

Lagrangian duality for robust problems with decomposable functions: the case of a robust inventory problem *

Filipe Rodrigues, Agostinho Agra, Cristina Requejo

*Department of Mathematics, University of Aveiro, Portugal
{fmgrdrigues, aagra, crequejo}@ua.pt*

Erick Delage

*Department of Decision Sciences, HEC Montréal, Montreal, Qc, H3T 2A7, Canada
Group for Research in Decision Analysis (GERAD), Montreal, Qc, H3T 1J4, Canada
erick.delage@hec.ca*

Abstract

We consider a class of min-max robust problems in which the functions that need to be *robustified* can be decomposed as the sum of arbitrary functions. This class of problems includes many practical problems such as the lot-sizing problem under demand uncertainty. By considering a Lagrangian relaxation of the uncertainty set we derive a tractable approximation, called the dual Lagrangian approach, that we relate with both the classical dualization approximation approach and an exact approach. Moreover we show that the dual Lagrangian approach coincides with the affine decision rule approximation approach.

The dual Lagrangian approach is applied to a lot-sizing problem, where demands are assumed to be uncertain and to belong to the uncertainty set with a budget constraint for each time period.

Using the insights provided by the interpretation of the Lagrangian multipliers as penalties in the proposed approach, two heuristic strategies, a new guided iterated local search heuristic and a subgradient optimization method, are designed to solve more complex lot-sizing problems where additional practical aspects, such as setup costs, are considered. Computational results show the efficiency of the proposed heuristics which provide a good compromise between the quality of the robust solutions and the running time required in their computation.

Keywords: Lagrangian relaxation; robust optimization; lot-sizing; demand uncertainty; affine approximation; budgeted uncertainty polytope

1 Introduction

Dealing with uncertainty is very important when solving practical problems where some decisions need to be taken before the real data is revealed. This is the case of inventory management problems where some decisions, such as the quantities to be produced or ordered, need to be taken without knowing the exact demands. A recent and popular approach to deal with

*Cite as: F. Rodrigues, A. Agra, C. Requejo, E. Delage. Lagrangian duality for robust problems with decomposable functions: the case of a robust inventory problem. *INFORMS Journal on Computing*, **33** (2), Pages 419-835, 2021. [doi: 10.1287/ijoc.2020.0978]

such uncertain optimization problems is Robust Optimization (RO). RO was first introduced by Soyster (1973) who proposed a model for linear optimization where constraints had to be satisfied for all possible data values.

Ben-Tal and Nemirovski (1999), El-Ghaoui and Lebret (1997), Bertsimas and Sim (2003, 2004) propose computationally tractable approaches to handle uncertainty and avoid excessive conservatism. For a recent paper on a less conservative variant of RO see Roos and den Hertog (2017). For general reviews on RO see Ben-Tal et al. (2009) and Bertsimas et al. (2011).

Although current research on RO is being very useful and different approaches have been proposed, there is a large gap in the research devoted to applying those approaches to complex mixed-integer problems. This is the case of many practical production planning problems with demand uncertainty, which motivated our work. While deterministic production planning problems have been extensively studied both from a practical and a theoretical viewpoint (Pochet and Wolsey 2006), robust applications are still scarce. Two seminal works on robust inventory problems consists of i) the study of robust basestock levels, by Bienstock and Özbay (2008), where a decomposition approach to solve the true min-max problem to optimality is proposed (henceforward denoted by BO approach), and ii) the dualization approach introduced by Bertsimas and Thiele (2006) (henceforward denoted by BT approach) to inventory problems adapted from the general approach proposed by Bertsimas and Sim (2004). The two approaches have been applied to more complex problems. The decomposition approach for the min-max problem using the budget polytope is investigated by Agra et al. (2016b) for a larger class of robust optimization problems where the first-stage decisions can be represented by a permutation. The general decomposition procedure, regarded as row-column generation, is described for general robust optimization problems by Zeng and Zhao (2013). The BO approach is also used to solve more complex inventory problems, for example, the robust maritime inventory problem (Agra et al. 2018a), and a production and inventory problem with the option of remanufacturing (Attila et al. 2017). The dualization approach is also very popular since it often leads to tractable models. Wei et al. (2011) extended the results presented by Bertsimas and Thiele (2006) to a production and inventory problem with remanufacturing, where uncertainty is considered on returns and demands. For the application of the dualization approach to a robust inventory routing problem see Solyali et al. (2012).

Solving the true min-max problem to optimality, using for instance a decomposition algorithm, can be impractical for many inventory problems, while the dualization approach may produce solutions which are too conservative. In order to circumvent both the difficulty of solving the min-max problem and the conservativeness of the dualization approach, other approaches, such as the use of affine decision rules (Ben-Tal et al. 2004, Chen and Zhang 2009) have been proposed. The affine decision rules often lead to less conservative solutions than the ones obtained with the dualization approach, and in some cases they can lead to optimal solutions, see Bertsimas et al. (2011), Bertsimas and Goyal (2012), Kuhn et al. (2011), Iancu et al. (2013). Furthermore, for special uncertainty sets the use of affine decision rules leads to computationally tractable affinely adjustable robust counterpart (AARC) models. In particular, when the uncertainty set is a polyhedron the resulting AARC model is linear, see Ben-Tal et al. (2004). Georghiou et al. (2019) propose an approach that combines the affine decision rules with the extreme point reformulation used in the exact decomposition methods.

Tractable AARC models can also be obtained for lot-sizing problems when the demands are uncertain and belong to the uncertainty set with a budget constraint for each time period. However, when additional aspects are included in the lot-sizing problems, such tractable models can become computationally hard to solve even for small size instances. The results in this

paper express how difficult can be to solve the AARC model when setup costs are considered in the basic lot-sizing problem. Such results justify the need for developing simpler tractable models as well as the use of approximation heuristic schemes. For a survey on adjustable robust optimization see Yanikoglu et al. (2018). For a deeper discussion of other conservative approximations for the min-max problem obtained through relaxations of the uncertainty set we refer to Ardestani-Jaafari and Delage (2016) and Gorissen and den Hertog (2013).

For many practical production planning and inventory management problems some data, such as demands, are not known in advance, and several decisions need to be taken before the data is revealed. Frequently, such decisions are taken before the start of the planning horizon and are not adjustable to the data when it is revealed. That is the case of decisions such as the amount of each item to produce in each time period, when complex aspects such as setups, sequence dependent changeovers, etc. are present (Pochet and Wolsey 2006). Adjusting the production to the known demands can imply new setups, creating different sequences of products that may not be implementable. Another example in inventory management occurs in maritime transportation, where the distribution must be planned in advance and can hardly be adjusted to the demands given the long transportation times. Motivated by such applications we focus on robust problems in which the functions to be *robustified* can be decomposed as the sum of arbitrary functions. This class of problems was also investigated by Delage et al. (2018). The authors proposed a new robust formulation for generic uncertainty sets where it is assumed that the functions to be *robustified* are decomposed into the sum of functions, each one involving a different nonadjustable variable, which is not the case we consider in this paper.

Contributions

In this paper, for a class of RO problems with decomposable functions, we propose a reformulation of the inner maximization subproblem occurring in a min-max model, known as adversarial problem. This reformulation starts by creating copies of both the uncertain variables and the uncertainty set in a way that the uncertainty set becomes constraint-wise independent. Further, a set of additional constraints is imposed enforcing that all the copies take identical values. By relaxing those constraints in the usual Lagrangian way, we obtain a mixed integer linear model, called Lagrangian dual model, that permits to directly relate the min-max approach with the dualization approach (obtained when such constraints are ignored). With the obtained model, it is possible to derive efficient heuristic approximation schemes that use the information from the Lagrangian multipliers to obtain solutions with lower true cost.

Our main contributions are the following:

1. Exploit the Lagrangian relaxation of the uncertainty set to obtain a tractable model for a class of RO min-max problems in which the function to be *robustified* is decomposable in the sum of the maximum of affine functions.
2. Provide a better theoretical understanding of the relations between several approaches for RO problems with decomposable functions. In particular we show that our Lagrangian dual model coincides with the AARC model and that the classical dualization approach results from the Lagrangian dual approach with all the Lagrangian multipliers null.
3. Provide computational results for the lot-sizing problem with setups showing the impact of the setup costs on the several approaches considered. In particular, when the setup costs increase, the quality of the solutions obtained by the BT approach deteriorates rapidly. This behaviour was not observed when using the proposed Lagrangian dual model. For large setup costs, the BT approach provides a bound that is up to 28% larger than the

optimal solution provided by the BO approach, while an optimal choice of multipliers can reduce it to near 6%. A similar reduction on the gap can be quickly achieved by solving the Lagrangian dual model with the multipliers fixed to their optimal value in the linear relaxation.

4. Design efficient heuristic schemes. We propose a new Guided Iterated Local Search heuristic and a Subgradient Optimization method that explicitly uses the interpretation of the Lagrangian multipliers as penalties. Comparing with other heuristics, for large size instances, the Subgradient Optimization method runs in a shorter time, and finds solutions with true costs that are i) strictly better for 91.8% of the instances used and ii) up to 18.4% better than those obtained by the BT approach.

The paper is organized as follows. In Section 2 a dual Lagrangian approach is presented for RO problems with decomposable functions and its relation with the known approaches is established. The dual Lagrangian approach is applied to the robust inventory problem in Section 3. Heuristics based on the interpretation of the Lagrangian multipliers, including a new Guided Iterated Local Search heuristic and a Subgradient Optimization method are also presented in Section 3. Computational tests are reported in Section 4 and final conclusions are given in Section 5.

2 Lagrangian duality for RO problems with decomposable functions

Consider the min-max robust model

$$R^* = \min_{u \in U} R(u)$$

with

$$R(u) = g(u) + \max_{\xi \in \Omega} \sum_{t \in T} f_t(u, \xi)$$

where U is a feasible set, $\Omega \subseteq \mathbb{R}^n$ is some compact uncertainty set, $T = \{1, \dots, n\}$ and each $f_t : U \times \Omega \rightarrow \mathbb{R}$ is an arbitrary continuous function. Variables u represent non-adjustable decisions. The decision maker chooses a vector u while an adversary determines the uncertain vector $\xi \in \Omega$ that is most unfavorable to the decision $u \in U$. Problem $R(u)$ is known as the adversarial problem (Bienstock and Özbay 2008) and it calculates what is called the *true cost* for the vector u .

Problem R^* can be rewritten as a two-stage robust problem by using adjustable variables θ_t , such that $\theta_t(\xi) : \Omega \rightarrow \mathbb{R}$, $t \in T$, as follows.

$$\begin{aligned} R^* &= \min_{u, \theta(\cdot)} \max_{\xi \in \Omega} g(u) + \sum_{t \in T} \theta_t(\xi) \\ \text{s.t.} \quad &\theta_t(\xi) \geq f_t(u, \xi), \quad \forall \xi \in \Omega, t \in T, \\ &u \in U. \end{aligned} \tag{2.1}$$

When particular functions $\theta_t(\xi)$ are considered, conservative approaches of R^* are obtained. In particular, the usual non-adjustable approach examines the case where $\theta_t(\xi) = \theta_t$, $t \in T$, that is,

$$\begin{aligned} C^* &= \min_{u, \theta} g(u) + \sum_{t \in T} \theta_t \\ \text{s.t.} \quad &\theta_t \geq f_t(u, \xi), \quad \forall \xi \in \Omega, t \in T, \\ &u \in U. \end{aligned} \tag{2.2}$$

It is known (Bienstock and Özbay 2008) that the gap between the two approaches can be large. However, there are cases with no gap between these approaches, that is, $R^* = C^*$ (see Ben-Tal et al. (2004), Marandi and den Hertog (2018) and El Housni and Goyal (2018)). In particular, this equality holds when the uncertainty region Ω is the Cartesian product of sets Ξ_t (that is $\Omega = \Xi_1 \times \dots \times \Xi_n$), and each function $f_t(u, \xi)$ in constraints (2.2) is only affected by the terms of ξ which lie in Ξ_t :

$$R^* = \min_{u \in U} \max_{\xi \in \Omega} g(u) + \sum_{t \in T} f_t(u, \xi_t) = \min_{u \in U} g(u) + \sum_{t \in T} \max_{\xi_t \in \Xi_t} f_t(u, \xi_t) = C^*.$$

Here we explore this property to derive a Lagrangian relaxation of the adversarial problem. First, for each constraint $t \in T$, create a list of copies $\{\zeta^t\}_{t \in T}$ of the variables ξ and a list of respective uncertainty sets $\{\Omega^t\}_{t \in T}$, such that each $\Omega \subseteq \Omega^t$ and $\bigcap_{t \in T} \Omega^t = \Omega$ (e.g. for simplicity one can use $\Omega^t := \Omega$). We further impose a set of constraints enforcing that all the copies must be equal. This leads to the following exact reformulation of $R(u)$:

$$\begin{aligned} R(u) = g(u) + \max_{\zeta^1, \dots, \zeta^n} \sum_{t \in T} f_t(u, \zeta^t) \\ \text{s.t. } \zeta^t = \zeta^1, \quad \forall t \in \{2, \dots, n\}, \\ \zeta^t \in \Omega^t, \quad \forall t \in T. \end{aligned} \quad (2.3)$$

Remark 1. *In relation to the set of equalities (2.3), it is important to notice that one could impose additional redundant equalities $\zeta^t = \zeta^\ell$ for $t \neq \ell$ or replace them with other equivalent sets of equations. For all those cases, the process derived next still holds.*

Attaching Lagrangian multipliers $\lambda^t \in \mathbb{R}^n$ to each constraint (2.3) and dualizing these constraints in the usual Lagrangian way, the following Lagrangian relaxation of $R(u)$ is obtained

$$\begin{aligned} \overline{LR}(u, \lambda) := g(u) + \max_{\zeta^1, \dots, \zeta^n} \sum_{t \in T} f_t(u, \zeta^t) - \sum_{t=2}^n (\lambda^t)^\top (\zeta^t - \zeta^1) \\ \text{s.t. } \zeta^t \in \Omega^t, \quad \forall t \in T. \end{aligned}$$

The multipliers λ penalize the use of different uncertainty vectors for different constraints. Imposing that $\lambda^1 := -\sum_{t=2}^n \lambda^t$ this model is equivalent to

$$\begin{aligned} LR(u, \lambda) := g(u) + \max_{\zeta^1, \dots, \zeta^n} \sum_{t \in T} f_t(u, \zeta^t) - \sum_{t \in T} (\lambda^t)^\top \zeta^t \\ \text{s.t. } \zeta^t \in \Omega^t, \quad \forall t \in T. \end{aligned}$$

By using the epigraph reformulation, model $LR(u, \lambda)$ can be written as follows.

$$\begin{aligned} LR(u, \lambda) = g(u) + \min_{\theta_1, \dots, \theta_n} \sum_{t \in T} \theta_t \\ \text{s.t. } \theta_t \geq f_t(u, \zeta^t) - (\lambda^t)^\top \zeta^t, \quad \forall \zeta^t \in \Omega^t, t \in T, \end{aligned}$$

For a given u and λ , the minimization problem in $LR(u, \lambda)$ can be separated into n independent subproblems, one for each $t \in T$:

$$\begin{aligned} LR_t(u, \lambda^t) = \min_{\theta_t} \theta_t \\ \text{s.t. } \theta_t \geq f_t(u, \zeta^t) - (\lambda^t)^\top \zeta^t, \quad \forall \zeta^t \in \Omega^t \end{aligned}$$

and $LR(u, \lambda) = g(u) + \sum_{t=1}^n LR_t(u, \lambda^t)$.

