

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 and Emergency Arrivals**

Herausgegeben vom
Konrad-Zuse-Zentrum für Informationstechnik Berlin
Takustraße 7
D-14195 Berlin-Dahlem

Telefon: 030-84185-0
Telefax: 030-84185-125

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^{*†}, Christoph Barner[‡], Ralf Borndörfer^{*}, Mickaël Grima[‡],
Mathees Seeling[‡], Claudia Spies[‡], Klaus Wernecke[†]

March 24, 2016

Keywords: 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. In this article, we formulate this problem as a job shop scheduling problem, in which the job durations follow a lognormal distribution. We propose to use a cutting-plane approach to solve a robust version of this optimization problem. To this end, we develop an algorithm based on fixed-point iterations to solve the subproblems that identify worst-case scenarios and generate cut inequalities. The main result of this article uses Hilbert’s projective geometry to prove the convergence of this procedure under some mild conditions. We also present two extensions of our model, that allow to deal with add-on jobs (emergency arrivals), and to balance the load over several planning period 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) the present robust optimization approach –that protects against likely scenarios from the lognormal durations– perform well compared to a previous model that ignored the distribution of case durations; (ii) the proposed model is able to efficiently reserve capacity for emergency arrivals; (iii) compared to an alternative stochastic programming approach based on the sample average approximation (SAA), robust optimization yields solutions that are a bit more robust against uncertainty, at a small price in terms of average cost; (iv) However the SAA approach is superior to the robust optimization approach for instances in which the total duration of all cases is very likely to exceed to total available time in all ORs.

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

^{*}Zuse Institute Berlin

[†]Charité Universitätsmedizin, Berlin

[‡]Technische Universität München

admissions, and as much as 40% of the total expenses [8]. The management of the OT 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 [12] for a comprehensive review of the operations research literature on OT management.

This paper focuses on the problem of allocating operating rooms to 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 two widespread planning processes, namely the *block-scheduling* and the *open-scheduling* systems. The former system assumes that individual surgeons or surgical teams have predefined slots of OR-time allocated in a periodic schedule, and cases are booked within these slots. In contrast, the latter system fills an empty schedule with cases until the day of operation, by taking into account medical priorities and total waiting times of the patients.

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 [10]. However, a few days before the day of surgery, the staff has already been scheduled, and so the authors of the previous article claim that under-utilized time does not cause a loss of revenue for the surgical suite. This is consistent with [17], where it is shown that on a short-term perspective, the goal is solely to minimize the overtime in the OR. 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 a case could have been scheduled on day j without causing overtime, but is scheduled on day $j + 1$ in overtime. To deal with this issue while keeping a one-day horizon, we will propose an extended version of our model with deferral costs in Section 4.2. Also, in this paper we consider a fixed cost for *opening an OR*, as proposed in the model of [9].

The benefits of robust optimization methods to handle uncertainty have been demonstrated in several articles that we will review in Section 1.1. However, one caveat of these approaches is that no specific assumption is made on the distribution of surgical durations. It is well-known that surgical durations can be well modelled by heavy tailed distributions, in particular there is a close fit with lognormal distributions, see e.g. [27, 14] and the references therein. Therefore, we expect departures from the nominal scenario to be highly nonlinear, a fact poorly captured by existing approaches.

In this paper, we propose to remedy this problem by using a cutting-plane approach to solve a robust optimization problem that protects against all scenarios in a confidence region of the lognormal distribution. The problem we study is a variation of a problem introduced in [9]. We present this problem and the cutting-plane approach in Section 2. Then, we study the adversarial problem in Section 3, that is, the problem that generates new cuts by identifying the worst case scenario for a given allocation. In particular, our main result is stated in Proposition 3.2; we show with the help of *Hilbert's projective metric* that fixed

point iterations can be used to solve very efficiently the adversarial problem, although this is a non-convex optimization problem. Two important extensions of our model are presented in Section 4: the first one allows one to reserve room capacity for emergency cases, and the second one can be used in a rolling horizon approach, in which we also select the patients to be operated next. Finally, we present numerical results for the application to OR management in Section 5.

Throughout this article, we adopt the terminology of the job shop scheduling literature because we believe that the problem studied here could have other fields of application. Hence, **patient blocks are called *jobs*** and **operating rooms are called *machines***.

1.1 Related work

To the best of our knowledge, one of the first paper to consider the problem of allocating operating rooms to a list of surgical procedures is [22], 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 in the present article (patients to be operated by the same surgical team are grouped in a block), for two main reasons:

1. When a surgeon perform 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, a feature desired by many OR planners, in particular at the Charité hospital in Berlin.
2. Mathematical formulations allowing a practitioner to change of OR within a day are much harder to solve, because they involve so-called *big-M* constraints to avoid that two procedures performed by the same surgeon take place at the same time.

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 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 [23, 24]. where MIPs are proposed to solve a resource constrained scheduling problem. However, these approaches only consider the deterministic case, probably because of the very large number of variables and constraints. Moreover, as mentioned above, these MIPs include particular constraints that cause branch-and-bound solvers to make slow progress. One exception is the work of [16], where a stochastic programming approach is proposed to schedule procedures for one surgeon operating in two ORs.

Since the end of the 90's, many papers have demonstrated the benefits of robust optimization (RO) to handle uncertainty. In many cases, it offers tractable mathematical formulations

which are much easier to solve than their stochastic programming counterpart [2, 4]. Moreover, RO offers the possibility to tune the *budget of uncertainty* to choose the trade-off between performance and robustness.

This trend was also observed in the literature on OT management. A non-exhaustive list of recent contributions using robust optimization follows: The authors of [13] allocate slack times in each OR to reduce the risk of overtime; A 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 [25], 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 [18], to select elective admission ratios in order to balance bed occupancy.

The present paper builds on a robust optimization problem introduced by [9]. This paper presented a MIP model (called MRORA) to find an optimal allocation of the ORs, robust against all duration scenarios \mathbf{d} for the patient blocks in the uncertainty set

$$\mathcal{D}_{\text{MRORA}} = \{\mathbf{d} \in \mathbb{R}^n : \forall i, \ell_i \leq d_i \leq u_i; \sum_i \frac{d_i - \ell_i}{u_i - \ell_i} \leq \tau\}, \quad (1)$$

where the parameter τ controls the *budget of uncertainty*. As mentioned earlier, we know that surgical durations have a good fit with the lognormal distribution. Therefore, we think that the budget of uncertainty should be spent *in a logarithmic fashion*, a fact not represented by the set $\mathcal{D}_{\text{MRORA}}$ or the assumptions of the other articles cited above. This motivates our approach to develop a RO model that is more suited for lognormal-like durations.

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 \mathbf{v} . The symbol $\text{Diag}(\mathbf{v})$ 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

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

Denote by T_m the time available on machine m (if it is activated), c_f^m the fixed cost for activating machine m and c_o^m 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 $(\mathbf{x}, \mathbf{z}) \in \mathcal{X}$ can be measured as

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

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

In this paper, we consider the problem of finding the allocation (\mathbf{x}, \mathbf{z}) minimizing the overtime, while protecting ourselves against a set of likely scenarios \mathcal{D} . This leads to the following robust optimization problem:

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

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

$$\min_{\mathbf{x}, \mathbf{z}, \Delta, \delta} \sum_{m \in \mathcal{M}} c_f^m z_m + \Delta \quad (3a)$$

$$\text{s.t.} \quad \delta_{im} \geq \sum_{j \in \mathcal{J}} x_{jm} d_j^{(i)} - z_m T_m, \quad \forall i \in \mathcal{S}, \forall m \in \mathcal{M}, \quad (3b)$$

$$\delta_{im} \geq 0, \quad \forall i \in \mathcal{S}, \forall m \in \mathcal{M}, \quad (3c)$$

$$\Delta \geq \sum_{m \in \mathcal{M}} c_o^m \delta_{im}, \quad \forall i \in \mathcal{S}, \quad (3d)$$

$$(\mathbf{x}, \mathbf{z}) \in \mathcal{X} \quad (3e)$$

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

We also point out that when several machines have the same values for c_f^m, c_o^m and T_m , it is possible to strengthen the above formulation by using symmetry-breaking constraints; see [9]. Also, note that we can drop the integer constraints on the variables z_m , because the continuous variable $z_m \in [0, 1]$ always takes the value 0 or 1 at the optimum.

