

# The L-Shaped Method for Stochastic Programs with Decision-Dependent Uncertainty

Giovanni Pantuso<sup>1</sup> and Mike Hewitt<sup>2</sup>

<sup>1</sup>Department of Mathematical Sciences, University of Copenhagen,  
Denmark , gp@math.ku.dk

<sup>2</sup>Department of Information Systems and Supply Chain Management,  
Loyola University Chicago, USA , mhewitt3@luc.edu

## Abstract

In this paper we extend the well-known L-Shaped method to solve two-stage stochastic programming problems with decision-dependent uncertainty. The method is based on a novel, unifying, formulation and on distribution-specific optimality and feasibility cuts for both linear and integer stochastic programs. Extensive tests on three production planning problems illustrate that the method is extremely effective on large-scale instances.

## 1 Introduction

A two-stage stochastic program can be concisely expressed as

$$\min_{x \in \mathcal{X}} c^\top x + \mathbb{E}_P[Q(x, \boldsymbol{\xi})] \quad (1a)$$

where  $\mathcal{X}$  is a subset of  $\mathbb{R}^{n_1}$  (potentially finite and countable),  $c \in \mathbb{R}^{n_1}$  a known vector, and  $\boldsymbol{\xi} : \Omega \rightarrow \mathbb{R}^N$  a random variable defined on a probability space  $(\Omega, \Sigma, \mu)$  with distribution  $P$  induced by  $\mu$ . Furthermore, given  $x$ ,  $\mathbb{E}_P[Q(x, \boldsymbol{\xi})] = \int_{\Xi} Q(x, \xi) dP(\xi)$ , where  $\Xi$  is the support of  $P$ , and

$$Q(x, \xi) = \min_{y \in \mathcal{Y}} \{q^\top y \mid Wy = h - Tx\}. \quad (1b)$$

represents the cost of second stage decisions  $y$  from  $\mathcal{Y} \subseteq \mathbb{R}^{n_2}$  given a first stage decision  $x$  and a realization  $\xi$  of random variable  $\boldsymbol{\xi}$  which collects the realizations of the random data of the problem, namely  $q \in \mathbb{R}^{n_2}$ ,  $h \in \mathbb{R}^{m_2}$ ,  $W \in \mathbb{R}^{m_2 \times n_2}$  and  $T \in \mathbb{R}^{m_2 \times n_1}$ . For given  $x$ ,  $Q(x, \boldsymbol{\xi})$  is a measurable function from  $(\Omega, \Sigma, \mu)$  into the extended real line equipped with the Borel  $\sigma$ -algebra  $\mathcal{B}$ . The extended real line carries the usual protocol that  $Q(x, \xi) = \infty$  captures infeasibility and  $Q(x, \xi) = -\infty$  captures unboundedness.

A fundamental assumption in problem (1) is that the distribution  $\mathbf{P}$  of  $\xi$  on  $\mathbb{R}^N$  is known and independent of  $x$ . This problem has been proven remarkably effective in modeling a vast variety decision problems under uncertainty. Many, however, do not fall under this umbrella.

In this paper we allow the probability distribution to depend on the decision vector  $x$ . Particularly, we consider a finite (but potentially large) collection  $(\mathcal{X}_d)_{d \in \mathcal{D}}$  of subsets of  $\mathcal{X}$  that defines a partition. That is,  $\bigcup_{d \in \mathcal{D}} \mathcal{X}_d = \mathcal{X}$  and  $\mathcal{X}_d \cap \mathcal{X}_{d'} = \emptyset$  for all  $d, d' \in \mathcal{D}, d \neq d'$ . We assume that the membership of  $x$  to a subset  $\mathcal{X}_d$  determines a distribution  $\mathbf{P}_d$  on  $\mathbb{R}^N$  for the uncertainties  $\xi$ . In other words, the non-trivial relationship between  $x$  and the resulting distribution is piecewise-constant, i.e., different sub-regions of the feasible space induce different distributions. This implies a finite, but potentially very large, number of distributions.

The choice of working with disjoint subsets of  $\mathcal{X}$  is driven by our conceptualization of the optimization process. Namely, that the optimizer makes decisions that in turn completely determine the relevant second stage distribution. If we allow the sets  $\mathcal{X}_d$  to intersect, such a determination can not be made and one is potentially left with the situation of modeling that the optimizer explicitly chooses the relevant second stage distribution. We do not believe such a process reflects the role optimization plays in real-world decision making. That said, if one is able to uniquely determine the relevant distribution for each point in a non-empty intersection, we believe our method applies with minor modifications. However, we do not address this issue in this paper.

We define  $\mathbb{1}_d : \mathcal{X} \rightarrow \{0, 1\}$  to be the characteristic function of  $\mathcal{X}_d$ . Namely,  $\mathbb{1}_d(x) = 1$  iff  $x \in \mathcal{X}_d$ , 0 otherwise. Given this notation, in this paper we focus on the following stochastic program with decision-dependent uncertainty

$$\min_{x \in \mathcal{X}} c^\top x + \sum_{d \in \mathcal{D}} \mathbb{1}_d(x) \mathbb{E}_{\mathbf{P}_d}[Q(x, \xi)] \quad (2)$$

Observe that  $x \mapsto \sum_{d \in \mathcal{D}} \mathbb{1}_d(x) = 1$ . This follows from the fact that the collection  $(\mathcal{X}_d)_{d \in \mathcal{D}}$  is a partition of  $\mathcal{X}$ . That is, the problem can be viewed as a type of disjunctive program [4]. Observe that in (2) infeasibility is tolerated (i.e.,  $Q(x, \xi) = \infty$  with probability larger than zero) under distributions other than the one enforced by  $x$ . We assume the problem is bounded from below.