The Lagrangian dual problem is $DLR(u) = \min_{\lambda} LR(u, \lambda)$. Hence we have $R(u) \leq DLR(u)$. Denoting by D the problem $D = \min_{u \in U} DLR(u)$, the following relation holds

$$R^* = \min_{u \in U} R(u) \leq \min_{u \in U} DLR(u) = D.$$

The Lagrangian dual model D can be written as follows:

$$\begin{aligned} D = \min_{u, \lambda, \theta} \quad & g(u) + \sum_{t=1}^n \theta_t \\ \text{s.t.} \quad & \theta_t \geq f_t(u, \zeta^t) - (\lambda^t)^\top \zeta^t, \quad \forall \zeta^t \in \Omega^t, t \in T, \\ & \lambda^1 = - \sum_{t=2}^n \lambda^t, \\ & u \in U. \end{aligned}$$

Define $D(\lambda) = \min_{u \in U} LR(u, \lambda)$. Hence, $D = \min_{\lambda} D(\lambda)$, and for given multipliers λ ,

$$\begin{aligned} D(\lambda) = \min_{u, \theta} \quad & g(u) + \sum_{t=1}^n \theta_t \\ \text{s.t.} \quad & \theta_t \geq f_t(u, \zeta^t) - (\lambda^t)^\top \zeta^t, \quad \forall \zeta^t \in \Omega^t, t \in T, \\ & u \in U. \end{aligned}$$

Noticing that C^* is obtained with $\lambda = 0$ and $\Omega^t = \Omega$, we have the following relation.

Theorem 2. *When $\Omega^t = \Omega$, we have $R^* \leq D \leq C^*$.*

Proof. Proof. $R^* = \min_{u \in U} R(u) \leq \min_{u \in U} \min_{\lambda} LR(u, \lambda) = D = \min_{\lambda} D(\lambda) \leq D(0) \leq C^*$. \square

Following the work of Ben-Tal et al. (2015), one can start identifying conditions under which D becomes tractable. In particular, Section 2.3 will focus on the case where each $f_t(u, \xi)$ captures the maximum of a finite set of affine functions and Ω^t is a famous polyhedral budgeted set.

Theorem 3. *Given that $g(u)$ is a convex function, that $f_t(u, \xi)$ is the maximum of K functions $h_{tk}(u, \xi)$ convex in u and concave in ξ , and that both U and each Ω^t are compact convex sets, then D can be reformulated as the following finite dimensional convex program:*

$$\begin{aligned} D = \min_{u, \lambda, \theta, v} \quad & g(u) + \sum_{t=1}^n \theta_t \\ \text{s.t.} \quad & \theta_t \geq \delta^*(v_{tk} | \Omega^t) - h_{tk^*}(u, v_{tk} + \lambda^t) \quad \forall k = 1, \dots, K, t \in T, \\ & \lambda^1 = - \sum_{t=2}^n \lambda^t, \\ & u \in U, \end{aligned} \tag{2.4}$$

where $\delta^*(v | \Omega^t) := \sup_{\xi \in \Omega^t} v^\top \xi$ is the support function for Ω^t , while $h_{tk^*}(u, v) := \inf_{\xi} \{v^\top \xi - h_{tk}(u, \xi)\}$ is the partial concave conjugate of $h_{tk}(u, \xi)$. Moreover, if the epigraph of functions $g(u)$, $\delta^*(v | \Omega^t)$, and $h_{tk^*}(u, v)$ are polyhedral representable and set U is a polyhedron, then problem (2.4) can be reduced to a linear program.

The proof of this theorem is given in Appendix A.

2.1 Dualization versus affine approximation

It is known that less conservative approaches than C^* to approximate R^* can be obtained assuming that $\theta_t(\xi)$ is the affine approximation $\theta_t(\xi) = \nu_0^t + (\nu^t)^\top \xi$, with $\nu_0^t \in \mathbb{R}$ and $\nu^t \in \mathbb{R}^n$. The resulting model, called Affinely Adjustable Robust Counterpart (AARC) model (Gorissen and den Hertog 2013) is

$$\begin{aligned} AARC = \min_{u, \nu_0, \nu, s} \quad & g(u) + s \\ \text{s.t.} \quad & s \geq \sum_{t=1}^n \left(\nu_0^t + (\nu^t)^\top \xi \right), \quad \forall \xi \in \Omega, \\ & \nu_0^t + (\nu^t)^\top \xi \geq f_t(u, \xi), \quad \forall \xi \in \Omega, t \in T, \\ & u \in U \end{aligned}$$

and it holds $R^* \leq AARC \leq C^*$.

Next we establish the main result of this section stating that the Lagrangian dual bound D obtained with $\Omega^t = \Omega$, $t \in T$ coincides with the affine approximation $AARC$.

Theorem 4. *When $\Omega^t = \Omega$, $t \in T$ we have that $AARC = D$.*

Proof. Proof. When $\Omega^t = \Omega$, model D can be obtained from model $AARC$ by replacing ν_0^t with θ_t and ν^t with λ^t , and adding the constraint $\sum_{t=1}^n \nu^t = 0$, hence $AARC \leq D$.

To prove that $AARC \geq D$, we can show that given any feasible solution (ν_0, ν) of $AARC$ that achieves a finite objective value, it is possible to construct a feasible solution for D that achieves the same objective value. To do so, set $m = \max_{\xi \in \Omega} \sum_{t=1}^n (\nu^t)^\top \xi$, $\lambda^1 = -\sum_{t=2}^n \nu^t$, and $\theta_1 = \nu_0^1 + m$, and for each $t \in \{2, \dots, n\}$ we define $\lambda^t = \nu^t$, and $\theta_t = \nu_0^t$. Clearly, the objective value for D is the same as achieved in $AARC$:

$$g(u) + \sum_{t=1}^n \nu_0^t + \max_{\xi \in \Omega} \sum_{t=1}^n (\nu^t)^\top \xi = g(u) + \sum_{t=1}^n \nu_0^t + m = g(u) + \sum_{t=1}^n \theta_t,$$

and the solution (θ, λ) is feasible for D , since for the constraint $t = 1$ (the remaining constraints are easily shown to be equivalent) and for each $\xi \in \Omega$, we have

$$\theta_1 + (\lambda^1)^\top \xi = \nu_0^1 + m - \sum_{t=2}^n (\nu^t)^\top \xi \geq \nu_0^1 + (\nu^1)^\top \xi \geq f_1(u, \xi)$$

where the first inequality follows from the definition of m since

$$m = \max_{\xi \in \Omega} \sum_{t=1}^n (\lambda^t)^\top \xi \geq \sum_{t=1}^n (\lambda^t)^\top \xi, \quad \forall \xi \in \Omega \Rightarrow m - \sum_{t=2}^n (\lambda^t)^\top \xi \geq (\lambda^1)^\top \xi, \quad \forall \xi \in \Omega,$$

and the second inequality follows from the feasibility of (ν_0, ν) in $AARC$. □

2.2 Extensions and related problems

Next we discuss three extensions of Theorem 4.

Two-stage robust linear programs with box uncertainty: First, we extended the result to two-stage robust linear programs where the uncertainty set is the box uncertainty set. Let us consider a two-stage robust linear program of the form

$$\begin{aligned} \min_{u \in U} \max_{\xi \in \Xi} \min_y \quad & g(u) + c^\top y \\ \text{s.t.} \quad & Au + By \leq D\xi. \end{aligned} \tag{2.5}$$

where $A \in \mathbb{R}^{m \times q}$, $B \in \mathbb{R}^{m \times p}$, $D \in \mathbb{R}^{m \times n}$ and $\Xi = \{\xi \geq 0 \mid \xi \leq d\}$.

By dualizing model (2.5) over the second-stage variables y and then over the uncertain variables ξ , Bertsimas and de Ruiter (2016) obtain the Dual Reformulation (DR) of model (2.5)

$$DR = \min_{u \in U} \max_{\zeta \in \Omega} \min_{z \geq 0} g(u) + \zeta^\top Au + d^\top z \quad (2.6)$$

$$s.t. \quad z \geq -D^\top \zeta.$$

where $\Omega = \{\zeta \geq 0 \mid c + B^\top \zeta = 0\}$. Bertsimas and de Ruiter (2016) proved that when affine decisions rules are applied to both the primal model (2.5) and the dual model (2.6), the optimal values of the resulting models coincide. The DR model (2.6) can easily be rewritten as an instance of the general model R^* defined in Section 2 as follows:

$$DR = \min_{u \in U, \theta(\cdot)} g(u) + \sum_{t=1}^n \theta_t(\zeta) \quad (2.7)$$

$$s.t. \quad \theta_t(\zeta) \geq \max(0, -d_t D_{:,t}^\top \zeta) + (1/n) \zeta^\top Au \quad \forall \zeta \in \Omega, t \in T,$$

where $D_{:,t}$ refers to the t -th column of D (see Zhen et al. (2018) for elimination of the adaptive variables). We can therefore state the following result as a consequence of Theorem 2 in Bertsimas and de Ruiter (2016).

Corollary 1. *The optimal value of the Lagrangian dual model of the dual reformulation model (2.7) coincides with the optimal value of the AARC applied either to the primal formulation (2.5) ($y(\xi) = y_0 + V\xi$) or to the dual formulation (2.6) ($\theta_t(\zeta) = \nu_0^t + (\nu^t)^\top \zeta$).*

Quadratic decision rules: In a different direction, a similar result to the one proved in Theorem 4 can be derived for the case of the quadratic decision rules. By defining for each $t \in T$, $\theta_t(\xi) = \mu_0^t + (\mu^t)^\top \xi + \xi^\top \Pi^t \xi$, with $\Pi^t \in \mathbb{R}^{n \times n}$, the obtained Quadratic Adjustable Robust Counterpart (QARC) model can be written as follows:

$$QARC = \min_{u, \nu_0, \nu, \Pi, s} g(u) + s$$

$$s.t. \quad s \geq \sum_{t=1}^n \left(\nu_0^t + (\nu^t)^\top \xi + \xi^\top \Pi^t \xi \right), \quad \forall \xi \in \Omega,$$

$$\nu_0^t + (\nu^t)^\top \xi + \xi^\top \Pi^t \xi \geq f_t(u, \xi), \quad \forall \xi \in \Omega, t \in T,$$

$$u \in U.$$

Following the idea of the proposed Lagrangian approach, in addition to the set of equality constraints (2.3), let us consider in model $R(u)$ the new set of equalities

$$\xi^t (\xi^t)^\top = \xi^1 (\xi^1)^\top, \quad t = 2, \dots, n \quad (2.8)$$

Attaching a matrix $\Lambda^t \in \mathbb{R}^{n \times n}$ of Lagrangian multipliers to each one of constraints (2.8), the Lagrangian Dual Quadratic (DQ) model becomes

$$DQ = \min_{u, \lambda, \Lambda} g(u) + \sum_{t=1}^n \theta_t$$

$$s.t. \quad \theta_t \geq f_t(u, \xi^t) - (\lambda^t)^\top \xi^t - (\xi^t)^\top \Lambda^t \xi^t, \quad \forall \xi^t \in \Omega^t, t \in T,$$

$$\lambda^1 = - \sum_{t=2}^n \lambda^t, \Lambda^1 = - \sum_{t=2}^n \Lambda^t,$$

$$u \in U.$$

We can once again establish a connection between this dual model and the use of quadratic decision rules.

Theorem 5. When $\Omega^t = \Omega$, for all $t \in T$, we have that $QARC = DQ$.

The proof of this result follows straightforwardly the steps of the proof of Theorem 3. In particular, for showing that $QARC \geq DQ$, one can define $m = \max_{\xi \in \Omega} \{\sum_{t=1}^n (v^t)^\top \xi + \xi^\top \Pi^t \xi\}$, $\lambda^1 = -\sum_{t=2}^n v^t$, $\Lambda^1 = -\sum_{t=2}^n \Pi^t$, $\theta_1 = \nu_0^1 + m$ and, for all $t \in \{2, \dots, n\}$, $\lambda^t = \nu^t$, $\Lambda^t = \Pi^t$, and $\theta_t = \nu_0^t$.

Distributionally robust optimization: Finally, the form of problem R^* is not limited to classical robust optimization but also emerges quite naturally when handling distributionally robust optimization (DRO) problems. In particular, consider the following general moment-based DRO problem:

$$\min_{u \in \mathcal{U}} \sup_{F \in \mathcal{D}(\mathcal{U})} \mathbb{E}_F \left[\sum_t f_t(u, \xi) \right],$$

where ξ is now considered to be drawn from a distribution F that lies in an ambiguity set \mathcal{D} defined as

$$\mathcal{D}(\mathcal{U}) := \left\{ F \mid \exists \mu \in \mathcal{U}, \begin{array}{l} \mathbb{P}_F(\xi \in \Omega) = 1 \\ \mathbb{E}_F[h_k(\xi)] = \mu_k, \forall k = 1, \dots, K \end{array} \right\},$$

with each $h_k(\xi)$ defining a moment function, and $\mathcal{U} \subset \mathbb{R}^K$ defining the set of possible moments. Exploiting a famous reformulation for moment problems, one can, under fairly general conditions, reformulate the inner supremum as follows:

$$\begin{aligned} \sup_{F \in \mathcal{D}(\mathcal{U})} \mathbb{E}_F \left[\sum_t f_t(u, \xi) \right] &= \sup_{\mu \in \mathcal{U}} \sup_{F \in \mathcal{D}(\{\mu\})} \mathbb{E}_F \left[\sum_t f_t(u, \xi) \right] \\ &= \sup_{\mu \in \mathcal{U}} \inf_q \sup_{\xi \in \Omega} \sum_t f_t(u, \xi) + \sum_{k=1}^K q_k (h_k(\xi) - \mu_k) \\ &= \inf_q \sup_{\mu \in \mathcal{U}, \xi \in \Omega} \sum_t f_t(u, \xi) + \sum_{k=1}^K q_k (h_k(\xi) - \mu_k). \end{aligned}$$

where the first step assumes that strong semi-infinite conic duality holds (see Shapiro (2001) for more details), followed by an application of Sion's minimax theorem as long as \mathcal{U} is convex and bounded. This gives rise to the following reformulation of the DRO problem:

$$\min_{u \in \mathcal{U}, q} \sup_{\mu \in \mathcal{U}, \xi \in \Omega} \sum_t f_t(u, \xi) + \sum_{k=1}^K q_k (h_k(\xi) - \mu_k), \quad (2.9)$$

One can directly see that this DRO reformulation takes the form of R^* . Theorem 4 therefore applies to the DRO problem reformulation (2.9).