Next, we introduce the adversarial problem, which, given a solution $(\mathbf{x}^*, \mathbf{z}^*)$ of the restricted master problem (3), finds the worst scenario within the uncertainty set \mathcal{D} ,

$$\max_{\mathbf{d} \in \mathcal{D}} F(\mathbf{x}^*, \mathbf{z}^*; \mathbf{d}). \quad (4)$$

The cutting plane algorithm to solve Problem (2) can be summarized as follows. Start with $\mathcal{D}^{(1)} = \{\bar{\mathbf{d}}\}$, where \bar{d}_i is the expected value of d_i . At iteration $k \in \mathbb{N}$, solve Problem (3) for $\hat{\mathcal{D}} = \mathcal{D}^{(k)}$ and set $(\mathbf{x}^{(k)}, \mathbf{z}^{(k)})$ to the optimal solution. Then, solve Problem (4) with

$(\mathbf{x}^*, \mathbf{z}^*) = (\mathbf{x}^{(k)}, \mathbf{z}^{(k)})$, insert the worst case scenario $\mathbf{d}^{(k)}$ in the restricted uncertainty set, $\mathcal{D}^{(k+1)} = \mathcal{D}^{(k)} \cup \{\mathbf{d}^{(k)}\}$, 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 $\mathbf{x}^{(j)} = \mathbf{x}^{(k)}$ with $j < k$ would indicate that the process has converged at iteration k ; cf. [19]. 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 this process can also be refined by generating worst-case scenarios directly at nodes of the branch-and-bound tree of the MIP (3); see [3] for more details.

Algorithm 1 (ROBUST_CUTS)

Input: Instance defined by: $\mathcal{J}, \mathcal{M}, \mathcal{D} \subseteq \mathbb{R}_+^{|\mathcal{J}|}$,

$\forall m \in \mathcal{M}, c_o^m, c_f^m, T_m \in \mathbb{R}_+$,

nominal scenario $\bar{\mathbf{d}} \in \mathcal{D}$,

tolerance parameter $\varepsilon > 0$;

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

- 1: $L \leftarrow 0$
 - 2: $U \leftarrow +\infty$
 - 3: $\hat{\mathcal{D}} \leftarrow \{\bar{\mathbf{d}}\}$
 - 4: **while** $U > (1 + \varepsilon)L$ **do**
 - 5: $(\mathbf{x}^*, \mathbf{z}^*) \leftarrow$ optimal solution of the restricted master problem (3)
 - 6: $F^* \leftarrow$ optimal value of the restricted master problem (3)
 - 7: $L \leftarrow \max(L, F^*)$
 - 8: $\mathbf{d}^* \leftarrow$ optimal solution of the adversarial problem (4)
 - 9: $U \leftarrow \min(U, F(\mathbf{x}^*, \mathbf{z}^*; \mathbf{d}^*))$
 - 10: $\hat{\mathcal{D}} \leftarrow \hat{\mathcal{D}} \cup \{\mathbf{d}^*\}$
 - 11: **end while**
 - 12: **return** $(\mathbf{x}^*, \mathbf{z}^*)$
-

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

3 Solving the adversarial problem

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 $\mathcal{D} := \{\mathbf{d} \in \mathbb{R}_+^n : \log(\mathbf{d}) \in \mathcal{E}\}$, where $\mathcal{E} := \{\mathbf{y} \in \mathbb{R}^n : \sum_{j=1}^n \sigma_j^{-2}(y_j - \mu_j)^2 \leq r^2\}$ for some $r > 0$. Note that the set \mathcal{D} defined above is simply a log-transformation of some confidence ellipsoid of the multivariate normal law $\mathcal{N}(\boldsymbol{\mu}, \text{Diag}(\boldsymbol{\sigma})^2)$.

Problem (4) may be reformulated as

$$\max_{\epsilon \in \{0,1\}^p} \max_{\mathbf{d} \in \mathcal{D}} \sum_{m \in \mathcal{M}} c_f^m z_m^* + \epsilon_m c_o^m \left(\sum_{j \in \mathcal{J}} x_{jm}^* d_j - T_m \right), \quad (5)$$

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 variable $y_j = \log d_j$. For a fixed $\boldsymbol{\epsilon}$, the value of the inner maximization problem equals

$$\sum_{m \in \mathcal{M}} c_f^m z_m^* - \epsilon_m c_o^m T_m + \max_{\mathbf{y} \in \mathcal{E}} \sum_{j \in \mathcal{J}} u_j e^{y_j}, \quad (6)$$

where we have set $u_j := \sum_{m \in \mathcal{M}} \epsilon_m c_o^m x_{jm}^* \geq 0$. If we put aside the trivial case $\mathbf{u} = \mathbf{0}$, the necessary Karush-Kuhn-Tucker (KKT) conditions for the maximization problem in (6) can be written as follows:

$$\exists \lambda > 0: \begin{cases} \forall j \in \mathcal{J}, \lambda (y_j - \mu_j) \sigma_j^{-1} = \sigma_j u_j e^{y_j} \\ \sum_{j \in \mathcal{J}} (y_j - \mu_j)^2 \sigma_j^{-2} = r^2, \end{cases} \quad (7)$$

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

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

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

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

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

The next results give a condition –almost always verified in practice, cf. discussion in Section 3.2– which guarantees that fixed point iterations of g converge, and we can use the fixed point to find a global optimum of (6). To do this, we prove the following result, which relies on *Hilbert's projective metric* d_H on the cone $K := \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} > \mathbf{0}\}$. It is defined by $\forall \mathbf{x}, \mathbf{y} \in K$, $d_H(\mathbf{x}, \mathbf{y}) := \log \max_i \frac{x_i}{y_i} + \log \max_j \frac{y_j}{x_j}$. 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 $K^{(1)} := \{\mathbf{x} \in K : \|\mathbf{x}\|_1 = 1\}$ or $K^{(2)} := \{\mathbf{x} \in K : \|\mathbf{x}\|_2 = 1\}$.

We list hereafter a few important properties of d_H , which are proved e.g. in [20]:

- (i) $\forall \mathbf{x}, \mathbf{y} \in K$, $d_H(\mathbf{x}, \mathbf{y}) = 0$ implies $\mathbf{x} = \alpha \mathbf{y}$ for some $\alpha > 0$
- (ii) $\forall \mathbf{x}, \mathbf{y} \in K$, $d_H(\mathbf{x}, \mathbf{y}) = d_H(\mathbf{y}, \mathbf{x})$
- (iii) $\forall \mathbf{x}, \mathbf{y} \in K$, $\forall \lambda > 0$, $d_H(\mathbf{x}, \lambda \mathbf{y}) = d_H(\mathbf{x}, \mathbf{y})$
- (iv) $\forall \mathbf{x}, \mathbf{y} \in K$, $\forall \mathbf{u} \in K$, $d_H(\mathbf{u} \circ \mathbf{x}, \mathbf{u} \circ \mathbf{y}) = d_H(\mathbf{x}, \mathbf{y})$
- (v) $(K^{(2)}, d_H)$ is a complete metric space.

Algorithm 2 (LOGNORMAL_ADVERSE)

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

Output: ν -approximate solution $\mathbf{d} \in \mathcal{D}$ of Problem (4).

```
1:  $OPT \leftarrow 0$ 
2:  $\mathbf{d}^* \leftarrow \mathbf{0}_n$ 
3: for all  $\epsilon \in \{0, 1\}^p, \epsilon \neq \mathbf{0}_p$  do
4:    $u_j \leftarrow \sum_{m \in \mathcal{M}} \epsilon_m c_o^m x_{jm}^*$  ( $\forall j \in \mathcal{J}$ )
5:    $\mathbf{w}^{(0)} \leftarrow \mathbf{0}_n$ 
6:    $\Delta \leftarrow +\infty$ 
7:    $i \leftarrow 0$ 
8:   while  $\Delta > \nu$  do
9:      $\mathbf{f} \leftarrow \boldsymbol{\sigma} \circ \mathbf{u} \circ \exp(\boldsymbol{\mu} + r\boldsymbol{\sigma} \circ \mathbf{w}^{(i)})$ 
10:     $\mathbf{w}^{(i+1)} \leftarrow \frac{\mathbf{f}}{\|\mathbf{f}\|_2}$ 
11:     $\Delta \leftarrow \|\mathbf{w}^{(i+1)} - \mathbf{w}^{(i)}\|_2$ 
12:     $i \leftarrow i + 1$ 
13:   end while
14:    $\mathbf{d} \leftarrow \exp(\boldsymbol{\mu} + r\boldsymbol{\sigma} \circ \mathbf{w}^{(i)})$ 
15:    $OPT_\epsilon \leftarrow \sum_{m \in \mathcal{M}} c_f^m z_m^* + \epsilon_m c_o^m (\sum_{j \in \mathcal{J}} x_{jm}^* d_j - T_m)$ 
16:   if  $OPT_\epsilon > OPT$  then
17:      $OPT \leftarrow OPT_\epsilon$ 
18:      $\mathbf{d}^* \leftarrow \mathbf{d}$ 
19:   end if
20: end for
21: return  $\mathbf{d}^*$ 
```

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

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

The proof of this result is included in the appendix of this article. 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 convergence of the fixed point iterations.

