relies on characteristics of the patient, operation, and surgical team. This model – which will be the object of a future publication – was tested using cross validation.

We construct a set $\mathcal{I}$ of $N = 348$ instances of the OR allocation problem for several regular working days (weekends and holidays excepted) of the period 2015–2016, so there is

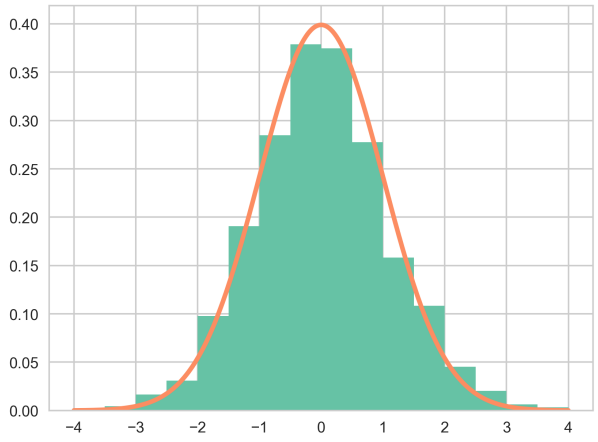


Figure 2: Distribution of the quantity $\frac{\log d_j^{\text{obs}} - \mu_j}{\sigma_j}$, for the 4028 patient blocks of the set of instances $\mathcal{I}$, together with the probability distribution function of a standard normal variable $\mathcal{N}(0,1)$.

no overlap with the training set used by our prediction model. To construct these instances, we group all patients operated by the same surgeon in a block. Note that the duration d_j includes both the total surgical duration for patients in block j and the turnover times, that is, the amount of time needed to clean-up and prepare the room between the patients. Our model gives a prediction for the mean and the variance of each activity within a block. Then, we use the Fenton-Willkinson approximation mentioned in the introduction to fit a lognormal distribution for the total duration of each patient block, $\log d_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$. To illustrate the quality of our duration model, denote by d_j^{obs} the true (observed) duration of block j. If the proposed model $\log d_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$ is correct, then the variables $z_j = \frac{\log d_j^{\text{obs}} - \mu_j}{\sigma_j}$ should be independent samples from the standard normal distribution $\mathcal{N}(0,1)$. The distribution of these variables for the 4028 patient blocks of our instances is depicted in Figure 2, and there is a close fit with the standard normal distribution indeed.

We point out that the Charité hospital in Berlin actually performs both elective and emergency surgery. To accommodate with our model for the allocation of ORs in an elective facility, we handle emergency patients as elective patients when creating our instances. The set of instances $\mathcal{I}$ can be downloaded[1] as a `.json` file which indicates, for each instance, the number p of available ORs and their capacity T_m. In addition, it gives the number n of patient blocks, and for each block, the parameters μ_j and σ_j resulting from our prediction model. The file also indicates the true (observed) duration d_j^{obs} of each patient block, as well as the duration d_j^{sched} that was initially planned for block j in the OR schedule. We point out that in the list of instances, the condition of Proposition 3.3 is satisfied 98% of the time at the confidence level $\alpha = 0.3$, and 75% of the time with $\alpha = 0.05$. Even when the condition

[1] http://page.math.tu-berlin.de/~sagnol/data/instances_OR_allocation_v2.json

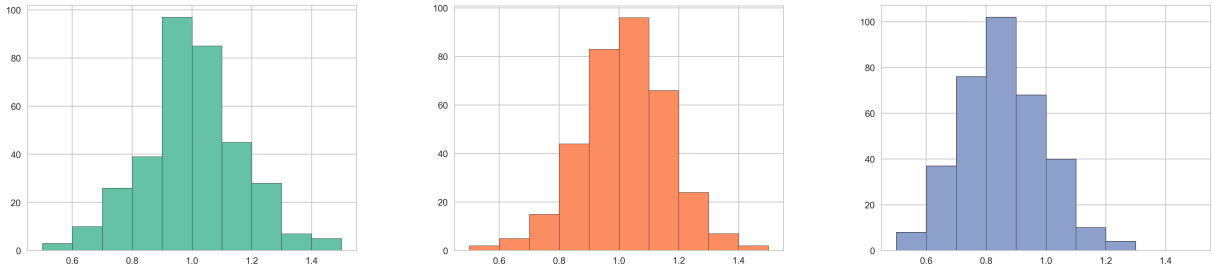


Figure 3: Distribution of the workload of the 348 instances without emergency. *Left:* expected workload $\hat{\rho}$. *Middle:* observed workload ρ^{obs}, according to the true durations. *Right:* scheduled workload ρ^{sched}, according to the planned durations.

is not satisfied, the fixed point iterations always converge in our experiments.

We denote by $\hat{d}_j = \exp\left(\mu_j + \frac{1}{2}\sigma_j^2\right)$ the expected duration of patient block j, and we define the *expected workload* of an instance as

$$\hat{\rho} = \frac{\sum_{j \in \mathcal{J}} \hat{d}_j}{\sum_{m \in \mathcal{M}} T_m},$$

which is a simple measure of how likely overtime will occur. Indeed, whenever $\hat{\rho} \geq 1$ the expected time required to perform all the operations is longer than the regular opening time of the ORs. The distribution of $\hat{\rho}$ over our 348 instances is displayed in Figure 3 (*left*). The expected workload seems to be centered at $\hat{\rho} = 1$, but it varies a lot, typically ranging in the interval $[0.7, 1.3]$. This is in accordance with the *observed workload* $\rho^{\text{obs}} = \frac{\sum_{j \in \mathcal{J}} d_j^{\text{obs}}}{\sum_{m \in \mathcal{M}} T_m}$, cf. Figure 3 (*middle*), which exhibits a similar distribution (although $\hat{\rho}$ is a bit more dispersed than ρ^{obs}). This indicates that the load of operating rooms could be better balanced among the operating days. One reason is that hospitals like the Charité in Berlin currently rely on surgeons' estimates of the durations to book OR time in the available slots. Figure 3 (*right*) shows that the *scheduled workload* $\rho^{\text{sched}} = \frac{\sum_{j \in \mathcal{J}} d_j^{\text{sched}}}{\sum_{m \in \mathcal{M}} T_m}$ is much smaller than the observed workload; it mostly lies in the interval $[0.6, 1.1]$. This is due to the fact that surgeons tend to underestimate the procedure durations (the average of the d_j^{sched}'s is 16% lower than the average of the true durations d_j^{obs}'s), so that their patients fit in the allowed OR time [24]. In contrast, the bias of our prediction model is of only 1.13%, and we think that the use of such models could lead to a much improved workload balance.

For all our instances, except in Section 6.5 where we evaluate the effect of the opening costs, we use a value of $c_m^f = 30$ and $c_m^o = 1$ ($\forall m \in \mathcal{M}$). In other words, opening a room yields the same cost as 30 minutes of overtime. The choice of this particular value is motivated by the payment policies of surgeries in the German healthcare system: For short-term planning, the cost of opening a new room for the hospital should be negligible, since the staffing is already fixed. However, health insurances pay a lump sum corresponding to 30 min. of OR time

to the hospital for each turnover [22], that is, the non-surgical time between two successive surgeries. Using one less room increases the number of turnover by one, and thus, allows the hospital to claim an additional reimbursement for 30 min. of overtime.

6.2 Solution methods

We shall next present detailed results for the $N = 348$ instances of the set $\mathcal{I}$. For each instance, we compare the quality of several scheduling strategies listed below; the first three solution methods also depend on a parameter α specifying the level of robustness.

- The `LRS` (lognormal robust schedule) is the schedule solving Problem (2) for an uncertainty set of the form $\mathcal{D}_r = \{\boldsymbol{d} \in \mathbb{R}^n_+ : \sum_{j=1}^n \sigma_j^{-2}(\log d_j - \mu_j)^2 \leq r^2\}$. We compute it by using Algorithm 1 with a tolerance parameter of $\varepsilon = 0.01$; unless stated otherwise, the separation subproblems are solved with a tolerance parameter $\nu = 10^{-6}$ in Algorithm 2. The parameter r defining the set $\mathcal{D}_r$ is set by solving the equation $P_n(r) = 1 - \alpha$, cf. (13).