2.3 Duality for the *B&T* budgeted set and for the maximum of affine functions

Here we consider the particular case that motivated our work, where functions $f_t(u, \zeta^t)$ are given by the maximum of affine functions. We consider the uncertainty set used by Bertsimas and Thiele (2006):

$$\Omega = \{ \xi \in [-1, 1]^n \mid \sum_{j=1}^t |\xi_j| \leq \Gamma_t, t \in T \},$$

where a budget constraint is imposed for each time period and we refer to this set as the *B&T* budgeted set.

We assume that $f_t(u, \zeta^t) = \max_{k \in K} \hat{f}_t^k(u, \zeta^t)$ where K is a finite set of indexes and

$$\hat{f}_t^k(u, \zeta^t) = L_t^k(u) + \sum_{j \in T} a_j^{tk} \zeta_j^t, \quad \forall t \in T, k \in K,$$

where $a_j^{tk} \in \mathbb{R}$ and $L_t^k(u) : U \rightarrow \mathbb{R}$ is an affine function for all $k \in K$ and $j, t \in T$.

When $\Omega^t = \Omega$, the Lagrangian dual problem takes the form

$$\begin{aligned} \min_{u, \theta, \lambda} \quad & g(u) + \sum_{t=1}^n \theta_t \\ \text{s.t.} \quad & \theta_t \geq L_t^k(u) + \left(\sum_{j \in T} a_j^{tk} \zeta_j^t - \sum_{j \in T} \zeta_j^t \lambda_j^t \right), \quad \forall \zeta^t \in \Omega^t, k \in K, t \in T, \\ & \lambda^1 = - \sum_{t=2}^n \lambda^t, \\ & u \in U. \end{aligned}$$

The uncertainty sets Ω^t are not sets of linear constraints due to the presence of the absolute value function in constraints $\sum_{\ell=1}^j |\zeta_\ell^t| \leq \Gamma_j$, $j \in T$. There are several ways of converting sets Ω^t into equivalent sets of linear constraints (see Ben-Tal et al. (2009)). Preliminary tests using different forms of conversion indicate that the best results are obtained by replacing ζ^t with $\zeta^{t+} - \zeta^{t-}$ such that $(\zeta^{t+}, \zeta^{t-}) \in \bar{\Omega}^t$ and

$$\bar{\Omega}^t = \left\{ (\zeta_j^{t+}, \zeta_j^{t-}) \in \mathbb{R}^t \times \mathbb{R}^t \mid \sum_{\ell=1}^j (\zeta_\ell^{t+} + \zeta_\ell^{t-}) \leq \Gamma_j, \quad j \in T, \right. \quad (2.10)$$

$$\left. \begin{aligned} & \zeta_j^{t+} + \zeta_j^{t-} \leq 1, & j \in T, \\ & \zeta_j^{t+}, \zeta_j^{t-} \geq 0, & j \in T \end{aligned} \right\}. \quad (2.11)$$

For practical reasons, to reduce the size of the resulting model, we assume henceforward that $\zeta_j^t = 0$ for $j > t$, which implies that constraints (2.10) and (2.11) can be disregarded for $j > t$. By considering the above linear transformation, one can apply linear programming duality to reformulate each robust constraint and obtain the following linear program:

$$\begin{aligned} D = \min_{u, \lambda, \theta, q, r} \quad & g(u) + \sum_{t=1}^n \theta_t \\ \text{s.t.} \quad & \theta_t \geq L_t^k(u) + \sum_{j=1}^t q_j^{tk} \Gamma_j + \sum_{j=1}^t r_j^{tk}, \quad \forall t \in T, k \in K, \end{aligned} \quad (2.12)$$

$$\sum_{\ell=j}^t q_\ell^{tk} + r_j^{tk} \geq (-1)^i (a_j^{tk} - \lambda_j^t), \quad \forall j, t \in T : j \leq t, k \in K, i \in \{1, 2\} \quad (2.13)$$

$$q_j^{tk}, r_j^{tk} \geq 0, \quad \forall j, t \in T : j \leq t, k \in K, \quad (2.14)$$

$$\lambda^1 = - \sum_{t=2}^n \lambda^t, \quad (2.15)$$

$$u \in U. \quad (2.16)$$

where the dual variables q_j^{tk} and r_j^{tk} are associated with constraints (2.10) and (2.11), respectively. In practice, when T is reasonably small, it can be interesting to rewrite D in a lower dimensional space by eliminating variables r_j^{tk} . The resulting model is called projected model and is given in Appendix B. Alternatively, one might improve numerical efficiency, albeit at the price of precision, by using a simpler set $\hat{\Omega}^t$, such that $\Omega^t \subseteq \hat{\Omega}^t$. In particular, the following form is a natural choice (see Remark 6):

$$\hat{\Omega}^t = \{ \zeta^t \in [-1, 1]^n \mid \sum_{j=1}^t |\zeta_j^t| \leq \Gamma_t \},$$

and leads to the following relation:

$$R^* \leq AARC = D \leq \hat{D} \leq \hat{D}(0),$$

where \hat{D} and $\hat{D}(\lambda)$ denote, respectively, D and $D(\lambda)$ when $\hat{\Omega}^t$ is considered instead of Ω^t .

Observe that since constraints (2.10) in set $\bar{\Omega}^t$ were disregarded for $j > t$, the version of the AARC approach equivalent to model D consists of using the affine policy $\theta_t(\xi) = \nu_0^t + (\nu^t)^\top \xi$ with $\nu_j^t = 0$ for $j > t$.

Remark 6. *In the case of the B&T budgeted set, the set of constraints in the adversarial problem is given by*

$$\{(\zeta^1, \dots, \zeta^n) \in [-1, 1]^n \mid \zeta^t = \zeta^\ell, \quad t, \ell \in T, t \neq \ell, \quad (2.17)$$

$$\sum_{j=1}^{\ell} |\zeta_j^t| \leq \Gamma_t, \quad t, \ell \in T\} \quad (2.18)$$

where constraints (2.18) for $\ell < t$ are redundant in the presence of constraints (2.17). However, when constraints (2.17) are relaxed, constraints (2.18) are no longer redundant for $\ell < t$ and the corresponding Lagrangian relaxation may differ.

Proposition 7. *Given any fixed λ such that $\sum_{t \in T} \lambda^t = 0$ and letting $\alpha_j^{tk}(\lambda) = |a_j^{tk} - \lambda_j^t|$ for all $j \in T$, the value of D is bounded above by*

$$\begin{aligned} \hat{D}(\lambda) = \min_u \quad & g(u) + \sum_{t=1}^n \theta_t \\ \text{s.t.} \quad & \theta_t \geq L_t^k(u) + A_t^k(\lambda), \quad \forall t \in T, k \in K, \\ & u \in U, \end{aligned}$$

where $A_t^k(\lambda) = \sum_{\ell=1}^{\lfloor \Gamma_t \rfloor} \alpha_{j(\ell)}^{tk}(\lambda) + (\Gamma_t - \lfloor \Gamma_t \rfloor) \alpha_{j(\lfloor \Gamma_t \rfloor)}^{tk}(\lambda)$ and $\alpha_{j(\ell)}^{tk}(\lambda)$ is the ℓ^{th} largest value among the values $\alpha_1^{tk}(\lambda), \dots, \alpha_n^{tk}(\lambda)$.

The proof is similar to the proof of Proposition 1 in Bertsimas and Sim (2004) so it is omitted.

Remark 8. *All the approximation models for R^* presented in this section overestimate the cost associated with each first-stage solution u . Hence, those approaches may lead to poor bounds based on good solutions and the following relation holds*

$$R^* = R(u^*) \leq R(u^J) \leq J,$$

where u^J denotes the first-stage solution obtained with model J , and $J \in \{D, \hat{D}, \hat{D}(\lambda)\}$.

3 The case of a robust inventory problem

In this section we particularize the results of the previous section for the case of the robust inventory problem that motivated this study and relate them with those known from the literature. We consider lot-sizing problems defined over a finite time horizon of n periods and define $T = \{1, \dots, n\}$. For each time period $t \in T$, consider the unit holding cost h_t , the unit backlogging cost b_t and the unit production cost c_t . The demand in time period t is given by d_t . Define x_t as the inventory at the beginning of period t (x_1 is the initial inventory level). In

case x_t is negative it indicates a shortage. Variables $u_t \geq 0$ indicate the quantity to produce in time period t . When the demand d_t is known and fixed we obtain a basic deterministic lot-sizing problem that can be modelled as follows:

$$\begin{aligned} \min_{u,x} \quad & \sum_{t=1}^n (c_t u_t + \max\{h_t x_{t+1}, -b_t x_{t+1}\}) \\ \text{s.t.} \quad & x_{t+1} = x_1 + \sum_{j=1}^t (u_j - d_j), \quad \forall t \in T, \\ & u_t \geq 0, \quad \forall t \in T. \end{aligned}$$

If $x_{t+1} \geq 0$, then $\max\{h_t x_{t+1}, -b_t x_{t+1}\}$ gives the holding cost $h_t x_{t+1}$, otherwise it gives the backlogging cost $-b_t x_{t+1}$ at the end of time period t .

Here we consider the case where the demands d_t are defined by $d_t := \mu_t + \delta_t z_t$, for each $t \in T$, where μ_t and δ_t are the nominal demand and the maximum allowed deviation in period t , respectively, and the uncertain variables z_t belong to the *B&T* budgeted set:

$$\Omega = \{z \in [-1, 1]^n \mid \sum_{j=1}^t |z_j| \leq \Gamma_t, \quad t \in T\}.$$

and assume that $0 \leq \Gamma_1 \leq \Gamma_2 \leq \dots \leq \Gamma_n$, $\Gamma_t \leq \Gamma_{t-1} + 1$ and $1 < t \leq n$.

The results presented in this section can easily be extended to accommodate other practical aspects such as setup costs and/or other production constraints. In that case, the objective function is

$$\min_{u,y} \sum_{t=1}^n (c_t u_t + S_t y_t + \max\{h_t x_{t+1}, -b_t x_{t+1}\}) \quad (3.1)$$

where y_t is the setup variable indicating whether there is a production setup in time period t , and S_t is the setup cost in time period t . A new set of constraints is also considered

$$u_t \leq P_t y_t, \quad y_t \in \{0, 1\}, \quad \forall t \in T, \quad (3.2)$$

where P_t is an upper bound on the production quantity at period t . In order to keep the notation easy, and since all the theoretical results presented hold for both cases (with and without setups) hereafter in the derivation of the theoretical results, we consider only the simplest case where no setup costs (and no setup variables) are considered. For the computational aspects (Sections 3.3 and 4) the more general case with setups is considered.

3.1 The Bienstock and Özbay and the Bertsimas and Thiele approaches

First, we review two of the main approaches for robust inventory problems: the decomposition approach introduced by Bienstock and Özbay (2008) to solve the problem written as a min-max problem (BO approach) and the dualization approach employed by Bertsimas and Thiele (2006) (BT approach). Bienstock and Özbay (2008) consider the robust inventory problem as a min-max problem where, for a given production vector u , the demand d_t is picked by an adversary problem. The min-max formulation is the following:

$$R^* = \min_{u \geq 0} R(u)$$

where

$$\begin{aligned}
R(u) &= \max_{x,z} \sum_{t=1}^n (c_t u_t + \max\{h_t x_{t+1}, -b_t x_{t+1}\}) & (3.3) \\
s.t. \quad x_{t+1} &= x_1 + \sum_{j=1}^t (u_j - \mu_j - \delta_j z_j), & \forall t \in T, \\
\sum_{j=1}^t |z_j| &\leq \Gamma_t, & \forall t \in T, \\
z_t &\in [-1, 1], & \forall t \in T.
\end{aligned}$$

Problem (3.3) corresponds to the general adversarial problem introduced in Section 2.3 with $K = \{1, 2\}$, $L_t^1(u) = h_t \left(x_1 + \sum_{j=1}^t (u_j - \mu_j)\right)$, $L_t^2(u) = -b_t \left(x_1 + \sum_{j=1}^t (u_j - \mu_j)\right)$, $g(u) = \sum_{t \in T} c_t u_t$ and $U = \mathbb{R}_+^n$.

Bienstock and Özbay (2008) solve the min-max problem using a decomposition approach where, in the master problem, a production planning problem is solved for a subset of demand scenarios, while in the subproblem (adversarial problem) the worst-case scenario is found for the current production plan and added to the master problem. A FPTAS is proposed in Agra et al. (2016b) where a similar decomposition approach is used and the adversarial problem is solved by dynamic programming.

The dualization approach introduced by Bertsimas and Sim (2004) was developed for the robust inventory problem by Bertsimas and Thiele (2006). The formulation is as follows

$$\begin{aligned}
\hat{C}^* &= \min_{u,z} \sum_{t=1}^n (c_t u_t + \theta_t) & (3.4) \\
s.t. \quad \theta_t &\geq h_t \left(x_1 + \sum_{j=1}^t (u_j - \mu_j) + A_t\right), & \forall t \in T, \\
\theta_t &\geq -b_t \left(x_1 + \sum_{j=1}^t (u_j - \mu_j) - A_t\right), & \forall t \in T, \\
u_t &\geq 0, & \forall t \in T,
\end{aligned}$$

where, for all $t \in T$,

$$A_t = \max_z \left\{ \sum_{j=1}^t \delta_j z_j \mid \sum_{j=1}^t |z_j| \leq \Gamma_t, \quad z_t \in [-1, 1] \right\}.$$

Notice that this approach is based on the supersets $\hat{\Omega}^t$, $t \in T$.

3.2 Lagrangian relaxation based approaches

To derive the Lagrangian relaxation of the adversarial problem (3.3) consider, for each time period $t \in T$, a copy v_j^t of each variable z_j with $j \leq t$. That is, consider new variables $v_j^t \in [-1, 1]$ which account for the deviation in period j affecting period t , $t \geq j$, and impose the constraints

$$v_t^t = v_t^j, \quad \forall j, t \in T, t < j. \quad (3.5)$$

With this set of equalities, constraints $\sum_{j=1}^t |z_j| \leq \Gamma_t$, $t \in T$ are replaced by constraints $\sum_{j=1}^{\ell} |v_j^t| \leq \Gamma_{\ell}$, $1 \leq \ell \leq t \leq n$ and the following approximation for the problem R^* is obtained.