Proposition 3.2. *Assume that for all $j \in \mathcal{J}, r\sigma_j < \sqrt{2}$. Then, there exists a point $\mathbf{w}^* \in K^{(2)}$ such that the fixed point iterations $g(g(\dots g(\mathbf{w}_0)))$ converge to \mathbf{w}^* for all $\mathbf{w}_0 \in \mathbb{R}^n$. Moreover, $\mathbf{y}^* := \boldsymbol{\mu} + r \text{Diag}(\boldsymbol{\sigma})\mathbf{w}^*$ is a global optimum of Problem (6).*

Proof. If $u_j = 0$ for some $j \in \mathcal{J}$, then it is clear that the fixed-point iterates $\mathbf{w}^{(k)} = g^k(\mathbf{w}_0)$ will satisfy $w_j^{(k)} = 0$ for all $k \geq 1$. So we assume without loss of generality that $u_j > 0$ 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, and any fixed point must lie in $K^{(2)}$. By using the properties of Hilbert's projective metric, we find that

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

In the above expression, the first equality follows from Properties (ii) and (iii), and the second equality follows from (iv). Therefore, Theorem 3.1 implies that g is contractant for the 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 fixed point theorem ensures the unicity of a fixed point \mathbf{w}^* and the convergence of fixed point iterations when $r\|\boldsymbol{\sigma}\|_\infty < \sqrt{2}$. In this case, $\mathbf{y}^* := \boldsymbol{\mu} + r\boldsymbol{\sigma} \circ \mathbf{w}^*$ is the unique solution of the necessary conditions (7), so \mathbf{y}^* maximizes $\sum_j u_j e^{y_j}$ over \mathcal{E} . \square

Our approach is summarized in Algorithm 2, which can be used to solve the adversarial problem at line 8 of Algorithm 1; The condition of Proposition 3.2 ensures that this procedure converges.

3.1 Choice of the parameter r

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

$$P_n(r) := \Phi(r)^n - \left(\Phi(r) - \frac{1}{2}\right)^n + \frac{1}{2^n} \sqrt{\chi_n^2(r)}, \quad (8)$$

where Φ is the standard normal cumulative distribution function (CDF), and χ_n^2 is the CDF of the χ^2 -distribution with n degrees of freedom. Indeed, the probability that \mathbf{d} lies in $\bar{\mathcal{D}}$ is the same as the probability that the vector with coordinates $\sigma_j^{-1}(\log d_j - \mu_j)$ —which has a standard multivariate normal distribution—lies in the grey area depicted in Figure 1. The formula of $P_n(r)$ expresses this probability as the difference between the probability $P_1 = \Phi(r)^n$ of the quadrant below the point A , and the probability P_2 of the region denoted with stripes in the figure. The probability of the striped region is equal to the difference between the probability of the cube with corners A and O , and the part of the ball of radius r that is situated in the positive quadrant: $P_2 = \left(\Phi(r) - \frac{1}{2}\right)^n - \frac{1}{2^n} \sqrt{\chi_n^2(r)}$.

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 $F(\mathbf{x}, \mathbf{z}; \mathbf{d})$:

Proposition 3.3. *Assume that the d_j 's follow independent lognormal distributions, with $\log d_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$. Let $(\mathbf{x}^*, \mathbf{z}^*)$ be the optimal solution of the robust job shop problem (2)*

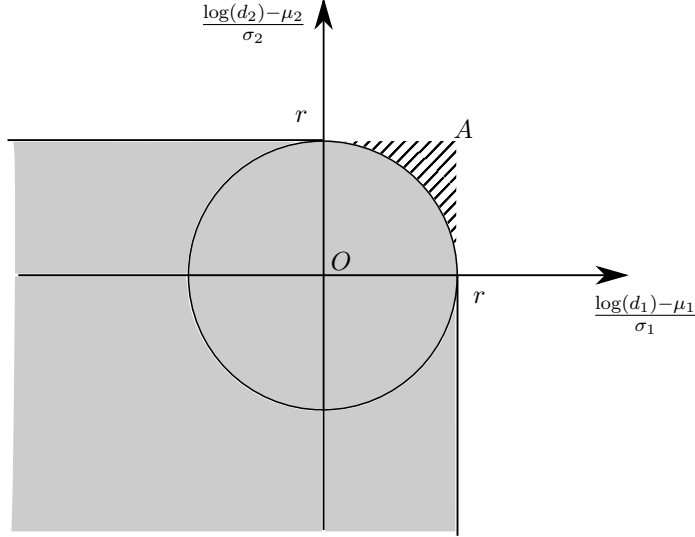


Figure 1: The probability $P_n(r)$ that $\mathbf{d} \in \bar{\mathcal{D}}$ is equal to the probability that a random vector $X \sim \mathcal{N}(0, \mathbf{I}) \in \mathbb{R}^n$ lies in the grey area. The formula (8) expresses this probability as the difference between the probability of the quadrant below A and the striped region.

and let $F^* := \max_{\mathbf{d} \in \mathcal{D}} F(\mathbf{x}^*, \mathbf{z}^*; \mathbf{d})$ be its optimal value, where

$$\mathcal{D} := \left\{ \mathbf{d} \in \mathbb{R}_+^n : \sum_{j=1}^n \sigma_j^{-2} (\log d_j - \mu_j)^2 \leq r^2 \right\} \quad \text{and} \quad P_n(r) = 1 - \alpha.$$

Then,

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

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

3.2 Discussion on the assumptions of Proposition 3.2

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 $[0.37m_j, 2.67m_j]$, 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 Extensions

4.1 Add-on cases (emergencies)

In this section, we propose an extension of our model to reserve capacity for additional jobs that must be inserted in the schedule during its execution. This is motivated by the application to surgery scheduling described in the introduction: many hospitals do not have an OR dedicated to handle emergency operations. Instead, emergency cases are allocated to one of the regular ORs after one of the ongoing operations has ended. This approach was recommended in a simulation study for a large teaching hospital [28].

We point out that the proposed approach does not represent the whole complexity of reserving capacity for emergency cases. In particular, our model assumes that capacity for the k th emergency arrival is reserved in room r_k . In life threatening situations however, it might not be possible to wait for the end of the ongoing operation in r_k , and the k th emergency might have to be handled in some other OR. However, our simulation study of Section 5.7 shows that we can operate the vast majority of add-on patients in room that was foreseen, which justifies the present approach.

Let N_A be a random variable indicating the number of additional jobs that arrive during the execution of the schedule. We assume that N_A has a known distribution over a finite support $\{0, \dots, N\}$:

$$\sum_{k=0}^N p_k = 1, \quad \text{where } p_k := \mathbb{P}(N_A = k).$$

Let $\mathcal{K} := \{1, \dots, N\}$. We introduce a set of decision variables $x_{km} \in \{0, 1\}$ ($\forall (k, m) \in \mathcal{K} \times \mathcal{M}$), such that $x_{km} = 1$ indicates whether the k th add-on job will be scheduled on machine m (in the event that $N_A \geq k$). Similarly, for all $k \in \mathcal{K}_0 := \{0\} \cup \mathcal{K}$ the variable $z_{km} \in \{0, 1\}$ indicates whether machine m is activated if $N_A \geq k$. The set of feasible solutions is thus as follows:

$$\hat{\mathcal{X}} = \left\{ (\mathbf{x}, \mathbf{z}) \in \{0, 1\}^{|\mathcal{J} \cup \mathcal{K}| \times |\mathcal{M}|} \times \{0, 1\}^{|\mathcal{K}_0| \times |\mathcal{M}|} : \begin{array}{l} \forall j \in \mathcal{J} \cup \mathcal{K}, \quad \sum_{m \in \mathcal{M}} x_{jm} = 1; \\ \forall j, m \in \mathcal{J} \times \mathcal{M}, \quad x_{jm} \leq z_{0m}; \\ \forall k, m \in \mathcal{K} \times \mathcal{M}, \quad x_{km} \leq z_{km}; \\ \forall k, m \in \mathcal{K} \times \mathcal{M}, \quad z_{(k-1)m} \leq z_{km} \end{array} \right\}.$$

Let δ_k denote the duration of the k th additional job. Our simplified model assumes that all add-on jobs can be inserted directly after an existing job, so if $N_A = k$ the cost function becomes

$$F_k(\mathbf{x}, \mathbf{z}; \mathbf{d}, \boldsymbol{\delta}) := \sum_{m \in \mathcal{M}} c_f^m z_{km} + c_o^m \left(\sum_{j \in \mathcal{J}} x_{jm} d_j + \sum_{i=1}^k x_{im} \delta_i - T_m \right)^+,$$

where by convention the sum indexed by i is equal to zero whenever $k = 0$. Searching for a solution (\mathbf{x}, \mathbf{z}) that protects against a set of likely scenarios $(\mathbf{d}, \boldsymbol{\delta})$ might lead to overconservatism. Indeed, we expect such a solution to focus on scenarios with a large number of

additional jobs, although this might occur with a very low probability. Therefore, we propose a hybrid approach between robust optimization and stochastic programming, in which we minimize the expected worst case costs, conditionally to the number of add-on jobs:

$$\min_{(\mathbf{x}, \mathbf{z}) \in \hat{\mathcal{X}}} F(\mathbf{x}, \mathbf{z}; \mathbf{d}, \boldsymbol{\delta}) := \sum_{k=0}^N p_k \max_{(\mathbf{d}_k, \boldsymbol{\delta}_k) \in \mathcal{D}_k} F_k(\mathbf{x}, \mathbf{z}; \mathbf{d}, \boldsymbol{\delta}),$$

where $\mathcal{D}_k \subset \mathbb{R}^{n+k}$ is a set of likely scenarios with exactly k additional jobs.