Problem (2) leaves sufficient freedom to model a large variety of configurations. For example, the subsets  $\mathcal{X}_d$  may be convex polyhedra specifying, e.g., admissible ranges of values on individual components of  $x$ . If  $\mathcal{X}$  is a subset of the  $n_1$ -dimensional integers, subsets  $\mathcal{X}_d$  may represent subsets of discrete choices, specific combinations, arrangements, permutations, etc. The resulting distributions  $\mathbf{P}_d$  may in turn be parametrized by a function  $\theta : \{0, 1\}^{|\mathcal{D}|} \rightarrow \mathbb{R}^P$  where  $P$  is the number of parameters of  $\mathbf{P}_d$ , or even assume a different functional form over the different subsets  $\mathcal{X}_d$ . Furthermore, in situations where the distribution is explicitly parameterized by the decision vector, e.g.,  $\theta(x)$ , this formulation allows approximating the distribution using some reference set of parameters in each subset, e.g.,  $\theta(\bar{x}_d)$  for some  $\bar{x}_d \in \mathcal{X}_d, d \in \mathcal{D}$ .

An example of an application that can be modeled as Problem (2) is a production planning problem in which an organization seeks to determine expected profit-maximizing production volumes in the presence of uncertainty in both supply and demand. Uncertainty in supply is endogeneous and depends on choices regarding the combination of facilities that produce

and the respective production volumes. More generally, such a problem is similar to those studied in the literature on lot sizing with random yields [33]. We will return to a similar problem in Section 4.

The primary contribution of this paper is an extension of the classical L-Shaped and Integer L-Shaped methods to efficiently solve problem (2). As the relevant second stage distribution depends on first stage decisions, underlying the method is the idea of *distribution-specific* feasibility and optimality cuts, which we derive and prove. The overall method is general but the form of such cuts depends on the nature of the second stage problem,  $Q(x, \xi)$ . In particular, we demonstrate how such cuts can be formulated depending on whether  $Q(x, \xi)$  forms a linear or a mixed-integer linear program. Furthermore, we propose inequalities that are distribution-independent and may speed up convergence. Beyond proving theoretical convergence of the proposed method, we demonstrate computationally on three variants of a production planning problem that it significantly outperforms a state-of-the-art benchmark.

The remainder of this article is organized as follows. We begin by summarizing the relevant literature on stochastic programs with endogenous uncertainty in Section 2. We then introduce the method in Section 3. In Section 4 we introduce three production planning problems we use in the computational study presented in Section 5. Finally, we draw conclusions in Section 6.

## 2 Literature and contribution

According to [9] there exist at least two ways in which decisions can determine the nature of uncertainties. The first possibility is that decisions determine the time when the uncertainty is resolved. A typical example is provided by [3]. An oil company has to decide which gas reservoirs to explore and when. Exploration is performed by installing drilling equipment. The size and quality of the reservoirs is uncertain. The uncertainty is resolved only after the sites have been explored. Stochastic programs of this kind are studied in a number of articles including [5, 27, 28, 6, 10, 23, 29, 3].

The second possibility – which is of interest in this paper – is that decisions alter the probability space underlying the random variables, hence their distributions. The methodological landscape for this type of problems is sparser. One of the first approaches dates back to [26] who consider a Markovian random process whose transition probabilities depend on the decision variables of the optimization problem. The author provides an algorithm that converges with probability one. Later, [18] consider the case where decisions influence both the probability measure and the timing of the observation. Their framework includes both two- and multi-stage problems. Nevertheless, the decisions that have an impact on the uncertainty are entirely made at the first decision stage. As in the case described in this article, the authors assume that the set of probability measures which can be enforced by decisions is finite and countable. The authors show that the problem can be recast as choosing the best among the stochastic programs determined by a choice of a distribution and propose an implicit enumeration algorithm. In a similar framework, [24] considers multi-stage problems and extends the model of [18] by allowing decisions at all stages to determine the probability measure for the later stages. The author provides a new scenario tree structure and mathematical formulation that avoids explicit statement of non-anticipativity constraints which are

typically linked to model size growth, see e.g., [3, 15, 16]. Furthermore, the author provides an exact solution method for a special case. Considering two-stage stochastic programs, [12] and [13] discuss several ways of modeling the interplay between the decision variables and the parameters of the underlying probability distributions. Particularly, the authors formulate two-stage models where prior probabilities are distorted through affine transformations, or combined using convex combinations of several probability distributions. Furthermore, the authors present models which incorporate the parameters of the probability distribution as first stage decision variables. Practical applications of problems with endogenous uncertainty can be found in [1, 31, 7, 25, 20, 11, 30, 8, 14, 21]. We do not discuss these papers in detail for sake of brevity.