Theorem 9. *Model D defined below is a tractable approximation for the problem R^* .*

$$D = \min_{u, \lambda, \theta, q, p, r, s} \sum_{t \in T} (c_t u_t + \theta_t) \quad (3.6)$$

$$s.t. \quad \theta_t \geq L_t^1(u) + h_t \left(\sum_{j=1}^t q_j^{t1} \Gamma_j + \sum_{j=1}^t r_j^{t1} \right), \quad \forall t \in T, \quad (3.7)$$

$$\theta_t \geq L_t^2(u) + b_t \left(\sum_{j=1}^t q_j^{t2} \Gamma_j + \sum_{j=1}^t r_j^{t2} \right), \quad \forall t \in T, \quad (3.8)$$

$$q_t^{t1} + r_t^{t1} \geq (-1)^i \left(\delta_t + \sum_{j=t+1}^n \frac{\lambda_j^t}{h_t} \right), \quad \forall i \in \{1, 2\}, t \in T, \quad (3.9)$$

$$\sum_{\ell=j}^t q_\ell^{t1} + r_j^{t1} \geq (-1)^i \left(\delta_j - \frac{\lambda_j^t}{h_t} \right), \quad \forall i \in \{1, 2\}, j, t \in T : j < t, \quad (3.10)$$

$$q_t^{t2} + r_t^{t2} \geq (-1)^i \left(\sum_{j=t+1}^n \frac{\lambda_j^t}{b_t} - \delta_t \right), \quad \forall i \in \{1, 2\}, t \in T, \quad (3.11)$$

$$\sum_{\ell=j}^t q_\ell^{t2} + r_j^{t2} \geq (-1)^i \left(\frac{\lambda_j^t}{b_t} + \delta_j \right), \quad \forall i \in \{1, 2\}, j, t \in T : j < t, \quad (3.12)$$

$$q_j^{t1}, r_j^{t1}, q_j^{t2}, r_j^{t2} \geq 0, \quad \forall j, t \in T : j < t. \quad (3.13)$$

The proof of this theorem is given in Appendix C and it directly follows from the application of the process described in Section 2 to the robust inventory problem. By replacing the sets Ω^t with the supersets $\hat{\Omega}^t$ we obtain model \hat{D} , which is used in the heuristics proposed in Section 3.3. Model \hat{D} corresponds to model D by setting variables $q_j^{tk} = 0$ for all $k \in \{1, 2\}$, $j, t \in T : j < t$.

Remark 10. *The Bertsimas and Thiele model (3.4) is a model having the form of model C^* where Ω is replaced in each constraint (2.2) by $\hat{\Omega}^t$. Hence, we have the relations $D \leq C^* \leq \hat{C}^*$ and $\hat{D} \leq \hat{D}(0) \leq \hat{C}^*$.*

The projected version of model D in a lower dimension space can be written as follows:

$$D = \min_{u, \lambda, \theta, q, p} \sum_{t \in T} (c_t u_t + \theta_t) \quad (3.14)$$

$$s.t. \quad \theta_t \geq L_t^1(u) + h_t \left(\sum_{j=1}^t q_j^{t1} \Gamma_j - \sum_{j=1}^t |\pi_j^t| \sum_{\ell=j}^t q_\ell^{t1} + \sum_{j=1}^{t-1} \pi_j^t \left(\delta_j - \frac{\lambda_j^t}{h_t} \right) + \pi_t^t \left(\delta_t + \sum_{j=t+1}^n \frac{\lambda_j^t}{h_t} \right) \right),$$

$$\forall \pi_i^t \in \{-1, 0, 1\}, 1 \leq i \leq t, t \in T, \quad (3.15)$$

$$\theta_t \geq L_t^2(u) + b_t \left(\sum_{j=1}^t q_j^{t2} \Gamma_j - \sum_{j=1}^t |\pi_j^t| \sum_{\ell=j}^t q_\ell^{t2} + \sum_{j=1}^{t-1} \pi_j^t \left(\delta_j + \frac{\lambda_j^t}{b_t} \right) + \pi_t^t \left(\delta_t - \sum_{j=t+1}^n \frac{\lambda_j^t}{b_t} \right) \right),$$

$$\forall \pi_i^t \in \{-1, 0, 1\}, 1 \leq i \leq t, t \in T, \quad (3.16)$$

$$q_j^{t1}, q_j^{t2} \geq 0, \quad \forall j, t \in T : j \leq t. \quad (3.17)$$

Remark 11. *The BT model can be obtained through the projected model (3.14)–(3.17) by setting $\lambda = 0$, $q_j^{t1} = q_j^{t2} = 0$, $j, t \in T : j < t$ and $q_t^{t1} = q_t^{t2}$.*

Although the number of constraints (3.15) and (3.16) in the projected model increases exponentially with the number n of time periods, most of these inequalities are redundant.

In fact, for each $k \in \{1, \dots, t\}$ such that $\sum_{j=1}^t |\pi_j^t| = k$, only one inequality (3.15) and one inequality (3.16) are non dominated for each $t \in T$. The projected model can be solved using a Benders decomposition approach together with a separation algorithm for the constraints (3.15) and (3.16) that can easily work by inspection. However, preliminary results reported in Section 4.1 show that many of such constraints need to be included.

The next proposition provides an efficient way to solve model \hat{D} when multipliers are fixed. Such result will be used in the next section to design efficient heuristics to find solutions with lower true cost.

Proposition 12. *For fixed multipliers λ , $\hat{D}(\lambda)$ is given as follows*

$$\begin{aligned} \hat{D}(\lambda) = \min_u \quad & \sum_{t=1}^n (c_t u_t + \theta_t) \\ \text{s.t.} \quad & \theta_t \geq h_t \left(x_1 + \sum_{j=1}^t (u_j - \mu_j) \right) + A_t^1(\lambda), \quad \forall t \in T, \\ & \theta_t \geq -b_t \left(x_1 + \sum_{j=1}^t (u_j - \mu_j) \right) + A_t^2(\lambda), \quad \forall t \in T, \\ & u_t \geq 0, \quad \forall t \in T, \end{aligned}$$

with

$$A_t^1(\lambda) = \sum_{\ell=1}^{\lfloor \Gamma_t \rfloor} \alpha_{j(\ell)}^t + (\Gamma_t - \lfloor \Gamma_t \rfloor) \alpha_{j(\lfloor \Gamma_t \rfloor)}^t \quad \text{and} \quad A_t^2(\lambda) = \sum_{\ell=1}^{\lfloor \Gamma_t \rfloor} \beta_{j(\ell)}^t + (\Gamma_t - \lfloor \Gamma_t \rfloor) \beta_{j(\lfloor \Gamma_t \rfloor)}^t,$$

where $\alpha_j^t = | -h_t \delta_j + \lambda_j^t |$ for $1 \leq j < t$, $\alpha_t^t = | -h_t \delta_t - \sum_{j=t+1}^n \lambda_j^t |$, $\beta_j^t = | b_t \delta_j + \lambda_j^t |$ for $1 \leq j < t$, and $\beta_t^t = | b_t \delta_t - \sum_{j=t+1}^n \lambda_j^t |$, where $\alpha_{j(\ell)}^t$ is the ℓ^{th} largest value among $\alpha_1^t, \dots, \alpha_t^t$ and $\beta_{j(\ell)}^t$ is the ℓ^{th} largest value among $\beta_1^t, \dots, \beta_t^t$.

The proof is a direct application of Proposition 7 so it will be omitted.

3.3 Heuristic schemes to improve the quality of solutions

Among all the models considered in this paper, model D , corresponding to the AARC approach, is the one that provides bounds closer to R^* . Another important concern is related with obtaining solutions \bar{u} such that $R(\bar{u})$ is close to R^* , that is, solutions with the best possible true cost. From a practical perspective, obtaining such solutions \bar{u} is more relevant than obtaining good bounds. Taking into account this more practical orientation, in this section, we develop iterative heuristic solution approaches, based on the interpretation of the Lagrangian multipliers as penalties associated with constraints violation, to obtain solutions with a lower true cost. With that purpose, for a given vector of multipliers, the value of the uncertain variables v_t^j , $1 \leq t \leq j \leq n$, must be computed at each iteration. Since such computation can easily be done by inspection in model \hat{D} but not in model D , we use model \hat{D} rather than model D . Besides, there are two more reasons to use model \hat{D} instead of model D . First, model \hat{D} is computationally easier to solve when the multipliers are fixed (see Proposition 12) and second, the results presented in the computational section for the instances solved to optimality suggest that there are no significant differences between the true cost of the solutions provided by both models \hat{D} and D .

Given that model \hat{D} (and also model D) is a pure linear model, one would expect to solve it to optimality even for large size instances. However, when other aspects are included, the

model can quickly become very large and the direct use of such model can be prohibitive. In order to take advantage of model \hat{D} we derive heuristic schemes that iteratively fix the value of the new variables (multipliers) leading to easier subproblems. The proposed heuristics are tested using the inventory problem when production setup costs are considered, that is, when the objective function is given by (3.1) and the set of constraints (3.2) is added.

3.3.1 Guided Iterated Local Search algorithm

The first heuristic approach that we propose is called Guided Iterated Local Search (GILS). The GILS heuristic can easily be used to solve other complex problems and it is inspired in the classical Iterated Local Search (ILS) heuristic based on the local branching scheme proposed by Fischetti and Lodi (2003). ILS heuristics have performed well in complex inventory problems with uncertainty, such as the Maritime Inventory Routing problem (Agra et al. 2016a, 2018a) and the Production Inventory problem (Agra et al. 2018b).

The main idea of the ILS heuristic is to restrict the search space of some integer variables (setup variables in our case) to a neighbourhood of a given solution. For a given positive integer parameter ρ , define the *neighborhood* $\mathcal{N}(\bar{y}, \rho)$ of \bar{y} as the set of feasible solutions of the model \hat{D} satisfying the additional local branching constraint (see Fischetti and Lodi (2003)):

$$\sum_{t \in T | \bar{y}_t = 0} y_t + \sum_{t \in T | \bar{y}_t = 1} (1 - y_t) \leq \rho. \quad (3.18)$$

The neighborhood $\mathcal{N}(\bar{y}, \rho)$ is the set of solutions that differ from the current solution \bar{y} by a maximum number of ρ values of the y_t variables. The linear constraint (3.18) limits to ρ the total number of binary variables y_t flipping their value with respect to the solution \bar{y} , either from 1 to 0 or from 0 to 1.

The GILS heuristic is a modified version of the ILS heuristic and can be seen as an improved version in which the search space is even more reduced through the inclusion of new constraints on the Lagrangian multipliers. Motivated by the fact that the Lagrangian multipliers are used to penalize the deviations between the copies of the uncertain variables of the adversarial problem, we impose, at each iteration, two types of constraints to guide the value of the multipliers as follows.

Type I: Constraint $\lambda_t^j \leq 0$ if $\bar{v}_t^t - \bar{v}_t^j < 0$ or constraint $\lambda_t^j \geq 0$ if $\bar{v}_t^t - \bar{v}_t^j > 0$.

Type II: Constraint $\lambda_t^j \leq \bar{\lambda}_t^j$ if $\bar{v}_t^t - \bar{v}_t^j < 0$ or constraint $\lambda_t^j \geq \bar{\lambda}_t^j$ if $\bar{v}_t^t - \bar{v}_t^j > 0$.

At each iteration, the current value of the uncertain variables v_t^j and the current value of the Lagrangian multipliers are denoted by \bar{v}_t^j and $\bar{\lambda}_t^j$, for all $1 \leq t \leq j \leq n$, respectively.

To start the GILS heuristic, an initial solution is required. Such solution can be found by solving the model \hat{D} and fixing the Lagrangian multipliers to their value in the linear relaxation of model \hat{D} . The full algorithm is described in Algorithm 1.

Algorithm 1 Guided Iterated Local Search

- 1: Solve the linear relaxation of model \hat{D}
 - 2: Solve the integer model \hat{D} , with the Lagrangian multipliers fixed to their values in the linear relaxation of model \hat{D}
 - 3: Save the solution \bar{y} obtained in the previous step
 - 4: **repeat**
 - 5: **for all** $t, j \in T$ such that $t \leq j$ **do**
 - 6: Compute the value of the uncertainty variables v_t^j and add either constraints of type I or of type II to the model according to a predefined rule
 - 7: **end for**
 - 8: Add constraint (3.18) to the model \hat{D} and solve it for γ seconds
 - 9: Update the solution \bar{y}
 - 10: Remove all the constraints added
 - 11: **until** the time limit of β seconds or a maximum number of iterations is reached
-

Steps 5 to 7 are used to guide the values of the Lagrangian multipliers as penalties for variable deviations. By ignoring Steps 5 to 7, Algorithm 1 becomes the classic ILS heuristic, that will be also tested in the computational section. In Step 6, several specific rules can be used to choose in each iteration the type of constraints added to the problem. Some of those rules will be discussed in the computational section. It is important to notice that the purpose of Steps 5 to 7 is not to accelerate the algorithm. Additionally, we may also expect to obtain worse bounds (based on the value of model \hat{D}) using the GILS heuristic than using the ILS heuristic since we are restricting the search space. By penalizing the differences between the copies of the uncertain variables, we aim to force the choice of a neighbor solution based on an estimation of the cost closer to the true one. With this technique we expect to obtain better quality solutions (with true cost close to the cost of the optimal solution).

3.3.2 Subgradient Optimization method

Since model \hat{D} is based on a Lagrangian relaxation, we adapt the subgradient method, frequently used to solve the dual problem of a Lagrangian relaxation, to solve model \hat{D} heuristically. The Subgradient Optimization (SO) method that we propose depends on two given *a priori* parameters, parameter It_{Lim} and parameter ϕ , and uses the following additional functions:

- $R(\bar{u})$: computes the true cost of a given production policy \bar{u} .
- $C_{deviations}(\bar{\lambda})$: given a vector $\bar{\lambda}$, computes the value \bar{v} of the deviation variables v .

The SO method starts by solving the linear relaxation of model \hat{D} to obtain the initial values for the Lagrangian multipliers λ . The optimal value of the linear relaxation is used to define a lower bound to the problem. In the loop (step 4 to step 28 of the Algorithm 2), model \hat{D} is solved with updated information and the corresponding bound as well as the true cost of the production policy are computed and compared with the current best values. The value of the Lagrangian multipliers is updated in steps 21 to 25 according to the interpretation of the multipliers as penalties associated with the violation of constraints (3.5), taking into account the value of the variables v_t^j and v_t^t .

At each iteration, model \hat{D} is solved with the Lagrangian multipliers fixed and all the remaining variables free, however, whenever a limit number of iteration (It_{Lim}) is reached

without a better bound or a better solution (solution with lower true cost) is obtained, the multipliers are left free and the setup variables are fixed. This strategy is used to escape from local minimums and hence explore new feasible regions of the search space.