It is straightforward to adapt the cutting plane approach of Algorithm 1 to solve the above optimization problem. In place of the adversarial problem (4), there are now $(N + 1)$ independent problems to solve at each iteration: for all $k \in \mathcal{K}_0$,

$$\max_{(\mathbf{d}_k, \boldsymbol{\delta}_k) \in \mathcal{D}_k} F_k(\mathbf{x}^*, \mathbf{z}^*; \mathbf{d}_k, \boldsymbol{\delta}_k). \quad (9)$$

The optimal solutions of these $(N + 1)$ problems are to be inserted in a restricted set of scenarios $\hat{\mathcal{D}}$. When the finite set $\hat{\mathcal{D}}$ is given, the restricted master problem

$$\min_{(\mathbf{x}, \mathbf{z}) \in \hat{\mathcal{X}}} \sum_{k=0}^N p_k \max_{(\mathbf{d}, \boldsymbol{\delta}) \in \hat{\mathcal{D}} \cap \mathbb{R}^{n+k}} F_k(\mathbf{x}, \mathbf{z}; \mathbf{d}, \boldsymbol{\delta})$$

$$\min_{\mathbf{x}, \mathbf{z}, \Delta, \mathbf{o}} \sum_{k=0}^N p_k \left(\sum_{m \in \mathcal{M}} c_f^m z_{km} + \Delta_k \right) \quad (10a)$$

$$\text{s.t.} \quad o_{im} \geq \sum_{j \in \mathcal{J}} x_{jm} d_j^{(i)} + \sum_{k=1}^{k(i)} x_{km} \delta_k^{(i)} - z_{k(i)m} T_m, \quad \forall i \in \mathcal{S}, \forall m \in \mathcal{M}, \quad (10b)$$

$$o_{im} \geq 0, \quad \forall i \in \mathcal{S}, \forall m \in \mathcal{M}, \quad (10c)$$

$$\Delta_{k(i)} \geq \sum_{m \in \mathcal{M}} c_o^m o_{im}, \quad \forall i \in \mathcal{S}, \quad (10d)$$

$$(\mathbf{x}, \mathbf{z}) \in \hat{\mathcal{X}}, \quad (10e)$$

where \mathcal{S} is the finite set of scenario indices, and for all $i \in \mathcal{S}$ the symbol $k(i)$ denotes the number of add-on jobs in scenario $(\mathbf{d}^{(i)}, \boldsymbol{\delta}^{(i)}) \in \hat{\mathcal{D}}$ (so we have $\boldsymbol{\delta}^{(i)} \in \mathbb{R}^{k(i)}$). In summary, at each iteration the cutting plane approach to handle add-on jobs solves the $(N + 1)$ adversarial problems (9). The extreme scenarios thus identified are added in $\hat{\mathcal{D}}$, which translates into adding cut inequalities of the form (10b)-(10d) in the restricted master problem (10).

For the application to allocation of operating rooms, we assume that all add-on cases have the same distribution: $\forall k \in \mathcal{K}$, $\log \delta_k \sim \mathcal{N}(\mu_0, \sigma_0^2)$ (and as before, $\log d_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$ for all $j \in \mathcal{J}$). Therefore, we shall consider uncertainty sets of the form

$$\mathcal{D}_k := \left\{ (\mathbf{d}, \boldsymbol{\delta}) \in \mathbb{R}^{n+k} : \sum_{j=1}^n \left(\frac{\log d_j - \mu_j}{\sigma_j} \right)^2 + \sum_{i=1}^k \left(\frac{\log \delta_i - \mu_0}{\sigma_0} \right)^2 \leq r_k^2 \right\},$$

for all $k \in \mathcal{K}_0$. Note that \mathcal{D}_k has the same form as the uncertainty set \mathcal{D} considered in Section 3, so we can solve the adversarial problem (9) with Algorithm 2.

Finally, we point out that the guidelines of Section 3.1 can be used to set the parameters r_0, \dots, r_N : given a robustness parameter α , solve the equations $1 - \alpha = P_{n+k}(r_k)$ (for all $k \in \{0, \dots, N\}$).

4.2 Job selection and cancellation

Our model can also easily be adapted to the situation in which the decider may decide to cancel some jobs, or on the contrary when he may accept more jobs than initially planned. This naturally occurs in settings where the job shop scheduling 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. This is the case, e.g., in the application to OR management that motivated this study.

This situation 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 for non-selected jobs. So the objective function becomes:

$$\min_{\mathbf{x}, \mathbf{z}, \Delta, \delta} \quad \sum_{m \in \mathcal{M}} c_f^m z_m + \Delta + \sum_{j \in \mathcal{J}} \lambda_j (1 - \sum_{m \in \mathcal{M}} x_{jm}). \quad (3a')$$

Finally, note that these modifications do not affect the adversarial 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.

5 Results for the application to allocation of ORs

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 5.2, and we define some performance metrics in Section 5.3; The influence of the *expected filling rate* of the instance is discussed in Section 5.4; The robust optimization approach of the present paper is compared to the previous robust optimization model of [9] in Section 5.5; In order to evaluate the price of robustness, we also provide a comparison to a stochastic programming approach in Section 5.6. Finally, we present a simulation study with emergency cases in Section 4.1.

5.1 Data and instances

We used maximum likelihood estimators to fit the parameters of a lognormal model for the durations of $N = 20.849$ 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 [26], 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 10-fold cross validation.

We constructed instances of the OR allocation problem for several regular working days (weekends and holidays excepted), by grouping all patients operated by the same surgeon in a patient block. We used the method of moments to fit a lognormal distribution for the total duration of each patient block, $\log d_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$. 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. We denote by $\bar{d}_j = \exp(\mu_j + \frac{1}{2}\sigma_j^2)$ the expected duration of patient block j , and by \hat{d}_j its true (observed) duration. We further define the *expected filling rate* of an instance as

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

which is a simple measure of how likely overtime will occur. Indeed, whenever $\rho \geq 1$ the expected time required to perform all the operations is longer than the regular opening time of the ORs.

For all our instances we used $c_f^m = 30$ and $c_o^m = 1$ ($\forall m \in \mathcal{M}$). In other words, opening a room yields the same cost as 30 minutes of overtime.

It is important to point out that both elective and emergency surgery is performed at the Charité hospital in Berlin. In Sections 5.2 to 5.6 we focus on the basic problem of allocating ORs to elective cases only, so in our instances we treated all emergency cases as elective ones, that is, we did as if these cases could have been planned on the day prior to operation. In Section 5.7 however, emergency cases are removed from the instance, and we use the approach of Section 4.1 to reserve capacity for add-on cases.

5.2 Solution methods

We shall next present detailed results for $N = 348$ instances of the OR allocation problem. For each instance, we compare the quality of three scheduling strategies listed below; the first two solution methods also depend on a parameter α specifying the level of robustness.

- The **LRS** (lognormal robust schedule) is the approach proposed in the present paper. It was obtained by using Algorithm 1 with a tolerance parameter of $\varepsilon = 0.01$; the adversarial subproblems are solved by Algorithm 2, with a tolerance parameter $\nu = 10^{-6}$. This procedure approximates the solution of the robust optimization problem (2) for an uncertainty set of the form $\mathcal{D}_{\text{LRS}} = \{\mathbf{d} \in \mathbb{R}_+^n : \sum_{j=1}^n \sigma_j^{-2} (\log d_j - \mu_j)^2 \leq r^2\}$. The parameter r was set by solving the equation $P_n(r) = 1 - \alpha$, cf. (8).

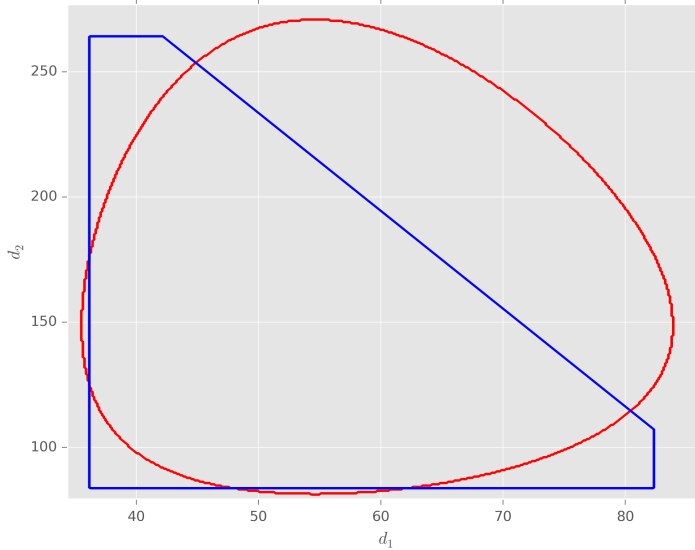


Figure 2: Comparison of the uncertainty sets \mathcal{D}_{LRS} (red) and $\mathcal{D}_{\text{MRORA}}$ (blue), for durations $\log d_1 \sim \mathcal{N}(4, 0.25^2)$, $\log d_2 \sim \mathcal{N}(5, 0.35^2)$ at a robustness level $1 - \alpha = 0.9$.