In this article we focus on stochastic programs of the second kind. The contribution of this article is twofold. First, we propose a general-purpose model for two-stage stochastic programs with decision-dependent uncertainty. Particularly, we are interested in the case where there exist finitely many distributions that can be enforced by the choice of  $x$ . Second, we provide a general purpose exact algorithm which extends the well-known L-Shaped method. Particularly, optimality and feasibility cuts are derived for problems which may include integer variables at both stages.

This article extends the available literature in several ways. In [12, 13] given reference discrete distributions are transformed by first stage decisions through affine transformations. Hence, their approach can handle all the (infinitely many) discrete distributions that may materialize as an affine transformation of the given reference distributions. We use a different approach. On the one hand, we restrict ourselves to finitely (though potentially combinatorially) many distributions. On the other hand, each one of these distributions is arbitrarily dependent on  $x$ . In particular, our model allows for arbitrary mappings of (subsets of) first stage decisions to probability distributions. When the number of probability distributions is naturally finite, our model is exact. This is the case, in particular, when the underlying distribution depends only on specific finite features of the first stage solutions,  $x$ . When  $\mathcal{X}$  is countable and bounded, our model is always exact.

The problem we study is in line with the work of [18] and [24] in that the set of potential probability distributions enforced by first stage decisions is finite and countable. In particular, in [18] the authors propose an implicit enumeration algorithm which relies on computing and storing bounds for each choice of a probability distribution. As the authors acknowledge, this solution strategy is viable only when the set of probability distributions is small. As will be more evident in Section 4.1, this may in general not be the case in large scale problems. To overcome these methodological limitations, we extend the L-Shaped method by deriving novel distribution-specific optimality and feasibility cuts. The method is generally applicable to stochastic programs, possibly with integer variables at both stages.

### 3 The L-shaped method for problems with endogenous uncertainty

In this section, we adapt the L-Shaped method to provide an exact algorithm for problem (2). As part of that adaptation we derive new distribution-specific optimality and feasibility

cuts. We note that the operations of the proposed method are based on the assumption that problem (2) is feasible and bounded from below. We treat, separately, two important cases. The first case, discussed in Section 3.1, is when  $\mathcal{Y} \subseteq \mathbb{R}^{n_2}$ . In the second, discussed in Section 3.2,  $\mathcal{Y} \subseteq \mathbb{Z}^{n_2^1} \times \mathbb{R}^{n_2^2}$ ,  $n_2^1 + n_2^2 = n_2$ . For each case we show there are families of functions,  $\mathcal{F} = \{\mathcal{F}_1, \dots, \mathcal{F}_{|\mathcal{D}|}\}$ ,  $\mathcal{O} = \{\mathcal{O}_1, \dots, \mathcal{O}_{|\mathcal{D}|}\}$ , that induce constraints such that the following is an equivalent reformulation of (2).

$$\min_{x \in \mathcal{X}} c^\top x + \mu \quad (3a)$$

$$x \in \mathcal{X}_d \implies \begin{cases} f_d(x) \leq 0 & \forall f_d(\cdot) \in \mathcal{F}_d \\ o_d(x) \leq \mu & \forall o_d(\cdot) \in \mathcal{O}_d \end{cases} \quad \forall d \in \mathcal{D} \quad (3b)$$

Namely, given an optimal solution  $(x^*, \mu^*)$  to (3),  $x^*$  is optimal for (2) and  $\mu^* = \sum_{d \in \mathcal{D}} \mathbb{1}_d(x^*) \mathbb{E}_{\mathbf{P}_d}[Q(x^*, \xi)]$ . We refer to the set  $\mathcal{F}_d$  as a set of distribution-specific (to  $d$ ) feasibility cuts and show that  $\mathcal{F}_d$  is finite. We show the same for the set of distribution-specific optimality cuts,  $\mathcal{O}_d$ .

Similar to the classical L-Shaped method, we consider subsets  $\underline{\mathcal{F}}_d \subseteq \mathcal{F}_d$ ,  $\underline{\mathcal{O}}_d \subseteq \mathcal{O}_d$ ,  $d \in \mathcal{D}$  of each set of functions and formulate the following *Relaxed Master Problem* (RMP).

$$\min_{x \in \mathcal{X}} c^\top x + \mu \quad (4a)$$

$$x \in \mathcal{X}_d \implies \begin{cases} f_d(x) \leq 0 & \forall f_d(\cdot) \in \underline{\mathcal{F}}_d, \\ o_d(x) \leq \mu & \forall o_d(\cdot) \in \underline{\mathcal{O}}_d. \end{cases} \quad (4b)$$

A high-level overview of the steps the method executes at an iteration is presented in Algorithm 1. At iteration  $i$ , the method solves problem (4) formulated with sets  $\underline{\mathcal{F}}_d^i, \underline{\mathcal{O}}_d^i$ , to obtain the solution  $(x^i, \mu^i)$ . We note that, at the initial iterations, without a valid lower bound for  $\mu$ , it may be necessary to exclude  $\mu$  from optimization and set its value to  $\infty$  to prevent unboundedness. The method then identifies the distribution  $d^i$  enforced by  $x^i$ . Next, the method checks whether  $x^i$  yields second stage feasible subproblems under distribution  $d^i$ . If it does not, the method identifies a feasibility cut from  $\mathcal{F}_{d^i} \setminus \underline{\mathcal{F}}_{d^i}^i$  and adds it to  $\underline{\mathcal{F}}_{d^i}^{i+1}$ .