Algorithm 2 Subgradient Optimization method

- 1: Initialization: $No_{Improve} := 0$, $Best_{bound} := \infty$, $Best_{value} := \infty$
 - 2: Solve the linear relaxation of model \hat{D} and store the multipliers $\bar{\lambda}$
 - 3: Set LB equal to the objective function value of the linear relaxation
 - 4: **repeat**
 - 5: **if** $No_{Improve} < It_{Lim}$ **then**
 - 6: Impose constraints $\lambda_t^j = \bar{\lambda}_t^j$ and make all the remaining variables free
 - 7: **else**
 - 8: Impose constraints $y_t = \bar{y}_t$ and make the Lagrangian multipliers free
 - 9: $No_{Improve} \leftarrow 0$
 - 10: **end if**
 - 11: Solve the integer model \hat{D} with the imposed constraints
 - 12: Set $Bound$ equal to the objective function value of model \hat{D}
 - 13: **if** $Bound < Best_{bound}$ **then**
 - 14: Update $Best_{bound}$
 - 15: $No_{Improve} \leftarrow 0$
 - 16: **end if**
 - 17: **if** $R(\bar{u}) < Best_{value}$ **then**
 - 18: Update $Best_{value}$
 - 19: $No_{Improve} \leftarrow 0$
 - 20: **end if**
 - 21: Compute \bar{v}_t^j , the value of the deviation variables v_t^j , for all $t, j \in T$ such that $t \leq j$ using function $C_{deviations}(\bar{\lambda})$
 - 22: Compute the subgradient $s_t^j := \bar{v}_t^t - \bar{v}_t^j$ for all $t, j \in T$ such that $t < j$
 - 23: Compute $norm := \sum_{t=1}^T \sum_{j=t+1}^T (s_t^j)^2$
 - 24: Define $step_{size} := \phi \frac{Bound-LB}{norm}$
 - 25: Update multipliers $\bar{\lambda}_t^j \leftarrow \bar{\lambda}_t^j + step_{size} \times s_t^j$
 - 26: Remove all the added constraints
 - 27: $No_{Improve} \leftarrow No_{Improve} + 1$
 - 28: **until** A time limit of β minutes is reached
-

4 Computational experiments

This section reports the computational experiments carried out to compare the BO approach, the BT approach, the Lagrangian Dual approach based on model \hat{D} (that is named by LD), and the approach based on model D . Since we have proved that this last approach coincides with the affinely adjustable robust counterpart approach, hereafter, it is denoted by AARC approach. A model equivalent to the AARC can be obtained by considering the dual reformulated model proposed in Bertsimas and de Ruiter (2016) solved through affine decision rules. However, preliminary results not reported here showed that it is not beneficial in our case to use such reformulation, since the computational times associated with this model are higher than the ones obtained by using the AARC model.

In Section 4.1 we report the results for medium size lot-sizing instances with 30 time periods,

for which all the optimal solutions can be obtained, while in Section 4.2 larger size instances with at most 100 time periods are considered.

Table 1 displays the total number of constraints and the total number of non integer variables of model D with 30 and 100 time periods. The reported results correspond to the cases where the setup costs are either considered or not (column *Setup*), and the cases where the Lagrangian multipliers are either free or fixed (column *#Multipliers*). In the column *#Variables*, the numbers in parenthesis indicate the total number of integer variables in model D associated with the use of setup costs. Notice that the number of constraints for model \hat{D} is exactly the same as for model D and the number of variables is approximately 2/3 of the number of variables in model D .

Table 1: Total number of variables and constraints of model D .

n	<i>Setup</i>	<i>#Multipliers</i>	<i>#Constraints</i>	<i>#Variables</i>
30	No	fix	2850	990
		free	2850	1425
	Yes	fix	2910	990 (+30)
		free	2910	1425 (+30)
100	No	fix	31100	10500
		free	31100	15550
	Yes	fix	31300	10500 (+100)
		free	31300	15550 (+100)

The computational experiments use instances generated as follows. For each time period $t \in T$, the nominal demand μ_t and the maximum allowed deviation δ_t are randomly generated in $[0, 50]$ and $[0, 0.2\mu_t]$, respectively. The maximum number of deviations in period t is computed using the relation $\Gamma_t = \Gamma_{t-1} + \tau$, with τ varying in $\{0, 1\}$ and Γ_0 is assumed to be zero. The initial stock level at the producer, x_1 , is randomly generated between 0 and 30 and the production capacity P_t is constant and equal to $\sum_{t=1}^n \mu_t$. The production, holding and backlog costs are the same as those used by Bertsimas and Thiele (2006), i.e., $c_t = 1$, $h_t = 4$, $b_t = 6$, respectively, for all $t \in T$. Throughout this section, we consider two variants of the robust inventory problem (with and without setup costs). The production setup costs occur in many practical inventory problems. However, the main goal of using instances with setup costs is to get harder instances, since the inclusion of integer setup variables results in a non linear model.

In order to compute the true cost $R(u)$ of a given solution u , preliminary tests were conducted using four approaches: the dynamic program proposed by Bienstock and Özbay (2008), the dynamic program proposed by Agra et al. (2016b), the mixed integer formulation with big-M constraints presented by Gorissen and den Hertog (2013), and the decomposition approach proposed by Bienstock and Özbay (2008). The dynamic program proposed by Bienstock and Özbay (2008) provided, in general, better results and solved all the adversarial problems in less than one second for instances with 100 time periods. Hereafter, for all the approaches considered in the computational experiments, the true cost of a solution is computed using the dynamic program proposed by Bienstock and Özbay (2008).

All tests were run using a computer with an Intel Core i7-4750HQ 2.00 GHz processor and 8 GB of RAM, and were conducted using the Xpress-Optimizer 28.01.04 solver with the default options.

4.1 Computational experiments for medium size instances

In this subsection all the reported results are based on instances with 30 time periods. Preliminary experiments on a set of 10 instances were conducted to compare the performance of model (3.6)–(3.13) against the projected model (3.14)–(3.17). The second model is solved through a Benders decomposition procedure, having a separation scheme for constraints (3.15) and (3.16). The average running time was 721 seconds and the required average number of iterations was 552. Using the model (3.6)–(3.13) the average running time was lower than 1 second. Note that model D could also be solved using the decomposition procedure proposed by Ardestani-Jaafari and Delage (2018). However preliminary experiments indicate that its performance is similar to the one observed when Benders decomposition is used to solve the projected model (3.14)–(3.17), since a large number of iterations is needed. Therefore, henceforward, we consider only model (3.6)–(3.13).

Now we analyse the impact of the setup cost in the presented approaches. Figures 1 to 4 report average results obtained for 16 different setup costs with values in $\{0, 10, \dots, 150\}$. For each setup cost, one hundred instances were randomly generated considering different samples of the nominal demand values. All obtained results are presented through their average values, therefore Mann-Whitney hypothesis tests are applied to find significant differences between the approaches. A significance level of 1% is used in all tests.

Figure 1 displays the average cost of the solutions obtained by the BO approach (optimal value) and the average objective function values corresponding to the LD, AARC and BT approaches (which are upper bounds for the value of the BO solution). The points marked with squares ($LD(u^{BT})$) represent the average cost of the solutions obtained by the LD approach for the production policy obtained by the BT approach, i.e., after obtaining the solution of the BT approach, the value of the production variables u_t , $t \in T$, is fixed and model \hat{D} is solved. The obtained results suggest that the BT approach is too conservative, since the quality of the

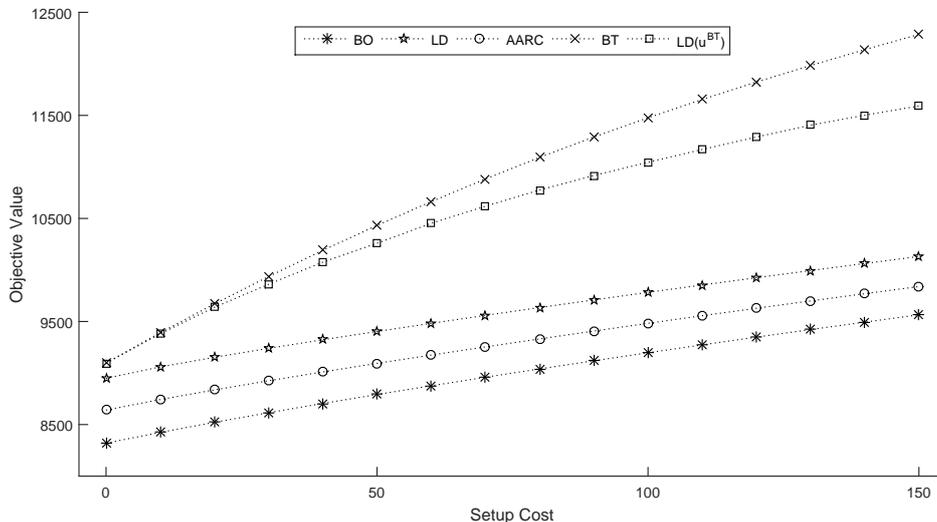


Figure 1: Cost of the solutions obtained by the different approaches in terms of the setup cost.

upper bound provided by this approach degrades rapidly as the setup cost increases. This is not the case of both LD and AARC approaches where the obtained upper bounds are close to the cost of the solution obtained by the BO approach, even when the setup cost increases. In fact, for large setup costs, the BT approach provides an optimal bound that is up to 28% larger than the true cost of the solutions provided by the BO approach while the gaps associated to both the LD and AARC approach are up to 6% and 3%, respectively, for all the setup costs

tested.

When comparing the displayed lines associated with $LD(u^{BT})$ and BT we observe that, in general, there is a gap (that is up to 6%) between the corresponding bounds. This means that the optimal value of the Lagrangian multipliers for the production policy obtained by the BT approach is usually different from zero (as considered in the BT approach). Hence, a better choice of the Lagrangian multipliers can be used to improve the quality of the upper bound provided by the BT approach.

A prevailing conclusion for all the setup costs tested is that the LD, the AARC and the BT approaches lead to solutions with average upper bounds significantly higher than the optimal value provided by the BO approach. Further, the average upper bounds obtained by the BT approach are significantly higher than the ones obtained by both the LD and the AARC approaches for setup costs greater than 10, and significantly greater than the ones obtained by the $LD(u^{BT})$ approach for high setup costs (greater than 110).

Figure 2 reports the average computational time in seconds required by each approach to find the solution, in terms of the setup cost.

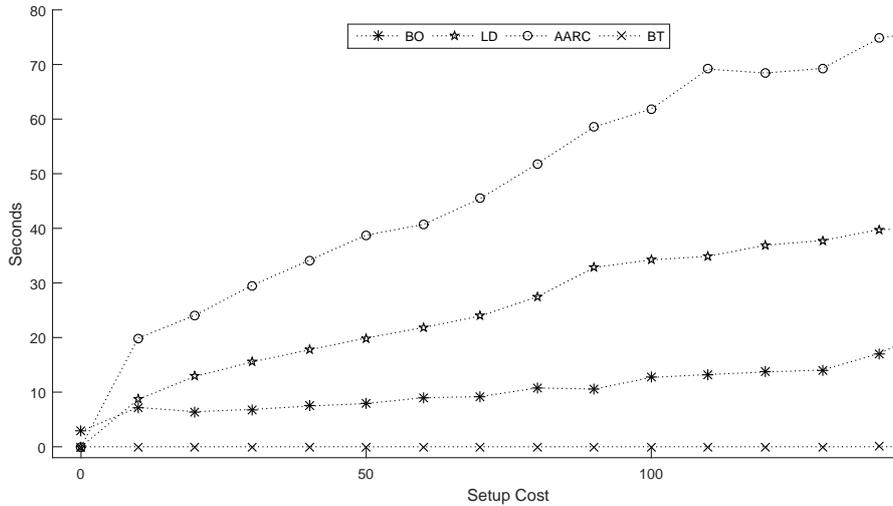


Figure 2: Average computational time associated to each approach in terms of the setup cost.

The computational time of the BT approach is always lower than one second. It can be observed that the exact BO approach is on average twice as faster as the LD approach. The computational time required by the AARC approach is approximately twice the computational time required by the LD approach. The average time required by the BO approach to solve each master problem ranges from 0 to 12 seconds while the computational time required to solve each adversarial problem is always lower than one second.

Figure 3 displays the average true cost of the production policy determined by the approaches LD ($R(u^{LD})$), AARC ($R(u^{AARC})$), BT ($R(u^{BT})$), and compare them with the cost of the optimal production policy obtained by the BO approach. Note that these values are not the upper bounds obtained by the LD, AARC and BT approaches directly. They are the true costs obtained by solving the adversarial problem for each solution obtained with the indicated approach. The behavior of the true cost of the production policy obtained by the LD, AARC and BT approaches resembles the trend observed for the upper bounds. However, when the setup costs are not considered, the true cost of the production policy obtained by the BT approach is, in general, lower than the one obtained by both the LD and AARC approaches. It is interesting to note that the true cost of the solutions determined by both the LD and AARC approaches are very close. In fact, the Mann-Whitney hypothesis tests reveal that, in

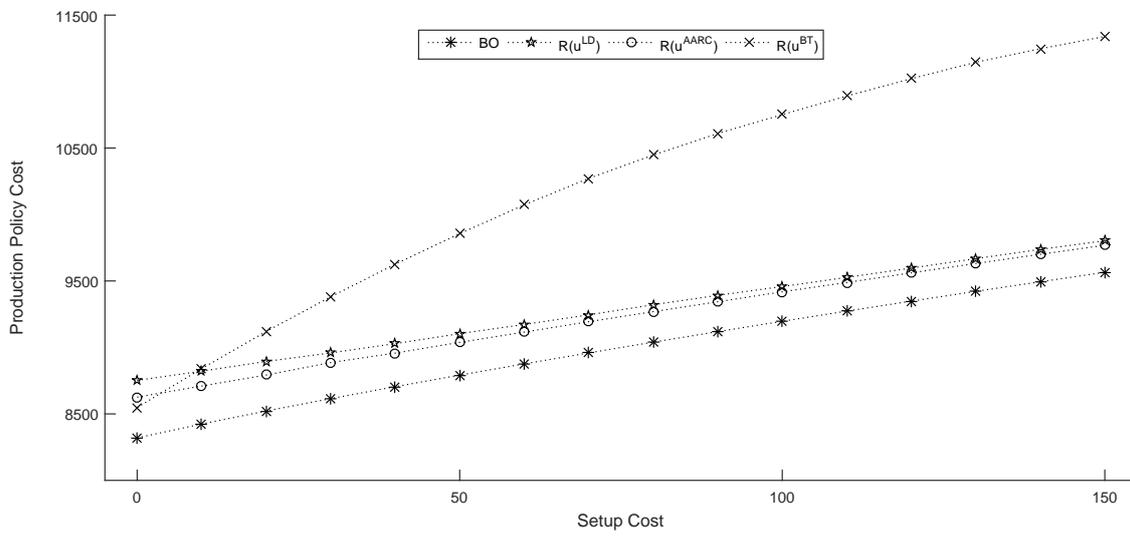


Figure 3: True average cost of the solution obtained by both the LD and the BT approaches compared with the cost of the optimal production policy obtained by the BO approach.

terms of the true cost of the production policy, the differences between both approaches are not significant. Moreover, the average true costs of the production policies determined by both LD and AARC approaches are not significantly different from the average costs of the optimal production policies. However, the average true cost of the production policies determined by the BT approach is significantly greater than the one determined by the LD and the AARC approaches for setup costs greater than 30.