- We have implemented the robust MIP called **MRORA** in [9]. This MIP solves Problem (2) for the uncertainty set $\mathcal{D}_{\text{MRORA}}$ defined in Equation (1). We followed the rule suggested in [9] to set the value of the parameters ℓ_j , u_j and τ defining $\mathcal{D}_{\text{MRORA}}$, 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 [9, Section 6.1].
- As a reference, we also used the solution provided by the longest processing time (LPT) heuristic for the nominal scenario $\bar{\mathbf{d}}$. It basically consists in sorting all cases by decreasing order of (expected) duration. Then, each case is allocated to one OR in a greedy fashion, by choosing the OR that causes the least increase in overtime; see [9] for a more detailed description. This solution is known to give excellent results when the goal is to minimize the expected value of $F(\mathbf{x}, \mathbf{z}; \mathbf{d})$ [9], and has an approximation guarantee of $\frac{13}{12}$ in the deterministic case [7].
- Finally, to study the tradeoff between stochastic programming and robust optimization, we have computed a solution by using the sample average approximation to stochastic programming (SAA), cf. Section 5.6.

Note that the solutions **MRORA** and **LRS** only differ by the underlying uncertainty set. Figure 2 illustrates \mathcal{D}_{LRS} and $\mathcal{D}_{\text{MRORA}}$ for an example in two dimensions.

5.3 Evaluation of solution quality

The quality of a solution (\mathbf{x}, \mathbf{z}) is evaluated by different statistics of the random variable $F(\mathbf{x}, \mathbf{z}; \mathbf{d})$, such as its mean or upper percentiles. We evaluated these statistics by means of Monte-Carlo simulations with $N = 10^6$ runs. In what follows, we denote by $\text{MEAN}(\text{SOL})$ the mean of $F(\mathbf{x}, \mathbf{z}; \mathbf{d})$, where (\mathbf{x}, \mathbf{z}) is the solution returned by the procedure $\text{SOL} \in \{\text{LRS}, \text{MRORA}, \text{LPT}, \text{SAA}\}$. Similarly $P_q(\text{SOL})$ is the q th percentile of $F(\mathbf{x}, \mathbf{z}; \mathbf{d})$. In order to compare two solutions called SOL and REF , we also introduce the notation

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

$$P_q(\text{SOL}|\text{REF}) := \frac{P_q(\text{SOL})}{P_q(\text{REF})}.$$

For example if $P_{90}(\text{LRS}|\text{LPT}) < 1$ for some instance, then the 90th percentile of $F(\mathbf{x}, \mathbf{z}; \mathbf{d})$ is lower for LRS than for the reference solution LPT.

5.4 Robustness vs. expected filling rate

We have plotted in Figure 3 the ratio of $P_{90}(\text{LRS})$ to $P_{90}(\text{LPT})$, where the LRS solution was computed at the robustness level $1 - \alpha = 0.90$. The plot indicates the value of this ratio for the $N = 348$ instances, plotted against the expected filling rate ρ . 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 LPT. The plot shows this trend indeed, for instances where ρ is small. When the expected filling rate is large however ($\rho \geq 1.1$), we observe that the LRS and LPT 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 ρ is large it is likely that all rooms are in overtime, in which case the cost equals $\hat{F}(\mathbf{d}) := \sum_m c_f^m + (\sum_j d_j - \sum_m T_m)$. So both $P_{90}(\text{LRS})$ and $P_{90}(\text{LPT})$ are very close to the 90th percentile of $\hat{F}(\mathbf{d})$. In other words, it is not possible to be protected against extreme scenarios in instances with large values of ρ . In the next two sections, we will therefore restrict our attention to instances satisfying $\rho \leq 1.1$.

5.5 Case study for 30 instances solved with LRS and MRORA

Table 1 presents results for the 30 instances of March and April 2013 with $\rho \leq 1.1$. Besides the size of the instance (n patient blocks and p ORs), we display the number of iterations that Algorithm 1 required to reach the tolerance ε . The next columns indicate the ratios $\text{MEAN}(\text{SOL}|\text{LPT})$, $P_{90}(\text{SOL}|\text{LPT})$ and $P_{98}(\text{SOL}|\text{LPT})$ for $\text{SOL} \in \{\text{LRS}, \text{MRORA}\}$ at a robustness level $1 - \alpha = 0.9$.

The table evidences that LRS is more robust than MRORA. On average, $P_{90}(\text{LRS})$ is 6.4% lower than $P_{90}(\text{LPT})$, while $P_{90}(\text{MRORA})$ is only 1.2% lower than for the reference solution. Moreover, the expected value of LRS is almost always better than the expected value of

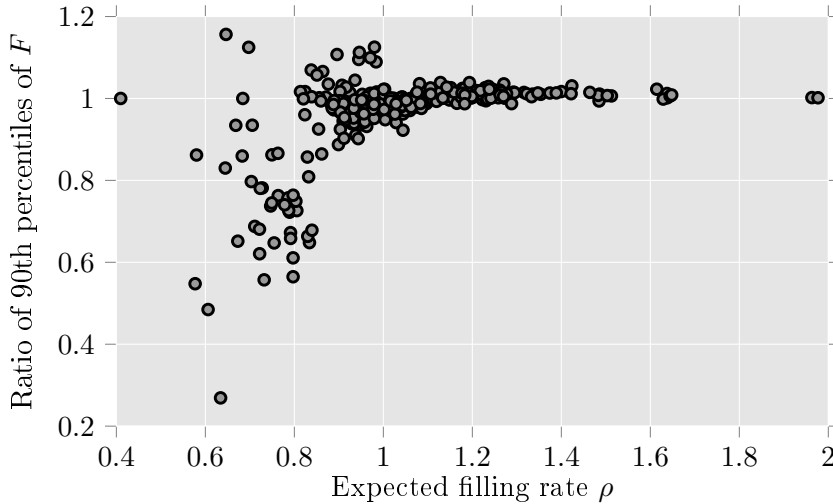


Figure 3: Value of the ratio $P_{90}(\text{LRS}|\text{LPT})$ for $N = 348$ instances, plotted against the expected filling rate ρ of each instance.

MRORA, and even significantly for certain instances. On average, LRS does a bit better than the reference solution (99.1%), while MRORA yields an increase of the expected cost of 8.9%.

Concerning the computing time, we point out that all MIPs were solved using CPLEX 12.6 [6] on a PC with 8 cores at 3.60 GHz. We observe that LRS requires a time similar to MRORA for most instances. However, some instances require many iterations to reach the desired tolerance, which affects the computing time. In practice, setting a limit on the number of iterations could help to keep the computing time within a few seconds, without impacting too much the quality of the returned solution.

5.6 Price of Robustness

It is also natural to compare the proposed robust optimization approach to stochastic programming, in which the expected value of the cost is minimized:

$$\min_{(\mathbf{x}, \mathbf{z}) \in \mathcal{X}} \mathbb{E}_{\mathbf{d}}[F(\mathbf{x}, \mathbf{z}; \mathbf{d})] \quad (11)$$

The authors of [9] have implemented the multicut L-shaped algorithm of Birge and Louveaux [5] to solve Problem (11). However, in our situation generating optimality cuts requires to evaluate $\mathbb{E}_{\mathbf{d}}[F(\mathbf{x}^*, \mathbf{z}^*; \mathbf{d})]$ and its subgradient, which is a costly process when durations d_j are lognormally distributed. To avoid this computational burden, we use the *sample average approximation* (SAA) method to approximate Problem (11). This approximation method has already been used in the context of OR management. This is the case, e.g. in [16], where the authors use it to set the starting time of surgical cases in a single OR. Theoretical convergence results for the SAA method were studied in [15], and are illustrated for a resource allocation problem presenting some similarities with Problem (11).