- We solve the robust MIP called `MRORA` in [19]. As shown in [3], this MIP minimizes an upper bound of Problem (2) for the uncertainty set $\mathcal{D}_{\texttt{MRORA}}$ defined in Equation (1). We follow the rule suggested in [19] to set the value of the parameters ℓ_j, u_j and τ defining $\mathcal{D}_{\texttt{MRORA}}$, as follows: For each $j \in \mathcal{J}$, we set ℓ_j and u_j to the $\frac{\alpha}{2}$- and $(1 - \frac{\alpha}{2})$-percentile of d_j, respectively, so $[\ell_j, u_j]$ is a $(1-\alpha)-$confidence interval for d_j. Then, τ is set using the *newsvendor rule* described in [19, Section 6.1].

- We also implement two algorithms that solve (exactly) Problem (2) for the uncertainty set $\mathcal{D}_{\texttt{MRORA}}$. On the one hand, the cutting plane procedure with the separation technique described in Section 3.1, and on the other hand, the reformulation as a MIP with $\mathcal{O}(2^p)$ constraints, cf. Section (4.1). We denote these two allocation strategies as `M-CP` and `M-MIP`, respectively.

- To study the tradeoff between stochastic programming and robust optimization, we compute a solution by using the sample average approximation to stochastic programming (`SAA`), as described in Section 4.2.

- As a reference, we use the solution provided by the longest expected processing time (`LEPT`) heuristic for the nominal scenario $\hat{\boldsymbol{d}}$. It basically consists in sorting all cases by decreasing order of (expected) duration $\hat{d}_j$. Then, each case is allocated to one OR in a greedy fashion, by choosing the OR that causes the least increase in overtime. We repeat this procedure by considering that only the k ORs with the most capacity are open, for $k = 1, \ldots, p$; see [19] for a more detailed description. This solution is known to give excellent results when the goal is to minimize the expected value of $F(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{d})$ [19], and yields an approximation guarantee of $\frac{13}{12}$ in the deterministic case, when all ORs have the same capacity T, and the fixed costs of opening are $c^f_m = T$ [17]. Note that in our setting (and in [19]), `LEPT` is a fixed-assignment policy, in which each block is allocated to an OR *before the execution of the schedule*. In some other work on

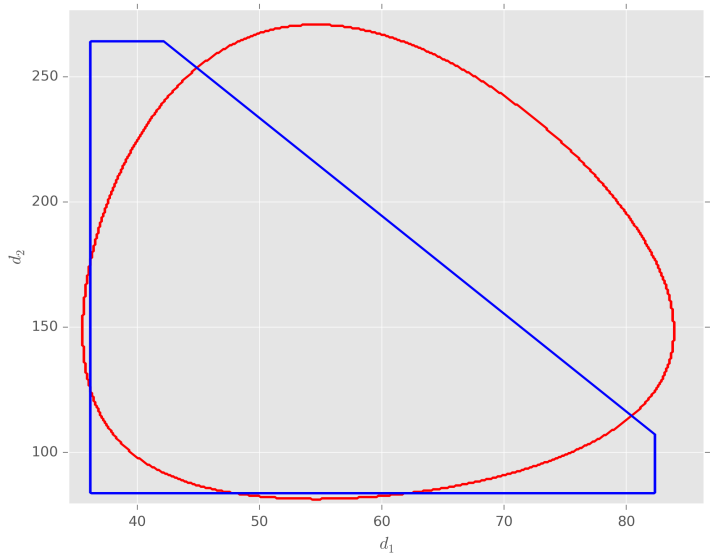


Figure 4: Comparison of the uncertainty sets $\mathcal{D}_r$(red) and $\mathcal{D}_{\texttt{MRORA}}$(blue), for durations $\log d_1 \sim \mathcal{N}(4, 0.25^2)$, $\log d_2 \sim \mathcal{N}(5, 0.35^2)$ at a robustness level $1 - \alpha = 0.9$.

stochastic scheduling on parallel machines, the term "LEPT" denotes the static priority rule in which jobs are started as early as possible, in the order of decreasing expected processing time, see e.g. [50].

- Finally, for the sake of comparison, we also computed other heuristic solutions, which are constructed exactly as above, but the jobs are considered in a different order: In `SV` (shorter variance first) and in `LV` (larger variance first), the jobs are sorted by increasing or decreasing order of variance $(e^{\sigma_j^2} - 1)e^{2\mu_j + \sigma_j^2}$, respectively. Also, to evaluate the benefits of using our statistical prediction model – compared to the situation in which the planner only knows the scheduled duration of the cases, we compute the `LSPT` heuristic (longest scheduled processing time first), in which the jobs are sorted in decreasing order of d_j^{sched}.

Note that the solutions `M-MIP` (or `M-CP`) and `LRS` only differ by the underlying uncertainty set. Figure 4 illustrates $\mathcal{D}_r$ and $\mathcal{D}_{\texttt{MRORA}}$ for an example in two dimensions.

6.3 Evaluation of solution quality

The quality of a solution $(\boldsymbol{x}, \boldsymbol{z})$ can be evaluated by different quality indicators, such as the mean or upper percentiles of the random variable $F(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{d})$. We evaluate these statistics by means of Monte-Carlo simulations with $N_{\text{mc}} = 10^6$ runs. In what follows, we denote by MEAN(`SOL`) the mean of $F(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{d})$, where $(\boldsymbol{x}, \boldsymbol{z})$ is the solution returned by the procedure `SOL` $\in \{\texttt{LRS}, \texttt{MRORA}, \texttt{M-CP}, \texttt{M-MIP}, \texttt{LEPT}, \texttt{LV}, \texttt{SV}, \texttt{LSPT}, \texttt{SAA}\}$. Similarly VaR_α(`SOL`) is the Value-at-risk α of the cost, that is, the $(1 - \alpha)$th percentile of $F(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{d})$.

In order to compare two solutions called SOL and REF, we also introduce the notation

$$\text{MEAN}(\texttt{SOL}|\texttt{REF}) := \frac{\text{MEAN}(\texttt{SOL})}{\text{MEAN}(\texttt{REF})},$$

$$\text{VaR}_\alpha(\texttt{SOL}|\texttt{REF}) := \frac{\text{VaR}_\alpha(\texttt{SOL})}{\text{VaR}_\alpha(\texttt{REF})}.$$

For example, if $\text{VaR}_{0.1}(\texttt{LRS}|\texttt{LEPT}) < 1$ for some instance, then the 90th percentile of $F(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{d})$ is lower for LRS than for the reference solution LEPT. The percentage of opened rooms in a solution is denoted by $\%\text{ROOM}(\texttt{SOL}) = \frac{1}{p}\sum_m z_m$. To evaluate the robustness of a solution $(\boldsymbol{x}, \boldsymbol{z})$, we also denote by $\text{maxF}(\mathcal{D})$ the objective function of Problem (2) for the uncertainty set $\mathcal{D}$. that is, $\max_{\boldsymbol{d}\in\mathcal{D}} F(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{d})$.

6.4 Robustness vs. expected workload

We plot in Figure 5 the ratio of $\text{VaR}_{0.1}(\texttt{LRS})$ to $\text{VaR}_{0.1}(\texttt{LEPT})$, where the LRS solution was computed at the robustness level $1-\alpha = 0.90$ and $1-\alpha = 0.70$. The plot indicates the value of this ratio for the $N = 348$ instances of the set $\mathcal{I}$, plotted against the expected workload $\hat{\rho}$. Since the goal of robust optimization is to protect against extreme scenarios, we expect the 90th percentile of the cost function to be lower for LRS than for LEPT. The plots show this trend indeed, especially for instances where $\hat{\rho}$ is small. There are only minor differences between the plots for the values $\alpha = 0.1$ and $\alpha = 0.3$; for instances with $\hat{\rho} \in [0.8, 0.1]$, it seems that the LRS solution yields better results more often with $\alpha = 0.3$ than with $\alpha = 0.1$. There is no contradiction to the result of Proposition 3.5, because LRS only minimizes an upper bound of VaR_α.