A key conclusion from Figure 3 is that in the case where the setup costs are not considered, the true average cost from the solutions obtained using the BT approach may give a fair approximation on the optimal value. However, when setup costs are high the BT approach can give poor bounds and, beyond that, it can also produce bad solutions (with costs up to 16% larger than the optimal true costs). This may indicate that for more complex inventory problems the overestimation of costs obtained by the BT approach may lead to poor decisions.

Figure 4 displays the average number of production periods associated with the production policy determined by the BO, the LD, the AARC and the BT approaches.

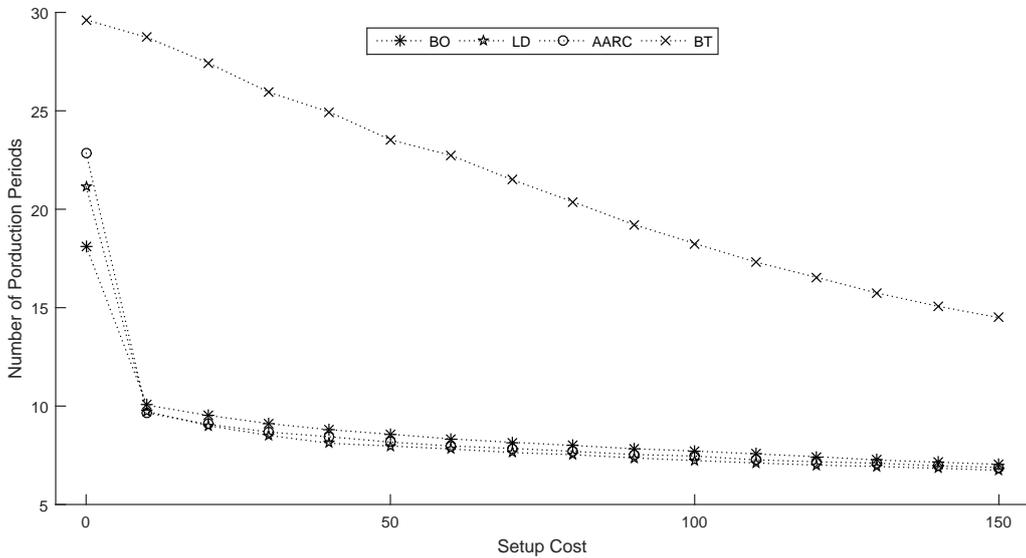


Figure 4: Average number of production periods in the production policies obtained by the different approaches in terms of the setup cost.

This figure can help to explain the results displayed in Figures 1 and 3, since the average numbers of production periods in the LD, AARC and BO approaches are similar. Notice that even when the setup cost is high, the number of production periods in the BT approach remains high, which may be justified by the fact that the BT approach tends to overestimate the contribution of the inventory costs in the objective function. The differences between the average number of production periods in the LD, AARC and BO approaches are not significant for any setup costs used while such differences between the BT and BO approaches are significant for all the setup cost tested.

We also analyse the performance of the LD, the AARC and the BT approaches regarding the maximum number of deviations Γ_n in the last time period. The obtained results are reported in Appendix D.

4.2 Computational experiments for large size instances

In this section we report the computational results for large size instances with up to 100 time periods. For these instances the exact BO approach cannot be solved to optimality within a reasonable time limit. Preliminary results showed that even for a small number of scenarios the master problem cannot be solved within eight hours. Similar difficulties were observed for a related lot-sizing problem in Attila et al. (2017). Furthermore, when the setup costs increase, model D (the one used in the AARC approach) becomes computationally harder to solve to optimality. For the instances with 100 time periods and setup costs greater than 70 we are not able to solve model D within a time limit of eight hours. Table 2 reports the average optimality gaps obtained with model D over a set of 10 instances with 100 time periods considering a time limit of two hours, for four different setup costs.

Table 2: Average optimality gaps obtained with model D with a time limit of two hours.

Setup Cost	50	150	450	750
Gap (%)	0.14	1.39	3.61	5.51

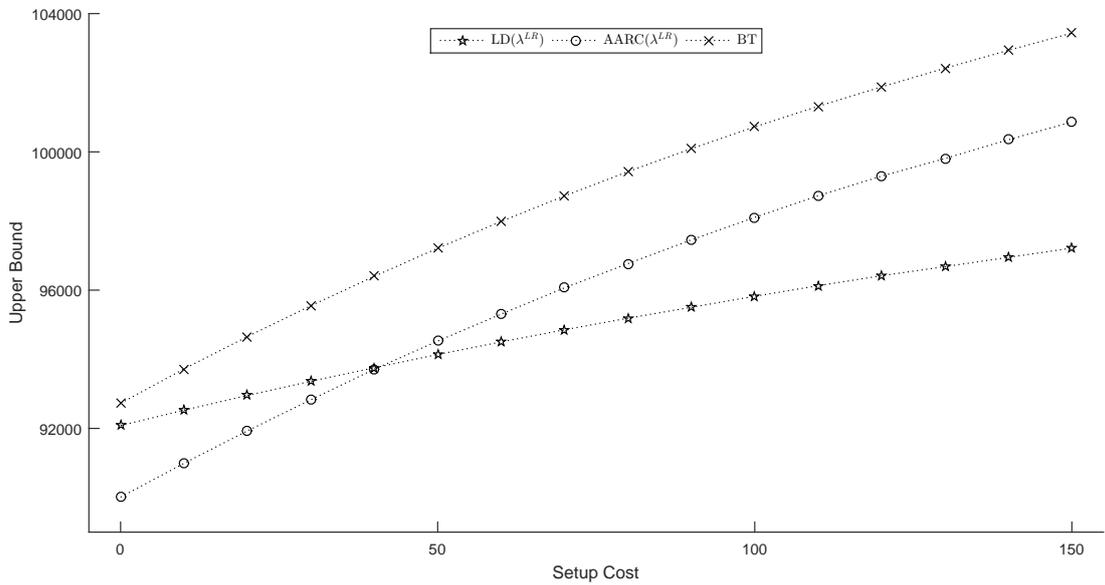


Figure 5: Average upper bound values obtained by the LD and AARC approaches considering the optimal Lagrangian multipliers of the linear relaxation of models \hat{D} and D , respectively, for different setup costs.

Table 2 shows that the instances become more difficult to solve when the setup cost increases. Results not reported here allow us to conclude that the optimal solution of the LD approach can be obtained in less than 2 hours for problem instances with up to 55 time periods while for the AARC approach only problem instances with up to 40 time periods can be solved to optimality within 2 hours. Hence, the main goal of this section is to test heuristic approaches that can be used on large size inventory models to obtain tight upper bounds as well as good solutions (with true cost close to the optimal value).

When the Lagrangian multipliers are fixed, models D and \hat{D} can be quickly solved, even if setup costs are considered. In particular, model \hat{D} with the multipliers fixed to zero, that corresponds to the BT approach, can be solved in less than 5 seconds. An initial value for the Lagrangian multipliers can easily be obtained by solving the linear relaxation of models D and \hat{D} , respectively. Figure 5 displays, for each setup cost in $\{0, 10, \dots, 150\}$, the average upper bound values over 100 randomly generated instances with 100 time periods obtained by both LD and AARC approaches when all the multipliers are fixed to their values in the optimal linear relaxations of models \hat{D} and D , respectively. The average upper bound values obtained by the BT approach are also displayed. Figure 5 shows opposite behavior of both LD and AARC approaches, when the Lagrangian multipliers are fixed to their values in the linear relaxation, comparing with the BT approach. While the gap between the lines associated with the AARC and the BT approaches tend to decrease as the setup cost increases, the gap between the lines associated with the LD and BT approaches tends to increase as the setup cost increases. The first gap varies between 2.6% and 3.0% while the second varies between 0.7% and 6.4%.

These results show that, when the value of the setup cost increases, tighter upper bounds can be obtained by considering the Lagrangian multipliers fixed to their values in the linear relaxation of model \hat{D} in the LD approach instead of considering all the multipliers equal to zero (as in the case of the BT approach). Furthermore, the difference between the computational time required to compute the upper bounds in both cases corresponds to the computational time required to solve the linear relaxation of model \hat{D} , which is always lower than 7 seconds for all the tested instances. This means that, in general, for large size instances, a better bound

than the one obtained by the BT approach can be quickly obtained by considering the optimal multipliers of the linear relaxation of model \hat{D} .

From the theoretical study we know that the upper bound corresponding to the optimal solution of the AARC approach is lower than or equal to the upper bound obtained by the LD approach. Moreover, the value of the linear relaxation is lower in the AARC approach than in the LD approach. Nevertheless, Figure 5 shows that when the multipliers are fixed to their value in the linear relaxation, the upper bounds provided by the LD approach tend to be better than the ones obtained with the AARC approach, when the value of the setup cost increases.

4.2.1 Evaluation of the proposed heuristics

In this section we analyse the performance of both the GILS heuristic and the SO method presented in Section 3.3. It is important to remind that these two heuristics were specifically designed to generate better solutions and not necessarily better bounds resulting from the objective function values of the considered models. As reference methods we use the heuristic that consists on solving the full model D with a time limit of one hour, and the ILS heuristic. The first heuristic will be called Full Model heuristic (FM heuristic).

Tuning of the parameters

We consider two variants of the ILS heuristic, one based on model \hat{D} and other based on model D , denoted by $ILS_{\hat{D}}$ and ILS_D , respectively. Both heuristics correspond to Algorithm 1 described in Section 3.3.1 without steps 5 to 8. However, instead of imposing a time limit or a maximum number of iterations, the algorithm stops when no improvement in the objective function value is observed. In both heuristics, the parameter ρ was set to 2 since with this parameter for the instances with 100 time periods, almost all the problems arising in each iteration of the ILS heuristics were solved to optimality in less than 150 seconds (the time limit imposed in each iteration).

For both the GILS heuristic and the SO method, a set of 20 randomly generated instances with 30 time periods was used to tune the values of the parameters. Since in the GILS heuristic we are imposing additional constraints on the Lagrangian multipliers, the value of the objective function in a given iteration can be worse than the one obtained in the previous iteration. So it does not make sense to stop the algorithm when there is no improvement in the objective function value. Hence, the stopping criteria for the GILS heuristic is defined using the number of iterations (that is limited to 15). Three rules were tested to choose the type of constraints added to the problem in each iteration: i) add only constraints of type I; ii) add only constraints of type II and iii) successively add constraints of type I k times and then add constraints of type II k times (with $k = 1, 2, 3$). Taking into account both the upper bounds and the true cost of the solutions, the best results were obtained when the third rule was used with $k = 2$, so this is the strategy used henceforward. To compare the GILS with both variants of the ILS heuristics we use the same time limit in each iteration (150 seconds) and also $\rho = 2$.

For the SO method, different values $\{0.25, 0.5, 1, 1.5, 2\}$ of ϕ and different values $\{5, 10, 15, 20\}$ of parameter It_{Lim} were tested. The best results were obtained when the values $\phi = 1$ and $It_{Lim} = 10$ were used. The time limit imposed in the SO method is 600 seconds.

Comparing upper bounds and true costs

Here we compare the performance of the heuristics in terms of the setup cost (for instances with 100 time periods) and also in terms of the number of periods (for instances with a setup cost equal to 150). Tables 3 and 4 present the average upper bounds obtained for each heuristic tested as well as the corresponding average computational time in seconds. Each line of the

tables reports average values obtained for a set of 10 instances. The best average upper bounds obtained for each set of instances are marked in bold. Furthermore, the numbers in parenthesis next to the bounds indicate the number of best bounds obtained by the corresponding heuristic.

Table 3: Average upper bounds obtained for the heuristics tested for instances with 100 time periods and setup costs varying from 25 to 150.

Setup Cost	FM		ILS _D		ILS _{\hat{D}}		GILS		SO	
	Bound	Sec.	Bound	Sec.	Bound	Sec.	Bound	Sec.	Bound	Sec.
25	92365 ⁽²⁾	3600	92338 ⁽⁸⁾	1364	94583	655	94620	525	94710	600
50	92756 ⁽⁵⁾	3600	92776 ⁽⁵⁾	1264	94943	762	95001	718	95097	600
75	93178 ⁽²⁾	3600	93120 ⁽⁸⁾	1206	95269	781	95365	801	95889	600
100	93474 ⁽⁴⁾	3600	93445 ⁽⁶⁾	1295	95594	836	95770	917	95893	600
125	93733 ⁽⁷⁾	3600	93789 ⁽³⁾	1011	95891	736	96080	1051	96188	600
150	94155 ⁽⁵⁾	3600	94131 ⁽⁵⁾	1032	96183	786	96286	1089	96592	600

Table 4: Average upper bounds obtained for the heuristics tested for instances with setup cost equal to 150 and time periods varying from 20 to 100.

n	FM		ILS _D		ILS _{\hat{D}}		GILS		SO	
	Bound	Sec.	Bound	Sec.	Bound	Sec.	Bound	Sec.	Bound	Sec.
20	5380 ⁽¹⁰⁾	18	5385 ⁽⁶⁾	8	5526 ⁽¹⁾	4	5539 ⁽¹⁾	13	5524 ⁽¹⁾	600
40	16673 ⁽⁹⁾	2309	16728 ⁽³⁾	58	17259	23	17294	52	17231	600
60	37273 ⁽⁴⁾	3600	37287 ⁽⁶⁾	192	38191	116	38258	188	38246	600
80	59383 ⁽¹⁾	3600	59261 ⁽⁹⁾	644	61425	425	61606	774	61548	600
100	94155 ⁽⁵⁾	3600	94131 ⁽⁵⁾	1032	96183	786	96286	1089	96592	600

The results presented in Tables 3 and 4 reveal that best upper bounds are obtained with both the FM and the ILS_D heuristics. These results agree with what was stated in the theoretical study since the best upper bounds are obtained by the heuristics based on model D . All the instances with 20 time periods and almost all the instances with 40 time periods are solved to optimality by the FM heuristic. This justifies that the best results for the instances with these time periods are obtained with the FM heuristic. However, when the number of periods increases, best upper bounds are in general obtained by the ILS_D heuristic.

In Tables 5 and 6 we compare the heuristics in terms of the true cost of the obtained solutions. As in the previous tables, the best average results are marked in bold and the number of best solutions (with the best true cost) appears in parenthesis.

At each iteration of the ILS heuristics, GILS heuristic and SO method, the true cost of the current solution is obtained and the best obtained value is reported. In the FM heuristic the true cost of all integer solutions found during the Branch-and-Bound process is computed and the best true cost is reported.

Table 5: Average true cost of the solutions obtained for the heuristics tested for instances with 100 time periods and setup costs varying from 25 to 150.