Ins.	n	p	#it.	CPU(s)		MEAN(SOL LPT)		$P_{90}(\text{SOL LPT})$		$P_{98}(\text{SOL LPT})$	
				LRS	MRORA	LRS	MRORA	LRS	MRORA	LRS	MRORA
I1	9	5	13	0.85	0.32	0.951	0.894	0.805	0.698	0.761	0.732
I2	10	5	20	5.31	0.38	0.989	1.169	0.987	1.098	0.985	1.049
I3	11	5	15	1.72	0.35	0.951	1.007	0.937	0.983	0.948	0.977
I4	11	5	2	0.07	0.38	0.981	1.360	0.973	1.148	0.975	1.050
I5	12	5	38	14.78	3.14	0.951	1.006	0.935	0.967	0.943	0.952
I6	10	4	2	0.06	0.14	0.978	1.110	0.975	1.060	0.978	1.026
I7	11	5	2	0.08	0.49	0.969	1.420	0.966	1.176	0.966	1.043
I8	12	5	15	5.67	0.51	1.027	0.999	0.993	0.968	0.940	0.937
I9	11	5	2	0.06	1.10	0.968	1.142	0.967	1.047	0.980	1.026
I10	12	5	39	7.01	5.34	1.030	1.034	0.717	0.732	0.733	0.750
I11	12	5	34	3.59	4.11	1.002	0.937	0.838	0.723	0.704	0.693
I12	11	5	2	0.07	1.08	0.973	1.003	0.976	0.998	0.981	0.997
I13	12	5	42	61.52	0.47	0.979	0.999	0.976	0.994	0.980	0.994
I14	9	5	3	0.12	0.15	0.996	1.929	0.995	1.675	0.994	1.531
I15	12	5	7	0.31	1.18	1.032	1.061	1.025	1.049	0.995	1.023
I16	11	5	2	0.07	0.43	0.968	1.006	0.976	0.999	0.984	0.998
I17	12	5	12	1.58	0.31	1.032	1.132	1.005	1.064	0.983	1.008
I18	11	5	2	0.05	0.44	0.969	1.046	0.987	1.024	0.996	1.010
I19	9	5	6	1.65	0.28	1.025	1.076	0.973	1.013	0.962	0.981
I20	12	5	41	8.80	3.02	0.754	0.740	0.655	0.629	0.665	0.646
I21	11	5	2	0.08	0.26	0.985	1.002	0.993	0.995	0.990	0.994
I22	12	5	27	8.13	0.86	0.969	1.087	0.940	1.022	0.931	0.977
I23	12	5	2	0.05	0.48	1.005	1.006	1.003	1.009	1.002	1.008
I24	12	5	2	0.11	0.45	0.970	1.066	0.973	1.026	0.978	1.010
I25	11	5	9	0.54	0.30	0.997	1.016	0.961	0.992	0.940	0.968
I26	9	5	2	0.06	0.22	0.977	1.114	0.993	1.064	0.998	1.041
I27	11	5	3	0.12	2.26	0.978	0.961	0.961	0.929	0.961	0.933
I28	12	5	2	0.04	0.25	0.975	1.145	0.986	1.057	0.989	1.019
I29	10	5	5	0.25	0.21	1.109	1.125	0.859	0.859	0.556	0.685
I30	9	4	22	0.77	0.14	1.055	1.061	0.652	0.652	0.511	0.547
mean	-	-	13.56	4.11	0.96	0.991	1.089	0.936	0.988	0.911	0.954
std. dev.	-	-	14.21	11.21	1.27	0.055	0.197	0.097	0.187	0.135	0.174

Table 1: Comparison of the expected value (*resp.* 90th, 98th percentile) of $F(\mathbf{x}^*, \mathbf{z}^*; \mathbf{d})$, where $(\mathbf{x}^*, \mathbf{z}^*)$ is the LRS or MRORA solution at robustness level $1 - \alpha = 0.90$, measured as a ratio to the expected value (*resp.* 90th, 98th percentile) for the reference solution (LPT).

We sample N_S duration scenarios $\mathbf{d}^{(1)}, \dots, \mathbf{d}^{(N_S)}$ and solve the following MIP, which is a small variation of (3), except that we minimize the average overtime instead of the worst case overtime.

$$\min_{\mathbf{x}, \mathbf{z}, \Delta, \delta} \quad \sum_{m \in \mathcal{M}} c_f^m z_m + \frac{1}{N_S} \sum_{i=1}^{N_S} \Delta_i \quad (12a)$$

$$\text{s.t.} \quad \delta_{im} \geq \sum_{j \in \mathcal{J}} x_{jm} d_j^{(i)} - z_m T_m, \quad \forall i \in \{1, \dots, N_S\}, \forall m \in \mathcal{M}, \quad (12b)$$

$$\delta_{im} \geq 0, \quad \forall i \in \{1, \dots, N_S\}, \forall m \in \mathcal{M}, \quad (12c)$$

$$\Delta_i \geq \sum_{m \in \mathcal{M}} c_o^m \delta_{im}, \quad \forall i \in \{1, \dots, N_S\}, \quad (12d)$$

$$(\mathbf{x}, \mathbf{z}) \in \mathcal{X} \quad (12e)$$

For each instance, we have computed a SAA solution for $N = 100$ samples of duration

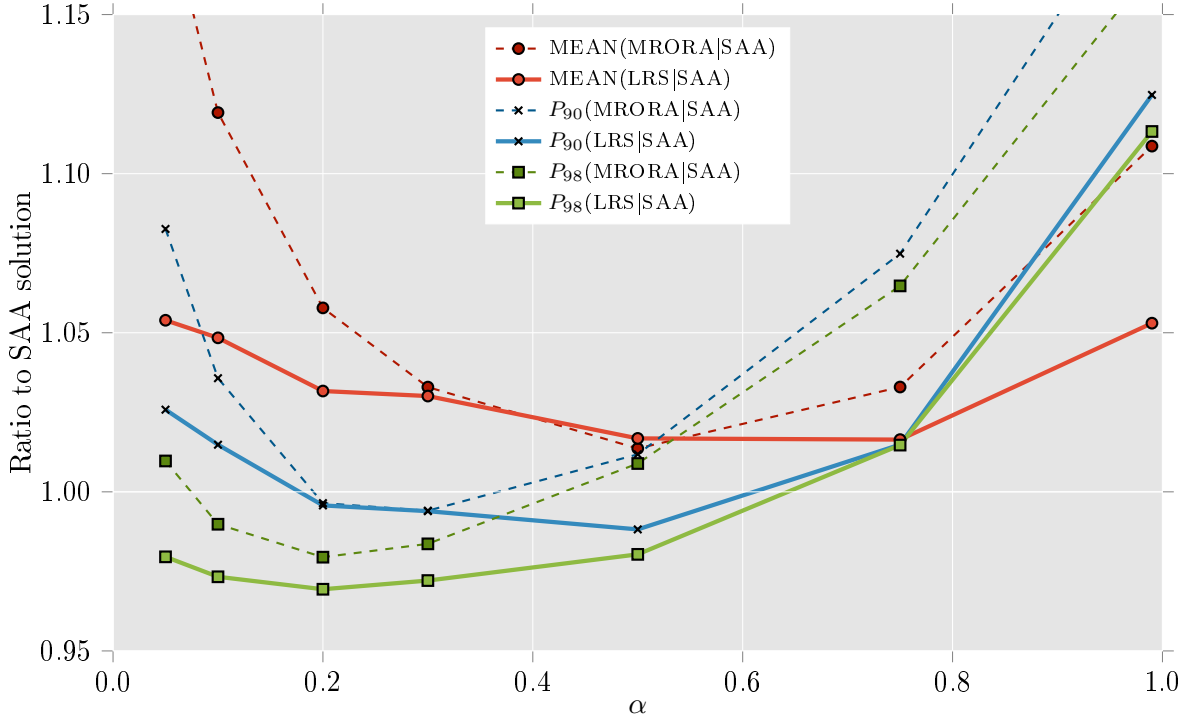


Figure 4: Mean value of the ratios $\text{MEAN}(\text{SOL}|\text{SAA})$, $P_{90}(\text{SOL}|\text{SAA})$ and $P_{98}(\text{SOL}|\text{SAA})$ over $N = 241$ instances with an expected filling rate $\rho \leq 1.1$, as a function of the robustness parameter α , for $\text{SOL} = \text{LRS}$ and $\text{SOL} = \text{MRORA}$.

scenarios. Figure 4 shows results for the 241 instances satisfying $\rho \leq 1.1$. The sensibility of LRS and MRORA 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, $\alpha \rightarrow 1$ so we basically consider the nominal scenario only). The y -axis shows the mean value (over the 241 instances) of the ratios $\text{MEAN}(\text{SOL}|\text{SAA})$, $P_{90}(\text{SOL}|\text{SAA})$ and $P_{98}(\text{SOL}|\text{SAA})$ for $\text{SOL} \in \{\text{LRS}, \text{MRORA}\}$.

On this plot, we observe that the LRS solution is better than MRORA in terms of upper percentiles, for all values of $\alpha \in [0, 1]$, and it is also better in terms of mean for most values of α . As expected, SAA is always better than both robust solutions (LRS and MRORA) 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 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 $\alpha = 0.5$ the LRS solution is, on average, 1.2% 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 large increase of the mean cost, especially for MRORA.

Priority level	Time window for operation
N0	Immediate intervention required, operation in next available OR
N1	Operation within 1h
N2	Operation within 6h
N3	Operation within 12h
N4	Operation within 24h
N5	Elective operation

Table 2: Priority scale for emergency patients used at the Charité hospital.