When the expected workload is large ($\hat{\rho} \geq 1.1$), we observe that both LRS and LEPT have a similar quality (with respect to 90th percentiles). One explanation is that for "overfilled" instances, both solutions tend to open all ORs, and balance the overtime between all rooms (because we have $\forall m \in \mathcal{M}$, $c^o_m = 1$). As a result, when $\hat{\rho}$ is large it is likely that all rooms are in overtime, in which case the cost equals $\hat{F}(\boldsymbol{d}) := \sum_m c^f_m + (\sum_j d_j - \sum_m T_m)$. So both $\text{VaR}_{0.1}(\texttt{LRS})$ and $\text{VaR}_{0.1}(\texttt{LEPT})$ is very close to the 90th percentile of $\hat{F}(\boldsymbol{d})$. In other words, it is not possible to be protected against extreme scenarios in instances with large values of $\hat{\rho}$. In the next sections, we will often restrict our attention to instances with a smaller expected workload. In particular, we denote by $\mathcal{I}'$ the subset of the $N' = 261$ instances satisfying $\hat{\rho} \leq 1.1$.

6.5 Effect of the fixed costs of opening c^f

When the opening costs and overtime costs are the same for all ORs ($\forall m \in \mathcal{M}$, $c^f_m = c^f$, $c^o_m = c^o$), we only need to study the effect of the ratio $\frac{c^f}{c^o}$ (because we can rescale the cost function $F(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{d})$, so its dependence in c^o and c^f only occur through this ratio). Hence, we assume $c^o = 1$ and we study the effect of the fixed opening costs c^f.

We plot the percentage of opened ORs (averaged over all instances of the set $\mathcal{I}'$) in Figure 6 (*upper left*), for three different solutions (the robust solutions LRS and M-MIP are

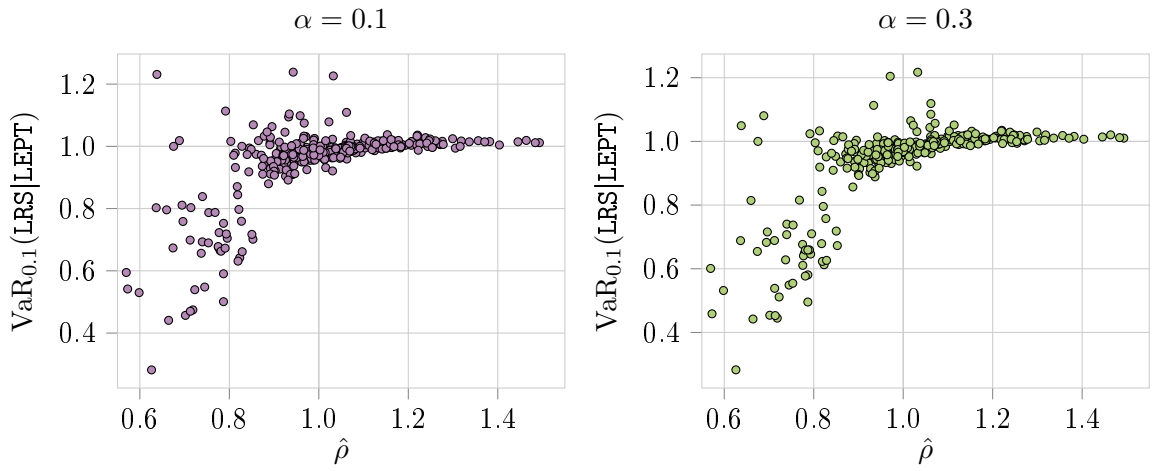


Figure 5: Value of the ratio $\mathrm{VaR}_{0.1}(\texttt{LRS}|\texttt{LEPT})$, where the LRS solution is computed at the robustness level $\alpha = 0.1$ (*left*) and $\alpha = 0.3$ (*right*), for $N = 348$ instances, plotted against the expected workload $\hat{\rho}$ of each instance.

computed for $\alpha = 0.3$). As could be expected, the number of utilized ORs decreases when the fixed cost c_f for opening an OR increases. However, we see that different solution concepts lead to different numbers of ORs. The stochastic programming solution SAA tends to use less rooms than the LRS and M-MIP solutions, which is an indicator for the robustness of LRS and M-MIP. We see on the lower left plot of Figure 6 that SAA is always better than LEPT in terms of means, while LRS only beats LEPT for small values of c_f. The solution LRS protects best against extreme scenarios for small values of c^f (Figure 6, lower right), but it seems that it becomes overconservative as c^f grows, which could explain the fact that LRS opens more rooms than M-MIP for large values of c^f. Finally, the geometric mean of CPU times is plotted in Figure 6 (*upper right*): SAA and LRS are much faster to compute than M-MIP, and the computing times are also less sensitive to the opening costs c^f. We point out that all MIPs are solved using CPLEX 12.6 [15] on a PC with 8 cores at 3.60 GHz.

6.6 Comparison of solution strategies

In this section, we first compare different quality indicators of the solutions, for three groups of instances with particular values of expected workload: the set $\mathcal{I}_1$ contains the $N_1 = 126$ instances such that $\hat{\rho} \leq 0.95$, $\mathcal{I}_2$ contains the $N_2 = 98$ instances such that $(0.95 < \hat{\rho} \leq 1.05)$, and $\mathcal{I}_3$ consists of the $N_3 = 122$ remaining instances. Table 1 indicates, for each group of instances and each solution SOL, the average number of iterations required by the cutting plane procedure (if any), the geometric mean of the CPU-time (in seconds), and the percentage of instances solved within a time limit of 10 minutes. Then, we show the average value of $\mathrm{MEAN}(\texttt{SOL})$, $\mathrm{VaR}_{0.1}(\texttt{SOL})$ and $\mathrm{VaR}_{0.05}(\texttt{SOL})$, and the worst case of the cost $F()$ over the uncertainty sets $\mathcal{D}_r$ and $\mathcal{D}_{\text{MRORA}}$. All robust solutions are computed at the risk level $\alpha = 0.3$.

The lognormal robust schedule LRS is computed for three different values of the tolerance parameter ν of the separation algorithm (Algorithm 2). The value of ν seems to have very

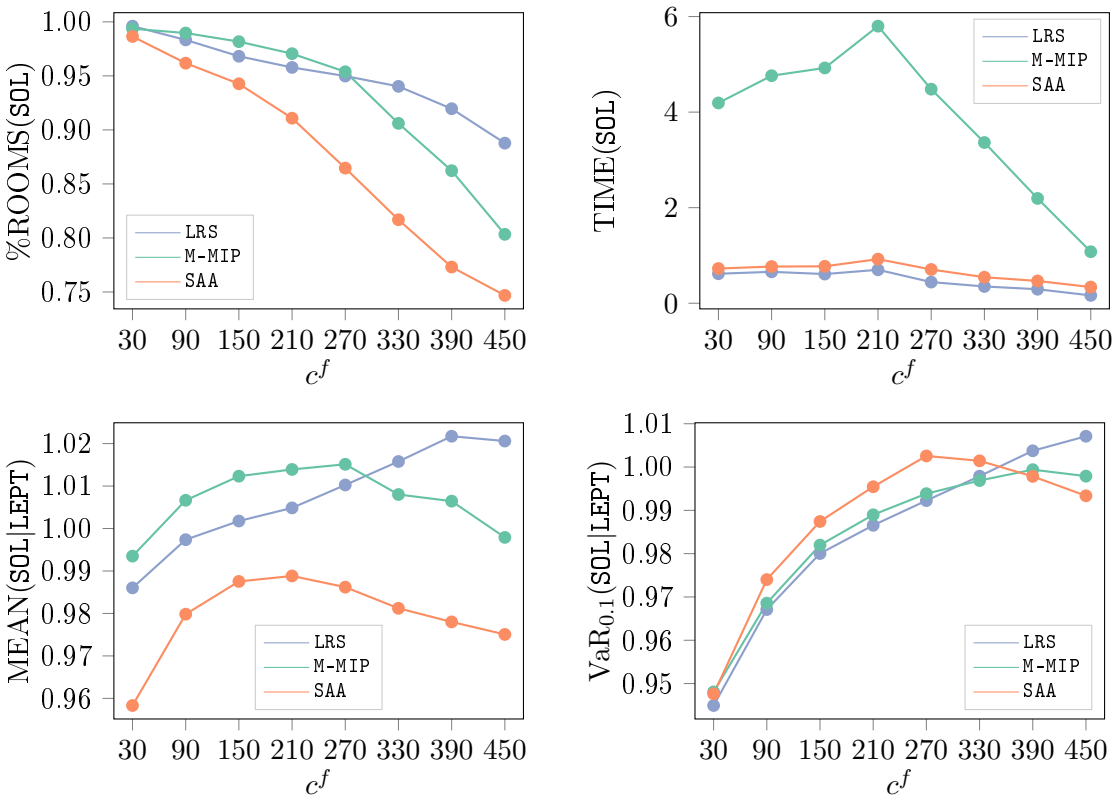


Figure 6: Evolution of the percentage of used ORs (*upper left*), geometric mean of CPU times (*upper right*), average ratio of the mean cost (*lower left*) and of the 90th percentile of the costs (*lower right*) to the reference solution `LEPT`, for three different solutions: `SOL`$\in$ {`LRS`, `M-MIP`, `SAA`}.

little effect on quality of the returned solution, and has only a limited effect on the CPU time, since most of the computing time is spent solving the MIPs, not the separation problems. Besides, for the three values of ν the algorithm returns exactly the same schedules for all instances of $\mathcal{I}_2$ and $\mathcal{I}_3$.