If the feasibility check is passed, the method moves on to assess the accuracy of  $\mu^i$ , the estimate of expected second stage costs. To do so, it computes the expected second stage cost  $\mathbb{E}_{\mathbf{P}_{d^i}}[Q(x^i, \xi)]$  under the distribution  $d^i$  enforced by  $x^i$ . If the approximation  $\mu^i$  is lower than this expected value it identifies an optimality cut from  $\mathcal{O}_{d^i} \setminus \underline{\mathcal{O}}_{d^i}^i$  and adds it to  $\underline{\mathcal{O}}_{d^i}^{i+1}$ . If the optimality check is passed, the solution  $x^i$  is provably optimal and the method terminates.

We note that the method described in Algorithm 1 is amenable to well-known enhancements of the classical L-Shaped method. For example, as (4) is defined over subsets of  $\mathcal{F}$  and  $\mathcal{O}$  it is a relaxation of (3) (and hence (2)). Hence, its objective value provides a lower bound on the optimal objective value to (2). Also, we note that when  $x^i$  passes the feasibility check,  $c^\top x^i + \mathbb{E}_{\mathbf{P}_{d^i}}[Q(x^i, \xi)]$  is an upper bound on the optimal objective value to (2). The method may then terminate when the gap between the two bounds falls below a pre-defined threshold. Furthermore, the implications (4b) imply that (4) is a MILP. Therefore, Algorithm 1 can be invoked at nodes in a branch-and-bound procedure that solves (4), see [19].

The following sections present appropriate families of functions  $\mathcal{F}, \mathcal{O}$  for different classes of second stage problems. For the sake of generality and ease of exposition, the constraints

---

**Algorithm 1** L-Shaped Method for (2)

---

```
 $i \leftarrow 0$ , solved  $\leftarrow$  false  
while not solved do  
  Solve (4). Let  $(x^i, \mu^i)$  be its optimal solution  
  Identify  $d^i \in \mathcal{D}$  such that  $x^i \in \mathcal{X}_{d^i}$   
  if  $x^i$  is second stage infeasible then  
    Add a feasibility cut from  $\mathcal{F}_{d^i} \setminus \underline{\mathcal{F}_{d^i}^i}$  to  $\underline{\mathcal{F}_{d^i}^{i+1}}$   
  else if  $\mu^i < \mathbb{E}_{P_{d^i}}[Q(x^i, \xi)]$  then  
    Add an optimality cut from  $\mathcal{O}_{d^i} \setminus \underline{\mathcal{O}_{d^i}^i}$  to  $\underline{\mathcal{O}_{d^i}^{i+1}}$   
  else  
    solved  $\leftarrow$  true  
  end if  
   $i \leftarrow i + 1$   
end while
```

---

based on those functions are expressed with indicator functions,  $\mathbb{1}_d(x)$ . Note, however, that a direct implementation of such indicator functions may require a large number of binary variables (at least  $|\mathcal{D}|$ ) in (4). At the end of the section we will show how, under rather general conditions, the implications  $x \in \mathcal{X}_d \implies f_d(x) \leq 0$  and  $x \in \mathcal{X}_d \implies o_d(x) \leq \mu$  can be modeled using fewer binary variables.

### 3.1 Continuous second stage

In this section, we consider the case in which the second stage problem can take the form of a linear program. Namely

$$Q(x, \xi) = \min_{y \geq 0} \{q^\top y \mid Wy = h - Tx\}. \quad (5)$$

The validity of the proposed optimality and feasibility cuts require the following assumption.

- A1** For each  $d \in \mathcal{D}$ , the uncertainty  $\xi$  has discrete distribution  $P_d = \sum_{s \in \mathcal{S}_d} \pi_{sd} \Delta_s$ , supported by  $\Xi_d = \{\xi_{1d}, \dots, \xi_{S_d}\}$  where  $\Delta_s$  is the Dirac measure that assigns total mass to realization  $\xi_{sd}$  for  $s$  in the set of scenarios  $\mathcal{S}_d$  and  $\pi_{sd}$  is the probability of realization  $\xi_{sd}$  under distribution  $d \in \mathcal{D}$ .

When  $\xi$  has a continuous distribution, assumption A1 can be satisfied by means of a discretization technique. In what follows, we begin by introducing distribution-specific feasibility cuts for problems without complete recourse. We then continue by introducing distribution-specific optimality cuts.