Setup Cost	FM	ILS _D	ILS _{\hat{D}}	GILS	SO
25	91875	92016	92445	92343	90729 ⁽¹⁰⁾
50	92223 ⁽²⁾	92428	92732	92582	91004 ⁽⁸⁾
75	92567 ⁽¹⁾	92736	93012	92666	91320 ⁽⁹⁾
100	92890	93089	93299	92768	91223 ⁽¹⁰⁾
125	93147	93414	93520	92929	91607 ⁽¹⁰⁾
150	93479	93725	93829	93203	91840 ⁽¹⁰⁾

Table 6: Average true cost of the solutions obtained for the heuristics tested for instances with setup cost equal to 150 and time periods varying from 20 to 100.

n	FM	ILS _{\mathcal{D}}	ILS _{$\hat{\mathcal{D}}$}	GILS	SO
20	5319 ⁽³⁾	5353 ⁽¹⁾	5359 ⁽¹⁾	5354 ⁽¹⁾	5301 ⁽⁵⁾
40	16534 ⁽¹⁾	16619	16661 ⁽¹⁾	16521	16155 ⁽⁹⁾
60	36988	37118	37210	36971	36066 ⁽¹⁰⁾
80	59061	59060	59518	59038	57728 ⁽¹⁰⁾
100	93479	93725	93829	93203	91840 ⁽¹⁰⁾

The results presented in Tables 5 and 6 clearly suggest that the best average true costs are in general obtained by the SO method. Only for 9 out of the 110 instances presented in these two tables the best solutions were not found by the SO method. Furthermore, the computational time of the SO method (600 seconds) is much lower than the one required by the remaining heuristics. The SO method allows us to obtain solutions with true costs that are, on average, 1.8% lower than the ones obtained with the FM heuristic (which is the heuristic closer to the AARC approach). Hence, among all the heuristic solutions tested, the SO method is the most efficient heuristic to obtain good solutions (with low true costs).

As expected, the upper bound values obtained by both ILS heuristics are better than the ones obtained by the GILS heuristic. However, in terms of the true cost of the obtained production policies, the best results are in general obtained using the GILS heuristic. In fact, among all the 60 instances with 100 time periods considered in Table 5, 45 of the best solutions were found by the GILS heuristic while 9 and 6 were found by the ILS _{\mathcal{D}} and the ILS _{$\hat{\mathcal{D}}$} heuristics, respectively. Among all the 50 instances with a setup cost equal to 150 considered in Table 6, 38 best solutions were found by the GILS heuristic while 9 and 3 were found by the ILS _{\mathcal{D}} and the ILS _{$\hat{\mathcal{D}}$} heuristics, respectively.

Looking deeper to the SO method

Since the true cost of the solutions obtained with the BT approach is much higher than those obtained by all the heuristics tested, such results were not reported the Tables 5 and 6. However, in Table 7 we report some gaps showing the improvements on the true cost of the solutions obtained by the SO method compared to the true cost of the solutions obtained by the BT approach. Columns 2 to 7 refer to the instances presented in Table 5, those with 100 time periods and setup costs varying between 25 and 150, while columns 8 to 12 refer to the instances presented in Table 6, the ones with a setup cost equal to 100 and time periods varying between 20 and 100. Remember that the SO method starts from the solution obtained with model \hat{D} with the multipliers fixed to their value in the linear relaxation of such model. Hence, the line *Initial Solution* reports the average gaps associated with the true cost of the initial solution used in the SO method comparing with the true cost of the solution obtained by the BT approach. The line *Best Solution* reports the average gaps associated with the true cost of the best solution found by the SO method comparing with the true cost of the solution obtained by the BT approach.

Table 7: Average gaps (in percentage) between the SO method and the BT approach in terms of the true cost of the solutions.

Setup Cost/Time periods	Instances from Table 5						Instances from Table 6				
	25	50	75	100	125	150	20	40	60	80	100
Initial Solution	0.3	1.2	2.2	2.8	3.3	3.8	7.3	5.1	4.5	4.2	3.8
Best Solution	2.4	3.8	5.1	6.5	7.3	7.9	18.4	15.4	11.8	9.9	7.9

We observe in Table 7 that the gap between the SO method and the BT approach, in terms of the true cost of the solutions, increases as the setup cost increases and decreases as the number of periods increases. For the hardest instances, the ones with 100 time periods and setup cost equal to 150, the BT approach provides solutions with true costs that are 7.9% larger than those obtained by the SO method.

Finally, in order to compare the quality of the solutions generated by the SO method with those resulting from the AARC method solved to optimality, we report in Table 8 the average optimality gaps associated with both the SO method and the AARC approach in terms of the true cost of the solutions, for instances with $n = \{10, 20, 30, 40\}$ time periods (those instances where the AARC method can be solved to optimality within reasonable amount of time). For each number n , 25 instances were used. The numbers in parenthesis next to the gaps indicate the number of best solutions obtained by the corresponding method. The average gaps, in percentage, were computed according to the formula:

$$Gap = \frac{R(u^J) - R(u^*)}{R(u^*)} \times 100,$$

where u^* is the optimal solution (obtained by the BO approach) and u^J is the solution obtained by approach J , with $J = AARC$ or $J = SO$. Table 8 suggests that for the instances solved to optimality the best solutions are on average obtained by the SO method since the gaps associated with this approach are lower than the ones associated with the AARC.

Table 8: Average optimality gaps associated to both the SO method and the AARC approach.

n	10	20	30	40
SO	0.46 ⁽¹⁵⁾	0.68 ⁽²³⁾	1.07 ⁽¹⁸⁾	1.03 ⁽²³⁾
AARC	0.80 ⁽¹⁰⁾	1.43 ⁽²⁾	2.06 ⁽⁷⁾	1.82 ⁽²⁾

Furthermore, the number of best solutions found is greater in the SO method than in the AARC approach.

5 Conclusion

In this paper we consider RO min-max problems with decomposable functions. Based on the dual Lagrangian problem resulting from a Lagrangian relaxation of the reformulation of the adversarial problem, we provide a compact formulation to approximate the true min-max problem and show that the Bertsimas and Thiele dualization approach is a particular case of this approach with the multipliers equal to zero. Additionally, we show that the new dual Lagrangian formulation coincides with an affine approximation.

The theoretical results are applied to the robust inventory problem where the demands are uncertain and the uncertain variables belong to the $B\&T$ budgeted set. Computational results have shown that when other complicating aspects such as setup costs are present, by overestimating the costs, the classical dualization approach from Bertsimas and Thiele (2006) can provide poor bounds and poor solutions. The dual Lagrangian formulation, which coincides with an affine approximation model, leads to bounds closer to the true min-max value even for those instances where the dualization from Bertsimas and Thiele (2006) provide worst bounds. However, although the dual Lagrangian formulation leads to tractable models, their size can be too large to be solved to optimality for real size instances. Taking advantage of regarding such models from the perspective of Lagrangian duality theory, we propose heuristics approaches that

consider the new multipliers as penalties for violation of the constraints of the adversarial problem. Thus, such penalties penalize the overestimation of the true cost of each feasible solution. Using such idea, we introduce a Guided Iterated Local Search heuristic and a Subgradient Optimization method to solve large size inventory models. The Subgradient Optimization method proved to be efficient to obtain better solutions than those obtained using other approximation approaches including the dual Lagrangian formulation.

ACKNOWLEDGMENT

The research of the first three authors is supported by the *Center for Research and Development in Mathematics and Applications (CIDMA)* through the Portuguese Foundation for Science and Technology (FCT - Fundação para a Ciência e a Tecnologia), references UIDB/04106/2020 and UIDP/04106/2020. The research of the first author has been supported by a research fellowship (grant PD/BD/114185/2016) through FCT.

References

- Agra A, Christiansen M, Hvattum LM, Rodrigues F (2016a) A MIP based local search heuristic for a stochastic maritime inventory routing problem. Paias A, Ruthmair M, Voß S, eds., *Lecture Notes in Computer Science, Computational Logistics*, volume 9855, (Springer International Publishing), 18–34.
- Agra A, Christiansen M, Hvattum LM, Rodrigues F (2018a) Robust optimization for a maritime inventory routing problem. *Transportation Science* 52:509–525.
- Agra A, Requejo C, Rodrigues F (2018b) An adjustable sample average approximation algorithm for the stochastic production-inventory-routing problem. *Networks* 72(1):5–24.
- Agra A, Santos MC, Nace D, Poss M (2016b) A dynamic programming approach for a class of robust optimization problems. *SIAM Journal on Optimization* 26(3):1799–1823.
- Ardestani-Jaafari A, Delage E (2016) Robust optimization of sums of piecewise linear functions with application to inventory problems. *Operations Research* 64(2):474–494.
- Ardestani-Jaafari A, Delage E (2018) The value of flexibility in robust location transportation problems. *Transportation Science* 52(1):189–209.
- Attila ÖN, Agra A, Akartunah K, Arulselvan A (2017) A decomposition algorithm for robust lot sizing problem with remanufacturing option. *Lecture Notes in Computer Science*, 684–695 (Springer).
- Ben-Tal A, Ghaoui LE, Nemirovski A (2009) *Robust optimization*. Princeton Series in Applied Mathematics (Princeton University Press).
- Ben-Tal A, Goryashko A, Guslitzer E, Nemirovski A (2004) Adjustable robust solutions of uncertain linear programs. *Mathematical Programming* 99(2):351–376.
- Ben-Tal A, Hertog D, Vial JP (2015) Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical Programming* 149:265–299.
- Ben-Tal A, Nemirovski A (1999) Robust solutions of uncertain linear programs. *Operations Research Letters* 25(1):1–13.
- Bertsimas D, Brown D, Caramanis C (2011) Theory and applications of robust optimization. *SIAM Review* 53:464–501.
- Bertsimas D, de Ruiter FJCT (2016) Duality in two-stage adaptive linear optimization: Faster computation and stronger bounds. *INFORMS Journal on Computing* 28(3):500–511.
- Bertsimas D, Goyal V (2012) On the power and limitations of affine policies in two-stage adaptive optimization. *Mathematical Programming* 134(2):491–531.
- Bertsimas D, Sim M (2003) Robust discrete optimization and network flows. *Mathematical Programming* 98:49–71.

- Bertsimas D, Sim M (2004) The price of robustness. *Operations Research* 52:35–53.
- Bertsimas D, Thiele A (2006) A robust optimization approach to inventory theory. *Operations Research* 54:150–168.
- Bienstock D, Özbay N (2008) Computing robust basestock levels. *Discrete Optimization* 5:389–414.
- Chen X, Zhang Y (2009) Uncertain linear programs: Extended affinely adjustable robust counterparts. *Operations Research* 57(6):1469–1482.
- Delage E, Gianoli LG, Sansob B (2018) A practicable robust counterpart formulation for decomposable functions: A network congestion case study. *Operations Research* 66(2):535–567.
- El-Ghaoui L, Lebret H (1997) Robust solutions to least-square problems with uncertain data matrices. *SIAM Journal on Matrix Analysis and Applications* 18:1035–1064.
- El Housni O, Goyal V (2018) Piecewise static policies for two-stage adjustable robust linear optimization. *Mathematical Programming* 169(2):649–665.
- Fischetti M, Lodi A (2003) Local branching. *Mathematical Programming* 98(1–3):23–47.
- Georghiou A, Tsoukalas A, Wiesemann W (2019) A primal-dual lifting scheme for two-stage robust optimization Available at Optimization Online.
- Gorissen B, den Hertog D (2013) Robust counterparts of inequalities containing sums of maxima of linear functions. *European Journal of Operational Research* 227(1):30–43.
- Iancu D, Sharma M, Sviridenko M (2013) Supermodularity and affine policies in dynamic robust optimization. *Operations Research* 61(4):941–956.
- Kuhn D, Wiesemann W, Georghiou A (2011) Primal and dual linear decision rules in stochastic and robust optimization. *Mathematical Programming* 130(1):177–209.
- Marandi A, den Hertog D (2018) When are static and adjustable robust optimization problems with constraint-wise uncertainty equivalent? *Mathematical Programming* 170(2):555–568.
- Pochet Y, Wolsey LA (2006) *Production Planning by Mixed Integer Programming* (Springer, New York).
- Roos E, den Hertog D (2017) A less conservative variant of robust optimization. *Optimization Online* .
- Shapiro A (2001) *On Duality Theory of Conic Linear Problems*, 135–165 (Boston, MA: Springer US).
- Solyali O, Cordeau J, Laporte G (2012) Robust inventory routing under demand uncertainty. *Transportation Science* 46(3):327–340.
- Soyster A (1973) Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research* 21:1154–1157.
- Wei C, Li Y, Cai X (2011) Robust optimal policies of production and inventory with uncertain returns and demand. *International Journal of Production Economics* 134(2):357–367.
- Yanikoglu I, Gorissen BL, den Hertog D (2018) A survey of adjustable robust optimization. *European Journal of Operational Research* To appear.
- Zeng B, Zhao L (2013) Solving two-stage robust optimization problems using a column-and-constraint generation method. *Operations Research Letters* 41(5):457–461.
- Zhen J, den Hertog D, Sim M (2018) Adjustable robust optimization via fourier-motzkin elimination. *Operations Research* 66(4):1086–1100.

Appendices

Appendix A. Proof of Theorem 3

Here we provide the proof of Theorem 3 presented in Section 2.

Proof. Proof. To obtain problem (2.4), we focus on reformulating each of the robust constraints indexed by $t \in T$:

$$\theta_t \geq \max_{k=1, \dots, K} h_{tk}(u, \zeta^t) - (\lambda^t)^\top \zeta^t, \quad \forall \zeta^t \in \Omega^t,$$

where we replaced $f_t(u, \xi) := \max_{k=1, \dots, K} h_{tk}(u, \xi)$. One can easily verify that this constraint is equivalent to:

$$\theta_t \geq h_{tk}(u, \zeta^t) - (\lambda^t)^\top \zeta^t, \quad \forall \zeta^t \in \Omega^t, k = 1, \dots, K.$$

Based on the work of Ben-Tal et al. (2015), we derive the Fenchel robust counterpart of this robust constraint to obtain the equivalent reformulation:

$$\theta_t \geq \delta^*(v_{tk} | \Omega^t) - \sup_{\zeta^t} \{v_{tk}^\top \zeta^t - h_{tk}(u, \zeta^t) + (\lambda^t)^\top \zeta^t\}, \quad \forall k = 1, \dots, K,$$

where for all $t \in T, k = 1, \dots, K, v_{tk} \in \mathbb{R}^n$, and which can be simplified to

$$\theta_t \geq \delta^*(v_{tk} | \Omega^t) - h_{tk^*}(u, v_{tk} + \lambda^t), \quad \forall k = 1, \dots, K.$$

□

Appendix B. Projected Model

Here we present the projected version of model D mention in Section 2.3 as well as the steps followed in its derivation.