We observe that `LRS` requires a shorter time than `MRORA` for instances with a large expected workload, and a longer time for instances with a small value of $\hat{\rho}$. Instances with a lower expected workload are harder to solve, and require more cutting planes, which affects the computing time. Although some instances require many iterations to reach the desired tolerance ε in Algorithm 1, we already get a good solution in a much shorter time. In practice, setting a limit on the computing time can be used to keep short computations, without impacting too much the quality of the returned solution. To check this fact, we recompute the solution `LRS`, by imposing a time limit of 10 seconds. The optimal solution (with respect to the `LRS`-criterion $\max F(\mathcal{D}_r)$) is found in 343 out of 348 instances. For the other 5 instances, the deviation to the optimal criterion is always less than 1.3% in our experiments.

instances $\mathcal{I}_1(\rho \leq 0.95)$

SOL	#it.	CPU (s)	freq CPU<lim	MEAN	$\text{VaR}_{0.1}$	$\text{VaR}_{0.05}$	$\max\text{F}(\mathcal{D}_r)$	$\max\text{F}(\mathcal{D}_{\text{MRORA}})$
LRS ($\nu = 10^{-6}$)	24.5	2.03	100.0%	220.2	333.2	398.9	296.9	341.6
LRS ($\nu = 10^{-10}$)	24.4	2.33	100.0%	220.2	333.4	399.1	296.9	340.4
LRS ($\nu = 10^{-14}$)	24.7	2.38	100.0%	220.0	332.8	**398.5**	**296.9**	339.1
MRORA	–	0.51	100.0%	217.4	**331.3**	403.8	334.6	306.7
M-MIP	–	3.29	100.0%	217.6	333.0	406.6	338.0	**300.0**
M-CP	76.2	9.53	90.4%	217.7	333.0	407.1	338.6	300.3
LPT	–	–	100.0%	230.9	385.5	469.7	409.0	418.7
SAA	–	0.52	100.0%	**214.3**	338.2	413.3	354.8	355.6

instances $\mathcal{I}_2(0.95 < \rho \leq 1.05)$

SOL	#it.	CPU (s)	freq CPU<lim	MEAN	$\text{VaR}_{0.1}$	$\text{VaR}_{0.05}$	$\max\text{F}(\mathcal{D}_r)$	$\max\text{F}(\mathcal{D}_{\text{MRORA}})$
LRS ($\nu = 10^{-6}$)	7.3	0.31	100.0%	376.4	**611.5**	**718.2**	**605.0**	665.1
LRS ($\nu = 10^{-10}$)	7.3	0.36	100.0%	376.4	**611.5**	**718.2**	**605.0**	665.1
LRS ($\nu = 10^{-14}$)	7.3	0.37	100.0%	376.4	**611.5**	**718.2**	**605.0**	665.1
MRORA	–	1.61	100.0%	386.9	617.8	723.8	617.4	622.6
M-MIP	–	7.14	100.0%	385.8	617.2	723.3	616.8	**616.0**
M-CP	95.2	65.80	77.5%	385.7	618.4	723.4	615.9	617.1
LPT	–	–	100.0%	380.5	625.1	735.8	648.8	689.1
SAA	–	0.94	100.0%	**369.9**	613.5	725.6	635.2	698.9

instances $\mathcal{I}_3(1.05 < \rho)$

SOL	#it.	CPU (s)	freq CPU<lim	MEAN	$\text{VaR}_{0.1}$	$\text{VaR}_{0.05}$	$\max\text{F}(\mathcal{D}_r)$	$\max\text{F}(\mathcal{D}_{\text{MRORA}})$
LRS ($\nu = 10^{-6}$)	2.1	0.06	100.0%	797.9	1158.1	1304.6	**1243.3**	1255.1
LRS ($\nu = 10^{-10}$)	2.1	0.07	100.0%	797.9	1158.1	1304.6	**1243.3**	1255.1
LRS ($\nu = 10^{-14}$)	2.1	0.07	100.0%	797.9	1158.1	1304.6	**1243.3**	1255.1
MRORA	–	1.31	100.0%	810.6	1160.7	1304.2	1244.8	1193.5
M-MIP	–	0.46	100.0%	811.3	1162.3	1305.8	1245.2	**1192.9**
M-CP	10.7	0.16	99.2%	812.0	1162.2	1305.2	1245.2	**1192.9**
LPT	–	–	100.0%	786.9	1149.8	1297.5	1245.1	1223.8
SAA	–	0.87	100.0%	**771.8**	**1147.3**	**1296.7**	1245.0	1249.8

Table 1: Comparison of different quality indicators for three groups of instances, for the robust solutions LRS, MRORA, M-MIP and M-CP at the level $\alpha = 0.3$, as well as the solutions LEPT and SAA. The LRS solution was computed for three different values of the tolerance parameter ν of the lognormal separation algorithm (Algorithm 2).

Interestingly, the cutting plane procedure over the uncertainty set $\mathcal{D}_{\text{MRORA}}$ (i.e., the solution M-CP) requires much more iterations than for the uncertainty set $\mathcal{D}_r$ of the lognormal law. We think that this is due to the combinatorial structure of the worst case scenarios in $\mathcal{D}_{\text{MRORA}}$ (roughly speaking, one extreme scenario for each subset of $\lceil \tau \rceil$ job durations which reach their upper bounds, see Section 3.1). Therefore, the reformulation approach M-MIP is faster than the cutting plane approach M-CP. For M-CP also, we see that instances with a medium expected workload are harder to solve. All but one instances of $\mathcal{I}_3$ can be solved within the time limit of 600 s, while we can solve only 77.5% of the instances of $\mathcal{I}_2$ in the same amount of time.

It is also interesting to look at the last column of the table, to compare the solutions MRORA, M-MIP and M-CP, which all aim at minimizing the quantity $\max\text{F}(\mathcal{D}_{\text{MRORA}})$. The solution M-MIP is always optimum. The value of $\max\text{F}(\mathcal{D}_{\text{MRORA}})$ for M-CP is a bit higher, which is explained