#### 3.1.1 Distribution-specific feasibility cuts

When complete recourse is not present, the method must be augmented with a technique for generating feasibility cuts to ensure convergence. We next define, for all  $d \in \mathcal{D}$  and  $s \in \mathcal{S}_d$

the set

$$\mathcal{K}_{sd} = \{x \in \mathbb{R}^{n_1} | \exists y \geq 0 : W_{sd}y = h_{sd} - T_{sd}x\}.$$

In words,  $\mathcal{K}_{sd}$  contains values of  $x$  such that the corresponding subproblem for scenario  $s$  of distribution  $d$  is feasible. Relatedly, we define for all  $d \in \mathcal{D}$ ,  $\mathcal{K}_d = \bigcap_{s \in \mathcal{S}_d} \mathcal{K}_{sd}$ . This allows us to define relatively complete recourse more precisely as  $\mathcal{X} \subseteq \bigcap_{d \in \mathcal{D}} \mathcal{K}_d$  and complete recourse as  $\mathcal{K}_d = \mathbb{R}^{n_1}$  for all  $d \in \mathcal{D}$ . Observe, that we rule out the possibility that  $\mathcal{K}_d = \emptyset$  given that we assume problem (2) is feasible.

To define feasibility cuts, we make, in addition to A1, the following assumption.

**A2** For all  $d \in \mathcal{D}$  and  $s \in \mathcal{S}_d$ , there exists a constant  $U_{ds} < \infty$  such that

$$U_{ds} \geq \max_{x \in \mathcal{X}} \min_{y \in \mathbb{R}_+^{n_2}, w^+, w^- \in \mathbb{R}_+^{m_2}} \{\mathbf{1}^\top w^+ + \mathbf{1}w^- : W_{sd}y + w^+ - w^- = h_{sd} - T_{sd}x\}$$

where  $\mathbf{1}^\top = (1, \dots, 1) \in \mathbb{R}^{m_2}$ .

The quantities  $U_{ds}$  enable the definition of feasibility cuts that are distribution-specific. Namely, they are binding only for first stage solutions that induce the same distribution for which they were generated and redundant otherwise.

**Theorem 1.** *Let  $(x^v, \mu^v)$  be a solution to (4) that induces distribution  $d^v$  such that  $x^v \notin \mathcal{K}_{sd^v}$  for some  $s \in \mathcal{S}_{d^v}$ . Then, solution  $x^v$  violates inequality*

$$(\rho^v)^\top (h_{sd^v} - T_{sd^v}x) \leq U_{d^v s}(1 - \mathbb{1}_{d^v}(x^v)) \quad (6)$$

where  $\rho^v$  is an optimal solution to the dual of

$$\Psi(x^v, \xi_{sd^v}) := \min_{y \in \mathbb{R}_+^{n_2}, w^+, w^- \in \mathbb{R}_+^{m_2}} \{\mathbf{1}^\top w^+ + \mathbf{1}w^- : W_{sd^v}y + w^+ - w^- = h_{sd^v} - T_{sd^v}x^v\} \quad (7)$$

*Proof.* We note that  $\Psi(x^v, \xi_{sd^v})$  is a feasible problem and that  $x^v \notin \mathcal{K}_{sd^v}$  implies that  $\Psi(x^v, \xi_{sd^v}) > 0$ . Letting  $\rho^v$  be an optimal solution to the dual of  $\Psi(x^v, \xi_{sd^v})$ , we have that  $\Psi(x^v, \xi_{sd^v}) = (\rho^v)^\top (h_{sd^v} - T_{sd^v}x^v) > 0$ . In turn, given that  $\mathbb{1}_{d^v}(x^v) = 1$  we have

$$\Psi(x^v, \xi_{sd^v}) = (\rho^v)^\top (h_{sd^v} - T_{sd^v}x^v) > 0 = U_{d^v s}(1 - \mathbb{1}_{d^v}(x^v))$$

as required.  $\square$

The following theorem shows that the proposed feasibility cuts are safe, in the sense that a cut generated for a solution  $x^v$  to (4) is satisfied by solutions  $(x^l, \mu^l)$  to (4), for which  $x^l \neq x^v$ , and the subproblems are feasible in the distribution enforced by  $x^l$ .

**Theorem 2.** *Let  $(x^l, \mu^l)$  be a solution to (4) such that  $x^l \in \mathcal{K}_d$ . Then, solution  $x^l$  satisfies inequality*

$$(\rho^v)^\top (h_{sd^v} - T_{sd^v}x) \leq U_{d^v s}(1 - \mathbb{1}_{d^v}(x))$$

where  $\rho^v$  is an optimal solution to the dual of  $\Psi(x^v, \xi_{sd^v})$  for some solution  $x^v$  to (4) and  $s \in \mathcal{S}_{d^v}$ .

*Proof.* We consider separately the cases in which  $d^l \neq d^v$  and in which  $d^l = d^v$ .

When  $d^l \neq d^v$ , we have  $\mathbb{1}_{d^v}(x^l) = 0$  and inequality (6) reduces to

$$(\rho^v)^\top (h_{sd^v} - T_{sd^v} x) \leq U_{d^v s}$$

Consider the set  $\mathcal{R}_{sd} := \{\rho \in \mathbb{R}^{m_2} \mid \rho^\top W_{sd} \leq 0, -\mathbf{1} \leq \rho \leq \mathbf{1}\}$  so that  $\rho^v \in \operatorname{argmax}_{\rho \in \mathcal{R}_{sd^v}} \{\rho^\top (h_{sd^v} - T_{sd^v} x^v)\}$ . Then we have

$$\begin{aligned} (\rho^v)^\top (h_{sd^v} - T_{sd^v} x^l) &\leq \max_{\rho \in \mathcal{R}_{sd^v}} \rho^\top (h_{sd^v} - T_{sd^v} x^l) \\ &\leq \max_{x \in \mathcal{X}} \max_{\rho \in \mathcal{R}_{sd^v}} \rho^\top (h_{sd^v} - T_{sd^v} x) \\ &= \max_{x \in \mathcal{X}} \min_{y \in \mathbb{R}_+^{n_2}, w^+, w^- \in \mathbb{R}_+^{m_2}} \{\mathbf{1}^\top w^+ + \mathbf{1} w^- : W_{sd^v} y + w^+ - w^- = h_{sd^v} - T_{sd^v} x\} \\ &\leq U_{d^v s} \end{aligned}$$

Hence, the inequality is satisfied.