Proposition 13. *Projecting out variables $r_j^{tk}, j, t \in T : j \leq t, k \in K$ model D can be written as follows.*

$$\begin{aligned} D = \min_{u, \lambda, \theta, q, r} & g(u) + \sum_{t=1}^n \theta_t \\ \text{s.t. } & \theta_t \geq L_t^k(u) + \sum_{j=1}^t q_j^{tk} \Gamma_j - \sum_{j=1}^t |\pi_j^t| \sum_{\ell=j}^t q_\ell^{tk} + \sum_{j=1}^t \pi_j^t (a_j^{tk} - \lambda_j^t), \\ & \forall \pi_i^t \in \{-1, 0, 1\}, t \in T, k \in K, \\ & q_j^{tk} \geq 0, \quad \forall j, t \in T : j \leq t, k \in K. \end{aligned} \quad (5.1)$$

Proof. Proof. Using Fourier-Motzkin elimination, we first project out variables r_t^{tk} . For each $k \in K$, we have from (2.13) and (2.14),

$$\begin{aligned} r_t^{tk} &\geq -q_t^{tk} + a_t^{tk} - \lambda_t^t, \\ r_t^{tk} &\geq -q_t^{tk} - a_t^{tk} + \lambda_t^t, \\ r_t^{tk} &\geq 0. \end{aligned}$$

Combining (2.12) with these inequalities we obtain

$$\theta_t \geq L_t^k(u) + \sum_{j=1}^t q_j^{tk} \Gamma_j + \sum_{j=1}^{t-1} r_j^{tk} - |\pi_t^t| q_t^{tk} + \pi_t^t (a_t^{tk} - \lambda_t^t), \quad \forall \pi_t^t \in \{-1, 0, 1\}, t \in T, k \in K.$$

By iteratively eliminating r_j^{tk} from $j = t - 1$ until $j = 1$ and by using (2.13) and (2.14) we obtain (5.1).

□

Appendix C. Proof of Theorem 9

Here we provide the proof of Theorem 9 presented in Section 3.2.

Proof. Proof.

We start by writing model $R(u)$ with the new variables v_j^t , for $j, t \in T : j \leq t$ as follows:

$$\begin{aligned}
R(u) = \max_{x,v} \quad & \sum_{t=1}^n (c_t u_t + \max\{h_t x_{t+1}, -b_t x_{t+1}\}) \\
s.t. \quad & x_{t+1} = x_1 + \sum_{j=1}^t (u_j - \mu_j - \delta_j v_j^t), \quad \forall t \in T, \\
& \sum_{j=1}^{\ell} |v_j^t| \leq \Gamma_{\ell}, \quad \forall \ell, t \in T : \ell \leq t, \\
& v_t^j = v_t^t, \quad \forall t, j \in T : t < j, \\
& v_t^j \in [-1, 1], \quad \forall t, j \in T : t \leq j.
\end{aligned} \tag{5.2}$$

Following the process described in Section 2, we attach a Lagrangian multiplier λ_t^j to each constraint (5.2) for $1 \leq t < j \leq n$, and dualize these constraints in the usual Lagrangian way. This leads to the following relaxed problem

$$\begin{aligned}
LR(u, \lambda) = \max_{x,v} \quad & \sum_{t=1}^n \left(c_t u_t + \max\{h_t x_{t+1}, -b_t x_{t+1}\} - \sum_{j=t+1}^n \lambda_t^j (v_t^t - v_t^j) \right) \\
s.t. \quad & x_{t+1} = x_1 + \sum_{j=1}^t (u_j - \mu_j - \delta_j v_j^t), \quad \forall t \in T, \\
& \sum_{j=1}^{\ell} |v_j^t| \leq \Gamma_{\ell}, \quad \forall \ell, t \in T : \ell \leq t, \\
& v_t^j \in [-1, 1], \quad \forall t, j \in T : t \leq j.
\end{aligned}$$

Rearranging the terms in the objective function by noticing that $\sum_{t=1}^n \sum_{j=t+1}^n \lambda_t^j v_t^j = \sum_{t=2}^n \sum_{j=1}^{t-1} \lambda_j^t v_j^t$, and eliminating variables $x_t, t > 1$, the relaxed problem can be written as follows

$$\begin{aligned}
LR(u, \lambda) = \max_v \quad & \sum_{t=1}^n \left(c_t u_t + \max \left\{ L_t^1(u) - h_t \sum_{j=1}^t \delta_j v_j^t, L_t^2(u) + b_t \sum_{j=1}^t \delta_j v_j^t \right\} - v_t^t \sum_{j=t+1}^n \lambda_t^j + \sum_{j=1}^{t-1} \lambda_j^t v_j^t \right) \\
s.t. \quad & \sum_{j=1}^{\ell} |v_j^t| \leq \Gamma_{\ell}, \quad \forall \ell, t \in T : \ell \leq t, \\
& v_j^t \in [-1, 1], \quad \forall j, t \in T : j \leq t.
\end{aligned}$$

where $L_t^1(u) = h_t \left(x_1 + \sum_{j=1}^t (u_j - \mu_j) \right)$ and $L_t^2(u) = -b_t \left(x_1 + \sum_{j=1}^t (u_j - \mu_j) \right)$.

For a given u and λ , the problem $LR(u, \lambda)$ can be separated into n independent subproblems, one for each time period. Hence, for a fixed time period $t \in T$, the corresponding subproblem

can be written as follows

$$\begin{aligned}
LR_t(u, \lambda^t) &= \min_{\theta_t} \quad c_t u_t + \theta_t \\
s.t. \quad \theta_t &\geq L_t^1(u) + \max_{v^t \in \Omega^t} \left\{ -h_t \sum_{j=1}^t \delta_j v_j^t - v_t^t \sum_{j=t+1}^n \lambda_j^t + \sum_{j=1}^{t-1} \lambda_j^t v_j^t \right\}, \\
\theta_t &\geq L_t^2(u) + \max_{v^t \in \Omega^t} \left\{ b_t \sum_{j=1}^t \delta_j v_j^t - v_t^t \sum_{j=t+1}^n \lambda_j^t + \sum_{j=1}^{t-1} \lambda_j^t v_j^t \right\},
\end{aligned}$$

where $\Omega^t = \{v^t \in [-1, 1]^t \mid \sum_{\ell=1}^j |v_\ell^t| \leq \Gamma_j, 1 \leq j \leq t\}$.

Linearizing $LR_t(u, \lambda^t)$ by writing variables v_j^t as $v_j^t = v_j^{t+} - v_j^{t-}$, for all $j, t \in T, j \leq t$ and rearranging the terms in the set of constraints, model $LR_t(u, \lambda^t)$ becomes

$$\begin{aligned}
LR_t(u, \lambda^t) &= \min_{\theta_t} \quad c_t u_t + \theta_t \\
s.t. \quad \theta_t &\geq L_t^1(u) + h_t \max_{(v_j^{t+}, v_j^{t-}) \in \bar{\Omega}^t} \left\{ - \left(\delta_t + \sum_{j=t+1}^n \frac{\lambda_j^t}{h_t} \right) (v_t^{t+} - v_t^{t-}) + \sum_{j=1}^{t-1} \left(\frac{\lambda_j^t}{h_t} - \delta_j \right) (v_j^{t+} - v_j^{t-}) \right\}, \\
\theta_t &\geq L_t^2(u) + b_t \max_{(v_j^{t+}, v_j^{t-}) \in \bar{\Omega}^t} \left\{ \left(\delta_t - \sum_{j=t+1}^n \frac{\lambda_j^t}{b_t} \right) (v_t^{t+} - v_t^{t-}) + \sum_{j=1}^{t-1} \left(\frac{\lambda_j^t}{b_t} + \delta_j \right) (v_j^{t+} - v_j^{t-}) \right\},
\end{aligned} \tag{5.3}$$

$$\theta_t \geq L_t^2(u) + b_t \max_{(v_j^{t+}, v_j^{t-}) \in \bar{\Omega}^t} \left\{ \left(\delta_t - \sum_{j=t+1}^n \frac{\lambda_j^t}{b_t} \right) (v_t^{t+} - v_t^{t-}) + \sum_{j=1}^{t-1} \left(\frac{\lambda_j^t}{b_t} + \delta_j \right) (v_j^{t+} - v_j^{t-}) \right\}, \tag{5.4}$$

where

$$\bar{\Omega}^t := \left\{ (v_j^{t+}, v_j^{t-}) \in \mathbb{R}^t \times \mathbb{R}^t \mid \sum_{\ell=1}^j (v_\ell^{t+} + v_\ell^{t-}) \leq \Gamma_j; v_j^{t+} + v_j^{t-} \leq 1; v_j^{t+}, v_j^{t-} \geq 0; 1 \leq j \leq t \right\}.$$

Associating the dual variables q_j^{t1} and r_j^{t1} to the constraints of the inner problem in the RHS of constraints (5.3) and the dual variables q_j^{t2} and r_j^{t2} to the constraints of the inner problem in the RHS of constraints (5.4) we obtain model D . □

Appendix D. Computational results with respect to Γ_n

Here we analyse the performance of the LD, the AARC and the BT approaches regarding the maximum number of deviations Γ_n in the last time period (see section 4.1). The maximum number of deviations in the previous periods is taken as follows. Given a value $k \in \{1, \dots, \Gamma_n\}$, the time periods $t \in T$ such that $\Gamma_t = k$ are the ones in the range

$$t = q(k-1) + \sum_{\ell=1}^{k-1} \alpha_\ell + 1, \dots, qk + \sum_{\ell=1}^k \alpha_\ell,$$

where $q = \lfloor \frac{n}{\Gamma_n} \rfloor$, and $\alpha_\ell = 1$ if $\ell \leq n - q\Gamma_n$ and $\alpha_\ell = 0$, otherwise. This rule ensures that given two values $k_1, k_2 \in \{1, \dots, \Gamma_n\}$ the difference between the number of periods having at most k_1 and k_2 deviations is either zero or one.

Figures 6 and 7 display the results for the case where Γ_{30} ranges from 0 (nominal case) to 30 (box-constrained case). For each value of Γ_{30} , we consider 100 randomly generated instances and the average gap

$$Gap = \frac{UB - BO}{BO} \times 100$$

is displayed, where UB is a given upper bound and BO is the optimal value obtained by using the BO approach. The lines associated with LD , $AARC$ and BT represent the average gap corresponding to the upper bounds obtained by the LD, AARC and BT approaches, respectively. The lines associated with $R(u^{LD})$, $R(u^{AARC})$ and $R(u^{BT})$ represent the average gap corresponding to the true cost of the solutions obtained by the LD, AARC and the BT approaches, respectively. In Figure 6 the setup costs are not considered, while in Figure 7 a setup cost of value 150 is considered.

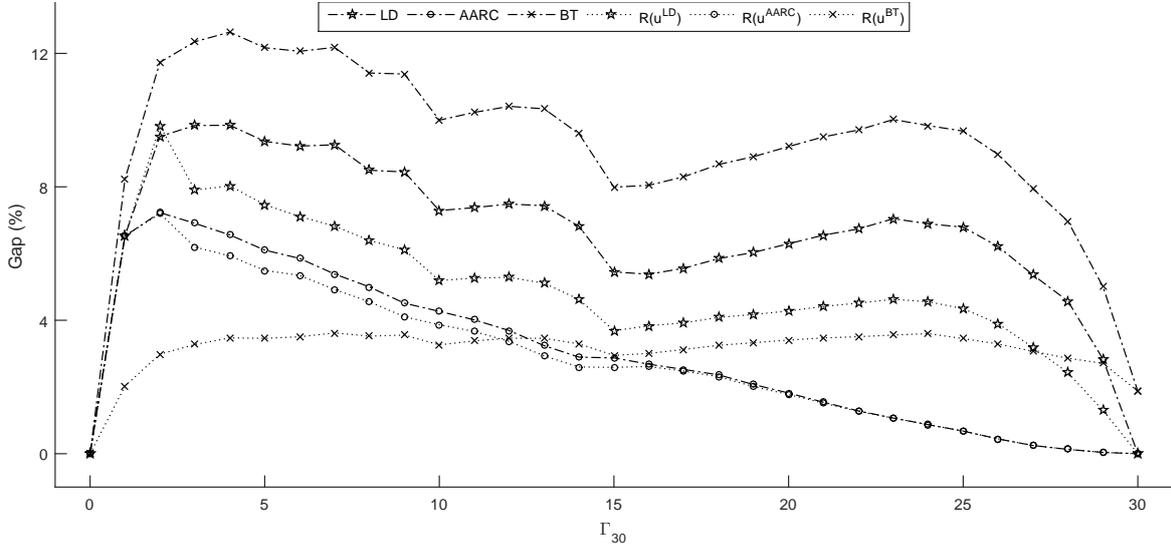


Figure 6: Average gaps in terms of the maximum number of deviations considering no setup costs.

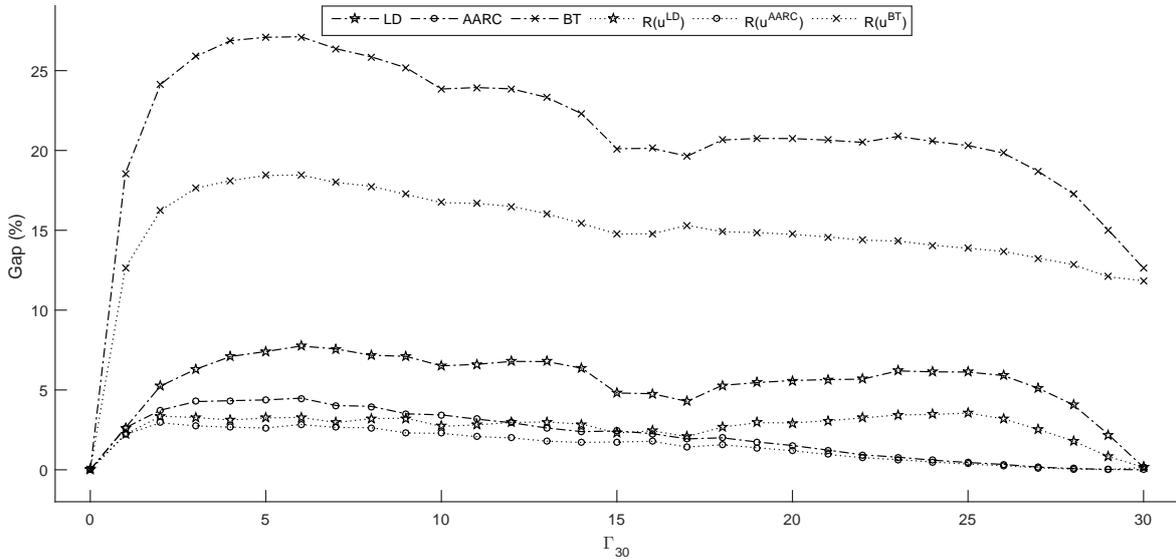


Figure 7: Average gaps in terms of the maximum number of deviations considering a setup cost equal to 150.

For both cases, the average gap associated with the LD and AARC approaches is always lower than 10% and 7%, respectively, while for the BT approach such gap can reach 28%. In particular, in our experiments, for the box-constrained case, there is no gap associated with the upper bounds obtained by both the LD and AARC approaches. In general, the average gap associated with the true cost of the solutions determined by both LD and AARC approaches tends to decrease as the number of deviations increases and it is zero for the box-constrained case.