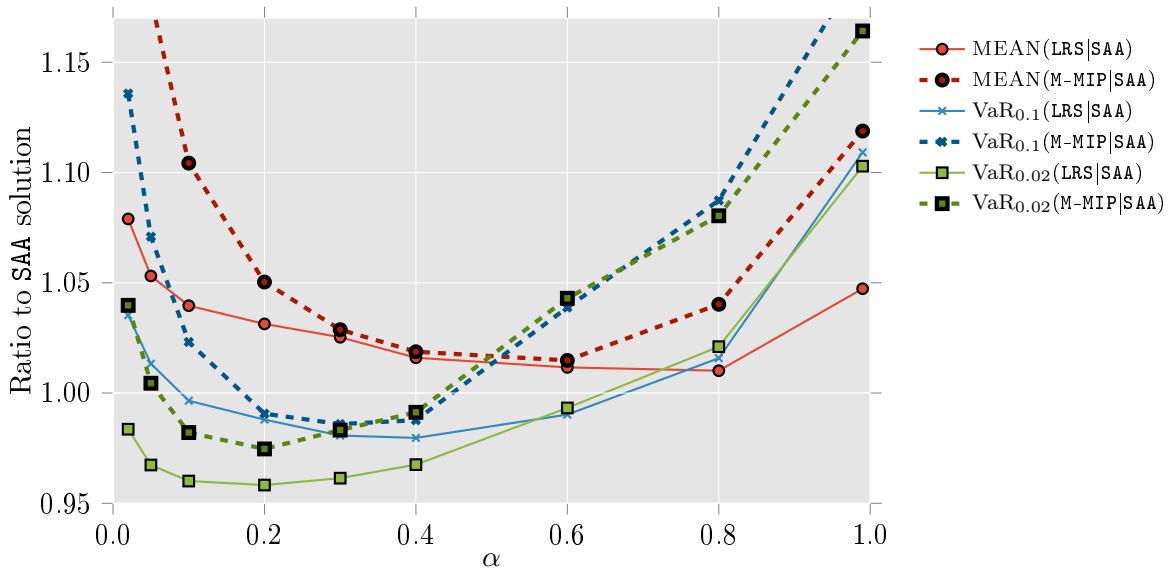
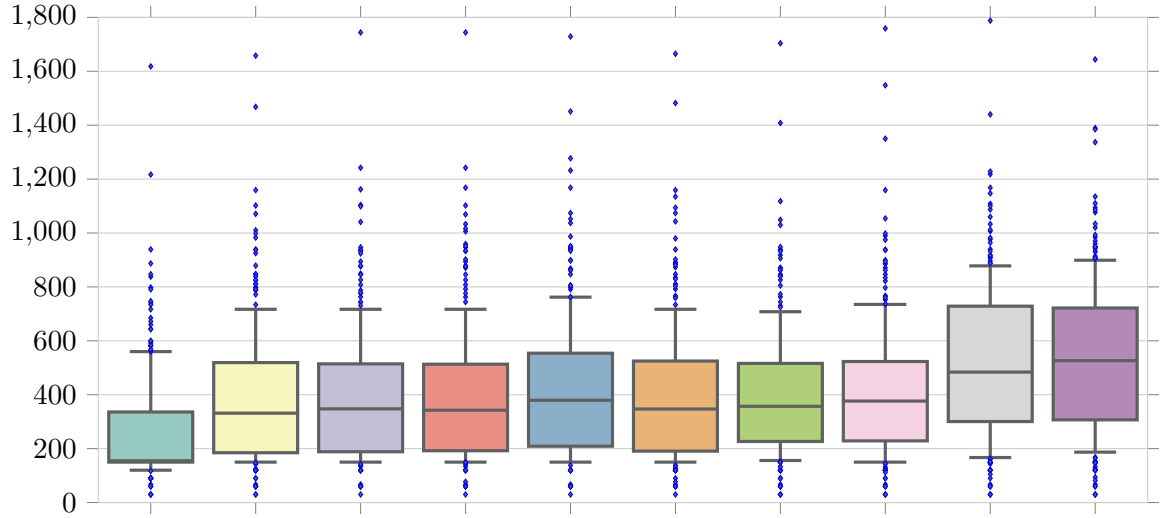


Figure 7: Mean value of the ratios MEAN(SOL|SAA), $\text{VaR}_{0.1}$(SOL|SAA) and $\text{VaR}_{0.02}$(SOL|SAA) over $N' = 261$ instances of $\mathcal{I}'$ (expected workload $\hat{\rho} \leq 1.1$), as a function of the robustness parameter α, for SOL = LRS and SOL = M-MIP.

because for some instances, there is still a small positive optimality gap after the time limit of 600 seconds. The average gap between MRORA and M-MIP is always small, which suggests that the MIP of [19] gives a very good approximation for the robust optimization problem over the uncertainty set $\mathcal{D}_{\text{MRORA}}$. This gap is the largest for the instances with a low expected workload; for $\mathcal{I}_1$, it is of approximately 2.3%.

Then, we study the effect of the robustness parameter α. Figure 7 shows results for the $N' = 261$ instances satisfying $\hat{\rho} \leq 1.1$. The sensibility of LRS and M-MIP to the robustness parameter α is shown on the x-axis (on the left-hand side of the figure, α is small and we protect ourselves against very unlikely scenarios, while on the right-hand side, $\alpha \to 1$ so we basically consider the nominal scenario only). The y-axis shows the mean value (over the 261 instances) of the ratios MEAN(SOL|SAA), $\text{VaR}_{0.1}$(SOL|SAA) and $\text{VaR}_{0.02}$(SOL|SAA) for SOL $\in$ {LRS, M-MIP}.

On this plot, we observe that the LRS solution is better than M-MIP in terms of mean and upper percentiles, for all values of $\alpha \in [0, 1]$. As expected, SAA is always better than both robust solutions (LRS and M-MIP) in terms of mean (the goal of stochastic programming is to minimize the expected value of the cost), but for some values of α, the robust solutions LRS and M-MIP are better in terms of upper percentiles, so they protect against extreme scenarios indeed. Nevertheless this gain seems to be rather marginal: for example, at $\alpha = 0.4$ the LRS solution is, on average, 2.1% better than SAA in terms of 90th percentile, at the price of an increase of 1.6% for the expected cost. As α approaches zero, the robust solutions tend to focus on very unlikely scenarios. This improves on SAA for very high percentiles, but yields a



	ORACLE	SAA	LRS $\alpha = 0.1$	LRS $\alpha = 0.3$	M-MIP $\alpha = 0.1$	M-MIP $\alpha = 0.3$	LEPT	LV	SV	LSPT
MEAN	265.6	**385.6**	391.8	392.9	421.8	398.1	400.2	414.7	520.5	530.7
$\text{VaR}_{0.5}$	155.0	**336.0**	348.0	343.0	380.0	347.0	359.0	377.0	484.0	528.0
$\text{VaR}_{0.25}$	336.0	520.0	516.0	**514.0**	560.0	528.0	517.0	524.0	730.0	724.0
$\text{VaR}_{0.1}$	561.0	734.0	731.0	**717.0**	763.0	734.0	725.0	737.0	887.0	905.0

Figure 8: Distribution of the *true costs* over the instances of $\mathcal{I}'$, for different solutions. The table below the figure indicates the mean, median, third quartile, and 9th decile of the costs, for each solution method.

large increase of the mean cost, especially for `M-MIP`.

6.7 Evaluation of allocation strategies with real observed time

We can evaluate the different allocation strategies by simulating the costs $F(\boldsymbol{x}, \boldsymbol{z}; \boldsymbol{d}^{\text{obs}})$ that would have occurred for the true duration scenario $\boldsymbol{d}^{\text{obs}}$. The distribution of the true costs over the $N' = 261$ instances of $\mathcal{I}'$ is depicted, for several solutions, on the box-plot of Figure 8. The large rectangle goes from the first quartile to the third quartile of each distribution, and the bar in the middle indicates the median. The smaller bars above and below the rectangle are located at the 1st and 9th decile of the distribution of the true costs. The blue dots indicate the costs of the 10% best scenarios and the 10% worst scenarios. The mean, median ($\text{VaR}_{0.5}$), third quartile ($\text{VaR}_{0.25}$), and 9th decile ($\text{VaR}_{0.1}$) of the costs of each solution is displayed under the x-axis. The `LRS` and `M-MIP` solutions are computed at both the risk levels $\alpha = 0.1$ and $\alpha = 0.3$. In addition, we have computed a solution called `ORACLE`, by solving the deterministic allocation problem for the true scenario $\boldsymbol{d}^{\text{obs}}$.

This simulation with real durations shows the same trend as the study of Section 6.6 with synthetic durations generated according to our predicted model: `LRS` gives better results than `M-MIP`; the solutions `SAA` and `LRS` (at both $\alpha = 0.1$ and $\alpha = 0.3$) have a very similar quality;

the average cost of the SAA solution is about 6 min. less than for LRS, but LRS is superior in terms of upper percentiles; despite its simplicity, the heuristic solution LEPT performs remarkably well, it even beats SAA on extreme scenarios.

The cost distribution for LV is just a bit worse as for LEPT. In fact, the variances of the durations are positively correlated with their expected values in our data set, so LV and LEPT often consider the operations in the same order. Our study shows that SV yields bad results and should be avoided. Finally, the comparison between LEPT and LSPT shows the importance of relying on a data-driven statistical model to assign patient blocks to ORs, rather than relying on surgeons' estimates only.

The ORACLE solution would allow a huge improvement compared to all other solutions (approx. 30%). Indeed, in almost half of the instances, there exists an allocation without any overtime. There is little hope to improve the existing solutions by using fixed assignment strategies (such as LRS or SAA), unless much progress is done for the prediction of surgery durations. Instead, we think that research should now focus on reactive allocation strategies, in which the OR allocation may be changed during the day when a case takes longer than expected.