When  $d^l = d^v$ , as  $x^l \in \mathcal{K}_{d^l}$  we have  $x^l \in \mathcal{K}_{sd^l} = \mathcal{K}_{sd^v}$ . Consequently,

$$\Psi(x^l, \xi_{sd^v}) = \min_{y \in \mathbb{R}_+^{n_2}, w^+, w^- \in \mathbb{R}_+^{m_2}} \{\mathbf{1}^\top w^+ + \mathbf{1} w^- : W_{sd^v} y + w^+ - w^- = h_{sd^v} - T_{sd^v} x^l\} = 0$$

By strong duality of  $\Psi(x^l, \xi_{sd^v})$  we can write

$$0 = \Psi(x^l, \xi_{sd^v}) = (\rho^l)^\top (h_{sd^v} - T_{sd^v} x^l)$$

where  $\rho^l$  is an optimal solution to the dual of  $\Psi(x^l, \xi_{sd^v})$ . As  $\rho^l$  is optimal for the dual while  $\rho^v$  is feasible we have, as required, that

$$U_{d^v s}(1 - \mathbb{1}_{d^v}(x^l)) = 0 = \Psi(x^l, \xi_{sd^v}) = (\rho^l)^\top (h_{sd^v} - T_{sd^v} x^l) \geq (\rho^v)^\top (h_{sd^v} - T_{sd^v} x^l)$$

□

The following proposition shows that the method will deliver second stage feasible solutions within a finite number of iterations.

**Lemma 3.** *There exist only finitely many cuts (6).*

*Proof.* This follows immediately from the finite number of distributions  $\mathcal{D}$ , finite number of realizations  $\mathcal{S}_d$  for all  $d \in \mathcal{D}$ , and finite number of extreme points  $\rho$  in the feasible region of the dual to  $\Psi(x, \xi_{sd})$  for all  $s \in \mathcal{S}_d$  and  $d \in \mathcal{D}$ . □

### 3.1.2 Distribution-specific optimality cut

The following proposition introduces a distribution-specific, duality-based, optimality cut. The validity of this cut relies on the existence of a constant  $U$  that satisfies the following inequalities.



**A3** There exists a constant  $U$  such that  $\infty > U \geq \max_{d \in \mathcal{D}} U_d - \min_{d \in \mathcal{D}} L_d$  where

$$U_d \geq \max_{x \in \mathcal{X}_d} \sum_{s \in \mathcal{S}_d} \pi_{sd} Q(x, \xi_{sd})$$

$$L_d \leq \min_{x \in \mathcal{X}_d} \sum_{s \in \mathcal{S}_d} \pi_{sd} Q(x, \xi_{sd})$$

We note that such a  $U$  exists when the second stage problem is bounded above and below, when feasible.

**Theorem 4.** *Let  $(x^v, \mu^v)$  be a solution to (4) and  $d^v \in \mathcal{D}$  such that  $x^v \in \mathcal{X}_{d^v}$ . Assume  $x^v \in \mathcal{K}_{d^v}$  and that  $\mu^v < Q_{d^v}(x^v)$  with*

$$Q_{d^v}(x^v) = \sum_{s \in \mathcal{S}_{d^v}} \pi_{sd^v} Q(x^v, \xi_{sd^v})$$

*Then, solution  $(x^v, \mu^v)$  violates inequality*

$$\mu \geq \sum_{s \in \mathcal{S}_{d^v}} \pi_{sd^v} (\rho_s^v)^\top (h_{sd^v} - T_{sd^v} x) - U(1 - \mathbb{1}_{d^v}(x)) \quad (8)$$

*where  $\rho_s^v$  is an optimal solution to the dual of  $Q(x^v, \xi_{sd^v})$ .*

*Proof.* Since  $\mu^v < Q_{d^v}(x^v)$  we can write

$$\begin{aligned} \mu^v &< Q_{d^v}(x^v) = \sum_{s \in \mathcal{S}_{d^v}} \pi_{sd^v} Q(x^v, \xi_{sd^v}) \\ &= \sum_{s \in \mathcal{S}_{d^v}} \pi_{sd^v} \left( (\rho_s^v)^\top (h_{sd^v} - T_{sd^v} x) \right) \\ &= \sum_{s \in \mathcal{S}_{d^v}} \pi_{sd^v} \left( (\rho_s^v)^\top (h_{sd^v} - T_{sd^v} x) \right) - U(1 - \mathbb{1}_{d^v}(x^v)) \end{aligned}$$

where the second equality holds by strong duality of  $Q(x^v, \xi_{sd^v})$  (recall that  $x^v \in \mathcal{K}_{d^v}$ ) and the third equality holds since  $\mathbb{1}_{d^v}(x^v) = 1$ .  $\square$