5.7 Simulation study with emergency arrivals

In this section, we evaluate the approach of Section 4.1 to handle add-on cases. For each of the instances considered above, we constructed a new modified instance, by removing the cases that were labelled as emergencies in our database. We will compare several solutions that were computed prior to the day of intervention, without knowing the number of emergency patients or their check-in times. To evaluate these solutions, we use a simulation procedure that uses simple rules to insert the add-on cases into the schedule after their arrival.

Empirical frequencies based on historical data led us to consider the following probabilities of arrival in the department of general surgery: $p_0 = 0.32$, $p_1 = 0.33$, $p_2 = 0.23$, $p_3 = 0.08$, $p_4 = 0.04$. In fact there is a small nonzero probability to observe $N_A > 4$ arrivals, but we neglected this fact and normalized the above probabilities so $\sum_{i=0}^4 p_i = 1$. There is an average of $\bar{N}_A = 1.19$ emergency cases per day. We can thus redefine the *expected filling rate* as

$$\rho = \frac{\sum_{j \in \mathcal{J}} \bar{d}_j + \bar{N}_A \bar{d}_0}{\sum_{m \in \mathcal{M}} T_m},$$

where $\bar{d}_0 = \exp(\mu_0 + \frac{1}{2}\sigma_0^2)$ is the expected duration of an emergency case.

For each instance, we computed:

- the lognormal robust schedule (LRS), which allocates ORs to elective cases only;
- the lognormal robust schedule with capacity for add-on cases (LRS_A0) described in Section 4.1 at robustness level $1 - \alpha = 70\%$
- a sample average approximation for a stochastic program that reserves capacity for add-on cases (SAA_A0) based on a slight variant of Problem (12). This procedure is detailed in Appendix B.

In addition, we used as a baseline (BAS) the true OR allocation that was used on the day of operation for elective patient blocks.

When an emergency occurs, the priority of the operation is evaluated based on the severity of the patient’s condition. Our database indicates both the time of arrival of the patients and their priority, on a scale ranging from N0 (vital emergency) to N5 (elective case), each level being associated with a permitted time window for the operation; see Table 2.

To evaluate the quality of the different solutions, we used a simplified simulation procedure which ignores the availability of other resources (such as anesthesiologists or recovery beds). First, the elective patient blocks are inserted one after the other in the OR where they were allocated. Then, for $k = 1, \dots, N_A$, the k th emergency patient is inserted in the schedule. Note that it is possible to know whether an OR will be free at time t somewhat ahead of t , because the room first needs to be cleaned up (recall that the clean-up time is included in the duration d_j of all patient blocks). Our simulation simply assumes that we know at time $t - t_0$ whether an insertion is possible in some OR at time t , where t_0 was set to 30 minutes. Denote by τ_k the check-in time of the k th emergency patient, and by ℓ_k the length of the permitted time window for this patient (as from Table 2, and ℓ_k was set to 5 minutes for patients of priority N0). To insert the k th emergency patient in the schedule, we first determine its *preferred* OR r_k :

- (i) For LRS and SAA, we store a queue Q of ORs with foreseen capacity for emergency arrivals, as specified by the variables x_{km} ($k \in \mathcal{K}, m \in \mathcal{M}$). Whenever Q is not empty, the preferred room r_k is popped from Q ; otherwise we use the rule (ii).
- (ii) For LPT and BAS, or when Q is empty, we evaluate the expected increase in cost associated with an insertion in room r , given that the durations of all operations performed until τ_k are known. Then, r_k is set to the OR with the least expected cost.

If r_k becomes free before $\tau_k + \ell_k$, then the k th emergency patient is inserted after the end of the current operation in room r_k . Otherwise, we operate the patient in the first OR that becomes available after time $\max(\tau_k, \tau_k + \ell_k - t_0)$, i.e., after the moment where we know that the patient cannot be operated in OR r_k , but not earlier than its arrival.

Table 3 shows some performance indicators for the solutions LRS, LRS_A0, SAA_A0, computed for the N=334 instances of this study for which the baseline solution BAS was available. The results are split for three subgroups of instances, corresponding to expected filling rates $\rho < 0.95$, $0.95 \leq \rho < 1.05$, and $\rho \geq 1.05$. These three groups respectively contain 162, 91 and 81 instances. The last two rows of the table show the fraction of all add-on cases that were allocated to the *preferred* OR r_k (according to the simulation procedure detailed above), as well as the fraction of emergency cases operated *in time* (before $\tau_k + \ell_k$). Observe that in the vast majority of the cases, it is possible to insert the emergency cases in the foreseen room. This justifies our approach that reserves capacity for eventual emergency arrivals with the help of variables x_{km} .

The table also indicates the mean cost for each subgroup of instances, as well as the frequency at which the overtime exceeds a certain threshold. Since all instances do not have the same number n of available rooms (which varies from 4 to 6), we standardize the overtime by number of rooms. More precisely, for an instance with n available ORs and a solution that yields a cost F , the standardized overtime is defined by

$$OVT_{st} = \frac{1}{n} \left(F - \sum_{m \in \mathcal{M}} c_m^f \right).$$

Note that this formula actually subtracts the fixed cost c_m^f from the overtime for each room m that is not used. In 97% of all instances however, all available ORs have been allocated to at least one surgery. In this situation, a value of $OVT_{st} = 60$ means that the average overtime is of one hour over the n ORs.

Concerning the mean cost, the table clearly shows that **LRS_AO** and **SAA_AO** beat the other two solutions for instances with a low or medium expected filling rate. For instances with a low expected filling rate ($\rho < 0.95$), **LRS_AO** and **SAA_AO** exhibit a similar mean cost; compared to the baseline solution **BAS**, the cost is reduced by a value corresponding to approximately 27 minutes per day. For instances with a large expected filling rate however, **SAA_AO** is the clear winner. Note also that **LRS_AO** is always better than **LRS**, which shows the importance of reserving capacity for emergency arrivals.

In terms of robustness against extreme scenarios, we observe the same trend: **LRS_AO** and **SAA_AO** perform better than **LRS** and **BAS** for small and medium values of ρ , and **SAA_AO** is the clear winner for larger values of ρ . For instances with a low expected filling rate ($\rho < 0.95$), **LRS_AO** seems to be a bit more robust than **SAA_AO**, while both solutions have the same mean cost. Indeed, the number of instances with an average overtime per room of 90 minutes or more (105, 120) is reduced by 18% (30%, 50%).

It is also noteworthy that the baseline solution can be regarded as a reactive strategy (i.e., **BAS** includes elements of real-time decision making because the allocation may have changed during the day of operation, after the planners observed that some cases took longer than expected), while **LRS_AO** and **SAA_AO** have been computed before the day of operation. Nevertheless, **LRS_AO** and **SAA_AO** beat the baseline, which shows the importance of using mathematical optimization for OR management.

6 Conclusion

This study was motivated by an application to OR management, in which a job shop scheduling problem with lognormally distributed durations must be solved. We presented a cutting plane approach to solve the robust counterpart of this problem. The main result of this article allows one 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 evaluated our approach on instances based on real data from the application to OR scheduling. Our results show that it is important to use uncertainty sets that rely on the log-normal assumption for robust OR allocation. Compared to the previous model of uncertainty of the **MRORA** approach [9], we obtained solutions that are better both in terms of expected value and robustness.

We also observed that the robust optimization approach only works well for instances with a low *expected filling rate*, 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 observed 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

	Exp. filling rate	LRS_A0	SAA_A0	LRS	BAS
Mean cost	$\rho < 0.95$	346.0	346.3	365.4	373.2
	$0.95 \leq \rho < 1.05$	414.6	403.6	437.4	453.0
	$1.05 \leq \rho$	548.6	507.9	554.0	574.8
Frequency $OVT_{st} \geq 60$	$\rho < 0.95$	22.8%	22.2%	27.8%	25.9%
	$0.95 \leq \rho < 1.05$	31.9%	29.7%	36.3%	34.1%
	$1.05 \leq \rho$	58.0%	50.6%	59.3%	65.4%
Frequency $OVT_{st} \geq 75$	$\rho < 0.95$	14.8%	17.9%	19.8%	15.4%
	$0.95 \leq \rho < 1.05$	19.8%	20.9%	26.4%	28.6%
	$1.05 \leq \rho$	40.7%	38.3%	45.7%	53.1%
Frequency $OVT_{st} \geq 90$	$\rho < 0.95$	8.6%	10.5%	16.0%	11.1%
	$0.95 \leq \rho < 1.05$	12.1%	13.2%	14.3%	20.9%
	$1.05 \leq \rho$	30.9%	23.5%	32.1%	39.5%
Frequency $OVT_{st} \geq 105$	$\rho < 0.95$	4.3%	6.2%	11.1%	8.0%
	$0.95 \leq \rho < 1.05$	8.8%	7.7%	9.9%	12.1%
	$1.05 \leq \rho$	18.5%	12.3%	23.5%	23.5%
Frequency $OVT_{st} \geq 120$	$\rho < 0.95$	1.9%	3.7%	6.8%	4.9%
	$0.95 \leq \rho < 1.05$	5.5%	4.4%	7.7%	9.9%
	$1.05 \leq \rho$	13.6%	6.2%	13.6%	12.3%
Fraction of emergency cases allocated to r_k		92.4%	92.6%	94.6%	93.9%
Fraction of emergency operations started before $\tau_k + \ell_k$		99.0%	99.0%	98.7%	99.2%

Table 3: Performance of 4 solution methods for the $N = 334$ instances of the simulation study with emergency arrivals, computed with respect to the true (observed) durations \hat{d}_j , grouped by expected filling rate.