7 Managerial implications

This extensive study allows us to draw a number of practical recommendations for the management of operating rooms.

First, our study gives evidence for the importance of using a good data-driven statistical prediction model for the durations. Compared to the LSPT solution (which uses surgeons' estimates of durations), the LEPT solution relying on our prediction model yields savings of around 25%. Of course, in the real-life savings might not be as large, since OR managers have the possibility to change the allocation during the execution of the schedule in order to react to unexpected events (even though they try to avoid it); however this study is a strong case for the use of accurate prediction models. This should motivate hospital managers to invest in automatic data collection systems that can lead to improved statistical models of surgical durations.

Second, the LEPT solution performs very well, and OR managers can compute this assignment very easily, *by hand*. This solution appears to be particularly robust to extreme scenarios, especially when the expected workload is not too small. Compared to LEPT, managers can expect an average improvement in the order of 2 to 4% by using an optimization algorithm for the assignment of patient blocks to ORs. The SAA solution or the LRS solution can be used, depending on whether the focus is on average performance or protection against extreme scenarios.

Third, on a short-term perspective, when the staffing is fixed and hence the costs of opening c_m^f are small, it is almost always optimal to use all available ORs, especially when the focus is on robustness and stability.

Fourth, there is a need to achieve a better balance of the expected workload across different operating days. Here again, a computer-based prediction model could be used when surgeons book OR-time in the available slots, which would result in a better estimation of the expected

workload. While this may rise an issue due to conflicting interests between surgeons and managers [24], it is in the interest of everyone to distribute more evenly through the year the amount of required OR-time. Another possibility would be to allow for a more flexible allocation of OR slots to different surgery specialties. The *modified block-scheduling* system described in [27] can deal with this situation, by reserving a certain amount of OR-time in the master surgery schedule (MSS) that is not bound to any particular surgeon or surgery specialty. On the day before surgery, the decision maker could then use a prediction model of surgery times to allocate the flexible OR-time slots between the different surgery specialties of the hospital, in order to get the best possible balance of expected workloads.

Fifth, our study shows the limits of proactive, fixed assignment strategies, which have an average cost about 45% higher than the best possible, `ORACLE` ideal solution. From a managerial point of view, this gives evidence for the necessity to allow for flexible allocation schemes, in which a patient block may be re-allocated to another OR when the execution of the schedule does not proceed as expected.

8 Conclusion

This study is motivated by an application to OR management, in which a parallel machines scheduling problem with lognormally distributed durations must be solved. We present a cutting plane approach to solve the robust counterpart of this problem. Our main result allows to efficiently solve the subproblem that generates cut inequalities, when the uncertainty set consists of duration scenarios in a confidence region of the lognormal distribution.

We evaluate our approach on instances based on real data from an application to OR scheduling. Our results show that it is important to use uncertainty sets that rely on the lognormal assumption for robust OR allocation. Compared to the previous model of uncertainty of the `MRORA` approach [19], we obtain solutions that are better both in terms of means and values-at-risk. In terms of computing times, it is also interesting to see that the cutting plane approach requires much less iterations to converge when the underlying uncertainty set is the confidence region $\mathcal{D}_r$ than for the polyhedral uncertainty set $\mathcal{D}_{\texttt{MRORA}}$. The overall computing time depends a lot on the number of required cuts. We think that this could greatly be improved, by using a branch-and-cut strategy with lazy constraints, as described in [8]. Another perspective for future research would be to investigate the development of an FPTAS to solve the restricted master problem, by using a similar approach as in [30] for the minimization of the sum of completion times.

We also observe that the robust optimization approach only works well for instances with a low *expected workload*, that is, instances for which it is likely that the total duration of all cases does not exceed the total time available in all operating rooms. For such instances, we observe that robust optimization is slightly better – in terms of upper percentiles – than a stochastic programming approach based on the sample average approximation (`SAA`), at a small cost in terms of expected value. Nevertheless the gain in terms of robustness is rather small, which shows that the `SAA` approach already provides quite robust solutions.

We draw a number of managerial implications in Section 7. One of them is that the fixed assignment policy `LEPT` performs very well, and can be implemented very easily. In future work

we would like to study the performance bound of this heuristic in the stochastic extensible bin packing model, similarly as was done in [5] for the online counterpart of this problem. At the same time, our study shows that this heuristic is quite sensible to the estimation of the mean job durations, since the policy `LSPT` (which relies on biased estimates) yields much more overtime. An interesting line of research would thus be to study the robustness of the `LEPT` policy, subject to misspecification of the mean durations.

Another perspective is to allow operations performed by the same surgeon to take place in different rooms. This makes the model much more complicated, since we must make sure that no surgeon operates simultaneously in two ORs. As stated in the introduction, this is a resource-constrained scheduling problem, which is already very hard in the deterministic case. However preliminary work shows that our cut generation procedure could be adapted, to find the worst case duration scenario for a given *Earliest-Start*-policy (ES-policy) of the stochastic Resource Constrained Project Scheduling Problem (RCPSP), which are policies that can be represented by a flow of resources through the activities to be scheduled, cf [34]. Then, the cutting-plane algorithm could be used to solve a robust version of the RCPSP over the space of ES-policies; this problem was first studied in [4], but so far only very small instances can be solved.

Finally, we recall that the allocation of ORs to patient blocks is just one of many steps involved in the the management of the operation theater. For future research, it would be necessary to evaluate the performance of the proposed method in a more complex and realistic environment that simulates, e.g., the availability of recovery beds and allocation of anesthesiologists.

Acknowledgement The first draft of this paper was prepared while the first author was affiliated with Charité–Universitätsmedizin Berlin.

The authors wish to thank three anonymous referees for their helpful and constructive comments that greatly contributed to improve both the content and the presentation of this manuscript.

References

References

[1] B. Addis, G. Carello, and E. Tànfani. A robust optimization approach for the operating room planning problem with uncertain surgery duration. In *Proceedings of the international conference on health care systems engineering*, pages 175–189. Springer, 2014.

[2] S. Alimoradi, M. Hematian, and G. Moslehi. Robust scheduling of parallel machines considering total flow time. *Computers & Industrial Engineering*, 93:152–161, 2016.

[3] A. Ardestani-Jaafari and E. Delage. Linearized robust counterparts of two-stage robust optimization problems with applications in operations management. 2016. preprint available at http://www.optimization-online.org/DB_HTML/2016/03/5388.html.

[4] C. Artigues, R. Leus, and F.T. Nobibon. Robust optimization for resource-constrained project scheduling with uncertain activity durations. *Flexible Services and Manufacturing Journal*, 25(1-2):175–205, 2013.

[5] , B. Berg, and B.T. Denton. Fast Approximation Methods for Online Scheduling of Outpatient Procedure Centers. *INFORMS Journal on Computing*, 29(4):631–644, 2017.

[6] S. Batun, B.T. Denton, T.R. Huschka, and A.J. Schaefer. Operating room pooling and parallel surgery processing under uncertainty. *INFORMS journal on Computing*, 23(2):220–237, 2011.

[7] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805, 1998.

[8] D. Bertsimas, I. Dunning, and M. Lubin. Reformulations versus cutting planes for robust optimization. *Computational Management Science*, 13:195––217, April 2016.

[9] D. Bertsimas and M. Sim. The price of robustness. *Operations research*, 52(1):35–53, 2004.

[10] J.R. Birge and F.V. Louveaux. A multicut algorithm for two-stage stochastic linear programs. *European Journal of Operational Research*, 34(3):384–392, 1988.

[11] M. Bougeret, A. Pessoa, and M. Poss. Robust scheduling with budgeted uncertainty, 2016. Preprint available on HAL: <hal-01345283>.

[12] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[13] P.J. Bushell. Hilbert's metric and positive contraction mappings in a banach space. *Archive for Rational Mechanics and Analysis*, 52(4):330–338, 1973.

[14] B.R Cobb, R. Rumi, and A. Salmerón. Approximating the distribution of a sum of log-normal random variables. *Statistics and Computing*, 16, 2012.

[15] IBM ILOG CPLEX. V12. 1 user's manual for cplex. Technical report, International Business Machines Corporation, 2009.

[16] G.B. Dantzig. Discrete-variable extremum problems. *Operations research*, 5(2):266–288, 1957.