Theorem 4 introduces an optimality cut that renders infeasible in (4) a pair  $(x^v, \mu^v)$  whenever  $\mu^v$  does not hold the correct expected cost conditional on  $x^v$ . It is well-known that, given  $d$ ,  $Q_d(x)$  is convex and piece-wise linear, see e.g., [32]. Furthermore,  $\sum_{s \in \mathcal{S}_d} \pi_{sd} \rho_s^v (h_{sd} - T_{sd} x)$  defines a supporting hyperplane to the epigraph of  $Q_d(x)$  at  $x^v$ . The last term in the optimality cut (8) ensures that the cut is effective only when  $x$  enforces distribution  $d^v$ .

Having established that the inequality renders the current solution  $(x^v, \mu^v)$  infeasible, we next prove that the inequality is satisfied by solutions  $(x^l, \mu^l)$ ,  $x^l \neq x^v$ , such that  $\mu^l$  does not under-estimate expected second stage costs.

**Theorem 5.** Let  $(x^l, \mu^l)$  be a solution to (4) such that  $x^l \in \mathcal{K}_{d^l}$  and for which  $\mu^l = Q_{d^l}(x^l)$ . Here  $d^l \in \mathcal{D}$  is such that  $x^l \in \mathcal{X}_{d^l}$ . Then, solution  $(x^l, \mu^l)$  satisfies inequality

$$\mu \geq \sum_{s \in \mathcal{S}_{d^v}} \pi_{sd^v} (\rho_s^v)^\top (h_{sd^v} - T_{sd^v} x) - U(1 - \mathbb{1}_{d^v}(x)) \quad (9)$$

where  $\rho_s^v$  is an optimal solution to the dual of  $Q(x^v, \xi_{sd^v})$  and  $d^v$  the distribution enforced by some solution  $x^v$  to (4).

*Proof.* There are two cases to consider. The first,  $d^l \neq d^v$ , reflects when the distribution enforced by solution  $x^l$  is different from the distribution enforced by the solution  $x^v$  for which the inequality was generated. The second,  $d^l = d^v$ , is when the two distributions are the same.

When  $d^l \neq d^v$  we have

$$\begin{aligned} \mu^l &\geq \sum_{s \in \mathcal{S}_{d^v}} \pi_{sd^v} \left( (\rho_s^v)^\top (h_{sd^v} - T_{sd^v} x^l) \right) - U \\ &= \sum_{s \in \mathcal{S}_{d^v}} \pi_{sd^v} \left( (\rho_s^v)^\top (h_{sd^v} - T_{sd^v} x^l) \right) - U(1 - \mathbb{1}_{d^v}(x^l)) \end{aligned}$$

where the inequality holds due to the definition of  $U$ , since

$$\sum_{s \in \mathcal{S}_{d^v}} \pi_{sd^v} \left( (\rho_s^v)^\top (h_{sd^v} - T_{sd^v} x^l) \right) \leq \max_{x \in \mathcal{X}} \sum_{s \in \mathcal{S}_{d^v}} \pi_{sd^v} Q(x, \xi_{sd^v})$$

and

$$\mu^l = Q_{d^l}(x^l) \geq \min_{x \in \mathcal{X}} \sum_{s \in \mathcal{S}_{d^l}} \pi_{sd^l} Q(x, \xi_{sd^l})$$

and the equality holds since the case implies that  $\mathbb{1}_{d^v}(x^l) = 0$ .

When  $d^l = d^v$  we have, given that  $\rho_s^l$  is an optimal solution to the dual of  $Q(x^l, \xi_{sd^l})$ ,

$$\begin{aligned} \mu^l &= \sum_{s \in \mathcal{S}_{d^l}} \pi_{sd^l} Q(x^l, \xi_{sd^l}) = \sum_{s \in \mathcal{S}_{d^v}} \pi_{sd^v} Q(x^l, \xi_{sd^v}) \\ &= \sum_{s \in \mathcal{S}_{d^v}} \pi_{sd^v} \left( (\rho_s^l)^\top (h_{sd^v} - T_{sd^v} x^l) \right) \\ &\geq \sum_{s \in \mathcal{S}_{d^v}} \pi_{sd^v} \left( (\rho_s^v)^\top (h_{sd^v} - T_{sd^v} x^l) \right) \\ &= \sum_{s \in \mathcal{S}_{d^v}} \pi_{sd^v} \left( (\rho_s^v)^\top (h_{sd^v} - T_{sd^v} x^l) \right) - U(1 - \mathbb{1}_{d^v}(x^l)) \end{aligned}$$

The second equality holds because the distributions  $d^l, d^v$  are the same. The third equality holds due to strong duality of  $Q(x^l, \xi_{sd^l})$ . The inequality follows from dual optimality of  $\rho_s^l$  and dual feasibility of  $\rho_s^v$ . The last equality holds as  $\mathbb{1}_{d^v}(x^l) = \mathbb{1}_{d^l}(x^l) = 1$ .  $\square$

**Lemma 6.** There exist only finitely many optimality cuts (8).

*Proof.* This follows immediately from the finite number of distributions  $\mathcal{D}$  and from the finite number of extreme points  $\rho$  in the feasible region of the dual to  $Q(x, \xi)$  for each  $\xi$ .  $\square$

### 3.1.3 Reformulation

In this section we present a reformulation of (2) in terms of the feasibility and optimality cuts just proposed. We also prove the validity of this reformulation. Namely, that an optimal solution to this reformulation can be used to derive an optimal solution to (2) with the same objective function value.