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.

Our approach can also be extended to deal with emergency cases, thanks to additional variables that allow one to reserve capacity for add-on patients. A simulation study relying on real surgical durations and emergency arrivals evidenced similar conclusions as in the situation with elective cases only: the robust optimization approach gives very good solutions for instances with a low expected filling rate. For other instances, we rather recommend to use a solution based on a SAA approach.

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.

Another direction of research 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, that is already very hard in the deterministic case. Nevertheless, we think that the solution proposed in the present article could serve as a good starting solution for variable neighborhood search metaheuristics, but this will require

new methods to quickly evaluate the delay propagation in activity graphs with lognormal durations; we intend to pursue this line of research in future work.

References

- [1] B. Addis, G. Carello, and E. Tãnfani. 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] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805, 1998.
- [3] D. Bertsimas, I. Dunning, and M. Lubin. Reformulations versus cutting planes for robust optimization. *Optimization Online*, 2014.
- [4] D. Bertsimas and M. Sim. The price of robustness. *Operations research*, 52(1):35–53, 2004.
- [5] 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.
- [6] IBM ILOG CPLEX. V12. 1 user’s manual for cplex. Technical report, International Business Machines Corporation, 2009.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] S. Gaubert and Z. Qu. Dobrushin ergodicity coefficient for markov operators on cones, and beyond. *arXiv preprint arXiv:1302.5226*, 2013.
- [12] 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.
- [13] E. Hans, G. Wullink, M. Van Houdenhoven, and G. Kazemier. Robust surgery loading. *European Journal of Operational Research*, 185(3):1038–1050, 2008.
- [14] E. Kayış, T.T. Khanliyev, 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.
- [15] 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.
- [16] 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.

- [17] 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.
- [18] 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.
- [19] A. Mutapcic and S. Boyd. Cutting-set methods for robust convex optimization with pessimizing oracles. *Optimization Methods & Software*, 24(3):381–406, 2009.
- [20] R.D. Nussbaum. *Iterated nonlinear maps and Hilbert’s projective metric, II*, volume 401. American Mathematical Soc., 1989.
- [21] 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.
- [22] I. Ozkarahan. Allocation of surgical procedures to operating rooms. *Journal of medical systems*, 19(4):333–352, 1995.
- [23] 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.
- [24] 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.
- [25] 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.
- [26] P.S. Stepaniak, C. Heij, and G. De Vries. Modeling and prediction of surgical procedure times. *Statistica Neerlandica*, 64(1):1–18, 2010.
- [27] 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.
- [28] G. Wullink, M. Van Houdenhoven, E.W. Hans, J.M. van Oostrum, M. van der Lans, and G. Kazemier. Closing emergency operating rooms improves efficiency. *Journal of Medical Systems*, 31(6):543–546, 2007.

A Proof of Theorem 3.1

We start to give a general result about the Lipschitz constant of a function over $K^{(2)} := \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} > \mathbf{0}, \sum_{i=1}^n x_i^2 = 1\}$ with respect to d_H . The proof of this result relies on the following property of Hilbert's projective metric, see [21]:

$$\forall \mathbf{x}, \mathbf{y} \in K, \quad d_H(\mathbf{x}, \mathbf{y}) = \inf_{\varphi} \int_{t=0}^1 \omega_{\varphi(t)}(\varphi'(t)) dt, \quad (13)$$

where the infimum is taken over all piecewise \mathcal{C}^1 -paths φ such that for all $t \in [0, 1]$, $\varphi(t) \in K$, $\varphi(0) = \mathbf{x}$, $\varphi(1) = \mathbf{y}$, and $\omega_{\mathbf{u}}(\mathbf{h})$ is the oscillation of \mathbf{h} and \mathbf{u} , defined by $\omega_{\mathbf{u}}(\mathbf{h}) := \max_i (h_i/u_i) - \min_j (h_j/u_j)$. The proof of the theorem mimics that of [21, Theorem 2.4], where a similar result is proved for $K^{(1)} := \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} > \mathbf{0}, \sum_{i=1}^n x_i = 1\}$, but we integrate over a different geodesic curve from \mathbf{x} to \mathbf{y} . A related result is also proved in [11], 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 $K := \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} > \mathbf{0}\}$. For all $\mathbf{x} \in G$, define*

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

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

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

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

Let $\mathbf{x}, \mathbf{y} \in G$. It is well known that the path $\varphi(t) = (1-t)\mathbf{x} + t\mathbf{y}$ is a geodesic curve from \mathbf{x} to \mathbf{y} for the Hilbert's projective metric (i.e., φ is a minimizer of expression (13)), see [21, Theorem 2.1]. It follows that for all functions $\alpha : [0, 1] \rightarrow (0, \infty)$ 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 \mathbf{x} to \mathbf{y} following the great circle, $\varphi_C(t) := \varphi(t)/\|\varphi(t)\|$ is a geodesic curve from \mathbf{x} to \mathbf{y} in the Hilbert's projective metric.

We can now use expression (13) with the path $t \mapsto f(\varphi_C(t))$ to obtain a bound of $d_H(f(\mathbf{x}), f(\mathbf{y}))$:

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

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

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

where the last expression follows from the fact that φ_C is a geodesic from \mathbf{x} to \mathbf{y} . \square

We are now ready to prove Theorem 3.1. By Theorem (A.1), the Lipschitz constant of the restriction of $\mathbf{x} \mapsto \exp \mathbf{x}$ to $K^{(2)}$ (with respect to d_H) is bounded from above by

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

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

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

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

Finally, we consider the maximization problem with respect to \mathbf{v} to find the value of λ_0 . Observe that we can assume without loss of generality that $\|\mathbf{v}\|_2 = 1$, because multiplying \mathbf{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 ℓ_∞ -norm, and a and b satisfy $a^2 + b^2 = \|\mathbf{v}\|_2^2 = 1$. We have shown above that the denominator is equal to $(ab)^{-1} \|\mathbf{v}\|_2^3 = (ab)^{-1}$. Hence,

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

□

B An SAA Formulation with capacity for add-on jobs

As for the situation with elective cases only, we obtain an SAA formulation for the stochastic program

$$\min_{(\mathbf{x}, \mathbf{z}) \in \mathcal{X}} \mathbb{E}_{(\mathbf{d}, \boldsymbol{\delta})} [F(\mathbf{x}, \mathbf{z}; \mathbf{d}, \boldsymbol{\delta})] \quad (14)$$

by sampling N_S duration scenarios $(\mathbf{d}^{(1)}, \boldsymbol{\delta}^{(1)}), \dots, (\mathbf{d}^{(N_S)}, \boldsymbol{\delta}^{(N_S)})$. To do this, for each scenario $i \in \mathcal{S} := \{1, \dots, N_S\}$ we first sample the number $k(i)$ of add-on cases according to the probabilities p_k . Then, we sample independent durations following $\log d_j^{(i)} \sim \mathcal{N}(\mu_j, \sigma_j^2)$ for

the elective patient blocks ($j \in \mathcal{J}$), and $\log \delta_k^{(i)} \sim \mathcal{N}(\mu_0, \sigma_0^2)$ for the emergency cases ($k = 1, \dots, k(i)$). Then, the MIP to solve is similar to (10), but we minimize the average cost instead of the worst-case cost:

$$\min_{\mathbf{x}, \mathbf{z}, \mathbf{o}} \quad \frac{1}{N_S} \sum_{i=0}^{N_S} \left(\sum_{m \in \mathcal{M}} c_f^m z_{k(i) m} + c_o^m o_{im} \right) \quad (15a)$$

$$\text{s.t.} \quad o_{im} \geq \sum_{j \in \mathcal{J}} x_{jm} d_j^{(i)} + \sum_{k=1}^{k(i)} x_{km} \delta_k^{(i)} - z_{k(i) m} T_m, \quad \forall i \in \mathcal{S}, \forall m \in \mathcal{M}, \quad (15b)$$

$$o_{im} \geq 0, \quad \forall i \in \mathcal{S}, \forall m \in \mathcal{M}, \quad (15c)$$

$$(\mathbf{x}, \mathbf{z}) \in \hat{\mathcal{X}}. \quad (15d)$$