[17] P. Dell'Olmo, H. Kellerer, M.G. Speranza, and Z. Tuza. A 1312 approximation algorithm for bin packing with extendable bins. *Information Processing Letters*, 65(5):229–233, 1998.

[18] B. Denton, J. Viapiano, and A. Vogl. Optimization of surgery sequencing and scheduling decisions under uncertainty. *Health care management science*, 10(1):13–24, 2007.

[19] B.T. Denton, A.J. Miller, H.J. Balasubramanian, and T.R.B Huschka. Optimal allocation of surgery blocks to operating rooms under uncertainty. *Operations research*, 58(4-part-1):802–816, 2010.

[20] F. Dexter and R.D. Traub. How to schedule elective surgical cases into specific operating rooms to maximize the efficiency of use of operating room time. *Anesthesia & Analgesia*, 94(4):933–942, 2002.

[21] F. Dexter, R.D. Traub, and P. Lebowitz. Scheduling a delay between different surgeons' cases in the same operating room on the same day using upper prediction bounds for case durations. *Anesthesia & analgesia*, 92(4):943–946, 2001.

[22] M. Diemer, C. Taube, J. Ansorg, J. Heberer, and W. Eiff. *Handbuch OP-Management*. MWV Medizinisch Wissenschaftliche Verlagsgesellschaft, 2015.

[23] L. Fenton. The sum of log-normal probability distributions in scatter transmission systems. *IRE Transactions on Communications Systems*, 8(1):57–67, 1960.

[24] A. Fügener, S. Schiffels, and R. Kolisch. Overutilization and underutilization of operating rooms-insights from behavioral health care operations management. *Health care management science*, 20(1):115–128, 2017.

[25] S. Gaubert and Z. Qu. Dobrushin ergodicity coefficient for markov operators on cones, and beyond. *arXiv preprint arXiv:1302.5226*, 2013.

[26] A. Granas and J. Dugundji. *Fixed point theory*. Springer Science & Business Media, 2013.

[27] F. Guerriero and R. Guido. Operational research in the management of the operating theatre: a survey. *Health care management science*, 14(1):89–114, 2011.

[28] E. Hans, G. Wullink, M. Van Houdenhoven, and G. Kazemier. Robust surgery loading. *European Journal of Operational Research*, 185(3):1038–1050, 2008.

[29] H. Hu, K.K.H. Ng, and Y. Qin. Robust parallel machine scheduling problem with uncertainties and sequence-dependent setup time. *Scientific Programming*, 2016, 2016.

[30] A. Kasperski, A. Kurpisz, and P. Zieliński. Parallel machine scheduling under uncertainty. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 74–83. Springer, 2012.

[31] E. Kayış, T.T. Khaniyev, J. Suermondt, and K. Sylvester. A robust estimation model for surgery durations with temporal, operational, and surgery team effects. *Health care management science*, pages 1–12, 2014.

[32] A.J. Kleywegt, A. Shapiro, and T. Homem-de Mello. The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, 12(2):479–502, 2002.

[33] O. Koné, C. Artigues, P. Lopez, and M. Mongeau. Event-based milp models for resource-constrained project scheduling problems. *Computers & Operations Research*, 38(1):3–13, 2011.

[34] R. Leus. Resource allocation by means of project networks: dominance results. *Networks*, 58(1):50–58, 2011.

[35] C. Mancilla and R.H. Storer. Stochastic sequencing of surgeries for a single surgeon operating in parallel operating rooms. *IIE Transactions on Healthcare Systems Engineering*, 3(2):127–138, 2013.

[36] C. McIntosh, F. Dexter, and R.H. Epstein. The impact of service-specific staffing, case scheduling, turnovers, and first-case starts on anesthesia group and operating room productivity: a tutorial using data from an australian hospital. *Anesthesia & Analgesia*, 103(6):1499–1516, 2006.

[37] F. Meng, J. Qi, M. Zhang, J. Ang, S. Chu, and M. Sim. A robust optimization model for managing elective admission in a public hospital. *Operations Research*, 63(6):1452–1467, 2015.

[38] A. Mutapcic and S. Boyd. Cutting-set methods for robust convex optimization with pessimizing oracles. *Optimization Methods & Software*, 24(3):381–406, 2009.

[39] R.D. Nussbaum. *Iterated nonlinear maps and Hilbert's projective metric, II*, volume 401. American Mathematical Soc., 1989.

[40] R.D. Nussbaum. Finsler structures for the part metric and hilbert's projective metric and applications to ordinary differential equations. *Differential and Integral Equations*, 7(5-6):1649–1707, 1994.

[41] I. Ozkarahan. Allocation of surgical procedures to operating rooms. *Journal of medical systems*, 19(4):333–352, 1995.

[42] D.W. Peterson. A review of constraint qualifications in finite-dimensional spaces. *SIAM Review*, 15(3):639–654, 1973.

[43] D. Pham and A. Klinkert. Surgical case scheduling as a generalized job shop scheduling problem. *European Journal of Operational Research*, 185(3):1011–1025, 2008.

[44] R. Pulido, A.M. Aguirre, M. Ortega-Mier, Á. García-Sánchez, and C.A. Méndez. Managing daily surgery schedules in a teaching hospital: a mixed-integer optimization approach. *BMC health services research*, 14(1):464, 2014.

[45] R.T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 2000.

[46] O.V. Shylo, O.A. Prokopyev, and A.J. Schaefer. Stochastic operating room scheduling for high-volume specialties under block booking. *INFORMS Journal on Computing*, 25(4):682–692, 2012.

[47] P.S. Stepaniak, C. Heij, and G. De Vries. Modeling and prediction of surgical procedure times. *Statistica Neerlandica*, 64(1):1–18, 2010.

[48] P.S. Stepaniak, C. Heij, G.H.H. Mannaerts, M. de Quelerij, and G. de Vries. Modeling procedure and surgical times for current procedural terminology-anesthesia-surgeon combinations and evaluation in terms of case-duration prediction and operating room efficiency: a multicenter study. *Anesthesia & Analgesia*, 109(4):1232–1245, 2009.

[49] J.-S. Tancrez, B. Roland, J.-P. Cordier, and F. Riane. How stochasticity and emergencies disrupt the surgical schedule. In *Intelligent patient management*, pages 221–239. Springer, 2009.

[50] L. Van Der Heyden, Scheduling jobs with exponential processing and arrival times on identical processors so as to minimize the expected makespan. *Mathematics of Operations Research*, 6(2):305–312, 1981.

[51] X. Xu, J. Lin, and W. Cui. Hedge against total flow time uncertainty of the uniform parallel machine scheduling problem with interval data. *International Journal of Production Research*, 52(19):5611–5625, 2014.

Appendix A Proof of Theorem 3.2

We start to give a general result about the Lipschitz constant of a function over $K^{(2)} := \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x} > \boldsymbol{0}, \sum_{i=1}^n x_i^2 = 1\}$ with respect to d_H. The proof of this result relies on the following property of Hilbert's projective metric, see [40]:

$$\forall \boldsymbol{x}, \boldsymbol{y} \in K, \quad d_H(\boldsymbol{x}, \boldsymbol{y}) = \inf_{\varphi} \int_{t=0}^{1} \omega_{\varphi(t)}(\varphi'(t))\, dt, \tag{19}$$

where the infimum is taken over all piecewise $\mathcal{C}^1$-paths φ such that for all $t \in [0,1]$, $\varphi(t) \in K$, $\varphi(0) = \boldsymbol{x}$, $\varphi(1) = \boldsymbol{y}$, and $\omega_{\boldsymbol{u}}(\boldsymbol{h})$ is the oscillation of $\boldsymbol{h}$ and $\boldsymbol{u}$, defined by $\omega_{\boldsymbol{u}}(\boldsymbol{h}) := \max_i(h_i/u_i) - \min_j(h_j/u_j)$. The proof of the theorem mimics that of [40, Theorem 2.4], where a similar result is proved for $K^{(1)} := \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x} > \boldsymbol{0}, \sum_{i=1}^n x_i = 1\}$, but we integrate over a different geodesic curve from $\boldsymbol{x}$ to $\boldsymbol{y}$. A related result is also proved in [25], but for functions f preserving the rays of K.

Theorem A.1. *Let f be a function of class $\mathcal{C}^1$, mapping a geodesically convex set $G \subseteq K^{(2)}$ to the cone $K := \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x} > \boldsymbol{0}\}$. For all $\boldsymbol{x} \in G$, define*

$$\lambda(\boldsymbol{x}) := \sup_{\{\boldsymbol{v}:\, \boldsymbol{v}^T\boldsymbol{x}=0, \boldsymbol{v}\neq\boldsymbol{0}\}} \frac{\omega_{f(\boldsymbol{x})}\big(f'(\boldsymbol{x})(\boldsymbol{v})\big)}{\omega_{\boldsymbol{x}}(\boldsymbol{v})} \in \mathbb{R} \cup \{+\infty\}.$$

Define further $\lambda_0 := \sup\{\lambda(\boldsymbol{x});\ \boldsymbol{x} \in G\}$. Then, we have

$$\forall \boldsymbol{x}, \boldsymbol{y} \in G, \quad d_H(f(\boldsymbol{x}), f(\boldsymbol{y})) \leq \lambda_0\, d_H(\boldsymbol{x}, \boldsymbol{y}).$$

Proof. Observe that $\lambda(\boldsymbol{x})$ is well defined for all $\boldsymbol{x} \in G$. Indeed, $\boldsymbol{v}^T\boldsymbol{x} = 0, \boldsymbol{v} \neq \boldsymbol{0}$ implies that $\boldsymbol{v}$ has at least one positive element, and at least one negative element, so $\omega_{\boldsymbol{x}}(\boldsymbol{v}) > 0$.

Let $\boldsymbol{x}, \boldsymbol{y} \in G$. It is well known that the path $\varphi(t) = (1-t)\boldsymbol{x} + t\boldsymbol{y}$ is a geodesic curve from $\boldsymbol{x}$ to $\boldsymbol{y}$ for Hilbert's projective metric (i.e., φ is a minimizer of expression (19)), see [40, Theorem 2.1]. It follows that for all functions $\alpha : [0,1] \to (0,\infty)$ of class $\mathcal{C}^1$ satisfying $\alpha(0) = \alpha(1) = 1$, the path $\psi(t) := \alpha(t)\varphi(t)$ is also a geodesic. Indeed, for all $t \in [0,1]$,

$$\omega_{\psi(t)}(\psi'(t)) = \max_i \frac{\alpha'(t)}{\alpha(t)} + \frac{\varphi'(t)_i}{\varphi(t)_i} - \min_i \frac{\alpha'(t)}{\alpha(t)} + \frac{\varphi'(t)_i}{\varphi(t)_i} = \omega_{\varphi(t)}(\varphi'(t)).$$

In particular, the path from $\boldsymbol{x}$ to $\boldsymbol{y}$ following the great circle, $\varphi_C(t) := \varphi(t)/\|\varphi(t)\|$ is a geodesic curve from $\boldsymbol{x}$ to $\boldsymbol{y}$ in the Hilbert's projective metric.

We can now use expression (19) with the path $t \mapsto f(\varphi_C(t))$ to obtain a bound of $d_H(f(\boldsymbol{x}), f(\boldsymbol{y}))$:

$$d_H(f(\boldsymbol{x}), f(\boldsymbol{y})) \leq \int_{t=0}^{1} \omega_{f(\varphi_C(t))}\Big(f'(\varphi_C(t))(\varphi_C'(t))\Big)\, dt.$$

The vectors $\varphi_C(t)$ and $\varphi_C'(t)$ are orthogonal for all $t \in [0,1]$, so by definition of λ_0,

$$d_H(f(\boldsymbol{x}), f(\boldsymbol{y})) \leq \int_{t=0}^{1} \lambda_0\, \omega_{\varphi_C(t)}(\varphi_C'(t))\, dt = \lambda_0\, d_H(\boldsymbol{x}, \boldsymbol{y}),$$

where the last expression follows from the fact that φ_C is a geodesic from $\boldsymbol{x}$ to $\boldsymbol{y}$. □

We are now ready to prove Theorem 3.2. By Theorem (A.1), the Lipschitz constant of the restriction of $\boldsymbol{x} \mapsto \exp \boldsymbol{x}$ to $K^{(2)}$ (with respect to d_H) is bounded from above by

$$\lambda_0 = \sup_{\boldsymbol{x} \in K^{(2)}} \sup_{\{\boldsymbol{v} \neq \mathbf{0} : \boldsymbol{v}^T \boldsymbol{x} = 0\}} \frac{\omega_{\exp(\boldsymbol{x})}\big(\operatorname{Diag}(e^{\boldsymbol{x}})\boldsymbol{v}\big)}{\omega_{\boldsymbol{x}}(\boldsymbol{v})} = \sup_{\boldsymbol{x} \in K^{(2)}} \sup_{\{\boldsymbol{v} \neq \mathbf{0} : \boldsymbol{v}^T \boldsymbol{x} = 0\}} \frac{\max_i v_i - \min_i v_i}{\max_i \frac{v_i}{x_i} - \min_i \frac{v_i}{x_i}}.$$

For a fixed vector $\boldsymbol{v}$, we start by minimizing the denominator of the above expression over the set $\{\boldsymbol{x} \in K^{(2)} : \boldsymbol{x}^T \boldsymbol{v} = 0\}$. Let $\mathcal{I}^+, \mathcal{I}^-, \mathcal{I}^0$ be the set of indices $i \in [n]$ such that $v_i > 0$, $v_i < 0$, and $v_i = 0$, respectively. Note that $\boldsymbol{v} \neq \mathbf{0}$ and $\boldsymbol{x}^T \boldsymbol{v} = 0$ for some $\boldsymbol{x} > \mathbf{0}$ implies that $\mathcal{I}^+$ and $\mathcal{I}^-$ are nonempty. The optimization problem with respect to $\boldsymbol{x}$ can be reformulated as

$$\inf_{\{\boldsymbol{x} \in K^{(2)} : \boldsymbol{x}^T \boldsymbol{v} = 0\}} \max_{i \in \mathcal{I}^+} \frac{v_i}{x_i} + \max_{i \in \mathcal{I}^-} \frac{(-v_i)}{x_i}.$$

Now, assume for simplicity that $\mathcal{I}^0 = \emptyset$ (the result for the case $\mathcal{I}^0 \neq \emptyset$ can be obtained by continuity). It is not hard to see that at the optimum, there must exist some constants $\alpha > 0$ and $\beta > 0$ such that $v_i / x_i = \alpha$ for all $i \in \mathcal{I}^+$ and $-v_i / x_i = \beta$ for all $i \in \mathcal{I}^-$. Let $a = (\sum_{i \in \mathcal{I}^+} v_i^2)^{1/2}$ and $b = (\sum_{i \in \mathcal{I}^-} v_i^2)^{1/2}$. The values of α and β are obtained by solving the system of equations

$$\left\{ \frac{a^2}{\alpha^2} + \frac{b^2}{\beta^2} = 1, \quad \frac{a^2}{\alpha} - \frac{b^2}{\beta} = 0 \right\},$$

where the first equation follows from $\|\boldsymbol{x}\| = 1$ and the second one from $\boldsymbol{v}^T \boldsymbol{x} = 0$. We find $\alpha = b/a \|\boldsymbol{v}\|_2$ and $\beta = a/b \|\boldsymbol{v}\|_2$, and so the value of the infimum is $\alpha + \beta = (ab)^{-1} \|\boldsymbol{v}\|_2^3$.

Finally, we consider the maximization problem with respect to $\boldsymbol{v}$ to find the value of λ_0. Observe that we can assume without loss of generality that $\|\boldsymbol{v}\|_2 = 1$, because multiplying $\boldsymbol{v}$ by a constant does not change the value of the ratio to maximize. The numerator is $\max_i v_i - \min_i v_i = \max_{i \in \mathcal{I}^+} v_i + \max_{i \in \mathcal{I}^-} (-v_i) \leq a + b$, where the inequality follows from the inequality between the ℓ_2-norm and the $\ell_\infty-$norm, and a and b satisfy $a^2 + b^2 = \|\boldsymbol{v}\|_2^2 = 1$. We have shown above that the denominator is equal to $(ab)^{-1} \|\boldsymbol{v}\|_2^3 = (ab)^{-1}$. Hence,

$$\lambda_0 \leq \sup\{(a+b)ab; \quad a > 0,\ b > 0,\ a^2 + b^2 = 1\} = \frac{1}{\sqrt{2}}.$$

□