The reformulation is as follows

$$\begin{aligned} \min_{x \in \mathcal{X}} \quad & c^\top x + \mu \\ & U_{sd}(1 - \mathbb{1}_d(x)) \geq (\sigma_s^v)^\top (h_{sd} - T_{sd}x) \quad \forall d \in \mathcal{D}, s \in \mathcal{S}_d, \sigma_s^v \in F_d^s, \\ & \mu \geq \sum_{s \in \mathcal{S}_d} \pi_{sd}(\rho_s^v)^\top (h_{sd} - T_{sd}x) - U(1 - \mathbb{1}_d(x)) \quad \forall d \in \mathcal{D}, \rho^v \in O_d. \end{aligned} \tag{10}$$

where  $F_d^s$  is the set of extreme points of the polyhedron  $\{\sigma : \sigma W_{sd} \leq 0, -\mathbf{1} \leq \sigma \leq \mathbf{1}\}$ ,  $\rho^v = (\rho_s^v)_{s \in \mathcal{S}_d}$ ,  $O_d^s$  is the set of extreme points of the polyhedron  $\{\rho : \rho W_{sd} \leq q_{sd}\}$ ,  $O_d = O_d^1 \times O_d^2 \times \dots \times O_d^{|\mathcal{S}_d|}$ , and  $U_{sd}$  and  $U$  are defined as above. We note that, given Lemma 3 and Lemma 6, there are a finite number of feasibility and optimality cuts used to define this formulation.

**Proposition 7.** *An optimal solution  $(x^*, \mu^*)$ ,  $x^* \in \mathcal{X}_{d^*}$  of (10) induces an optimal solution  $x^*$  to (2) of value  $c^\top x^* + \mu^*$  and  $\mu^* = \mathbb{E}_{\mathbf{P}_{d^*}}[Q(x^*, \xi)]$ .*

*Proof.* The proof can be obtained by contradiction. First, assume that the solution  $x^*$  to (10) is such that  $x^* \notin \mathcal{K}_{d^*}$ . In other words, for the relevant distribution  $d^*$ ,  $x^*$  induces an infeasible subproblem for some scenario  $s \in \mathcal{S}_{d^*}$ . However, by Theorem 1 this implies a feasibility cut of the form (6) can be generated and added to (10), contradicting the premise that (10) is formulated with all such (finitely many) cuts (6). Next, suppose  $\mu^* < \sum_{d \in \mathcal{D}} \mathbb{1}_d(x^*) \mathbb{E}_{\mathbf{P}_d}[Q(x^*, \xi)] = \mathbb{E}_{\mathbf{P}_{d^*}}[Q(x^*, \xi)] = Q_{d^*}(x^*)$ . However, by Theorem 4 this implies an optimality cut of the form (8) can be generated and added to (10), contradicting the premise that (10) is formulated with all such (finitely many) cuts.  $\square$

**Theorem 8.** *Assume (2) is feasible and bounded. Then the L-Shaped algorithm with distribution-specific cuts converges to an optimal solution in a finite number of iterations.*

*Proof.* The observation that there are a finite number of cuts of the form (6) and (8), coupled with Proposition 7, provides a sufficient condition for the convergence of an L-Shaped method that iteratively adds, and does not remove, cuts of those forms to (4).  $\square$

## 3.2 Integer second stage

In this section, we consider the case where the expected second stage cost may take an arbitrary form. That is, given  $x$ , the quantity  $Q_d(x) = \mathbb{E}_{\mathbf{P}_d}[Q(x, \xi)]$  can be computed. A case of practical interest is when the second stage problem takes the form of a Mixed Integer Linear Program (MILP), that is

$$Q(x, \xi) = \min_y \{q^\top y \mid Wy = h - Tx, y \in \mathcal{Y}\} \tag{11}$$

and  $\mathcal{Y}$  imposes integrality restrictions on at least one second stage variable. As such, duality theory is not available to generate the cuts presented above for the case of a continuous second stage. The following assumption is required for both the optimality and feasibility cuts proposed in this section.

**A4**  $\mathcal{X} = \{0, 1\}^{n_1}$ .

In other words, the first stage decision variables that have an impact on second stage costs must take on binary values. Relatedly, both inequalities rely on identifying the set  $\mathcal{I}^v = \{i \in \{1, \dots, n_1\} | x_i^v = 1\}$  for solutions  $x^v$  to (4).

### 3.2.1 Distribution-specific feasibility cut

The inequality we present in this section can render any solution infeasible in (4), including those that induce an infeasible second stage subproblem of the relevant distribution. As for the case of continuous second stage problems, we define for all  $d \in \mathcal{D}$  and  $s \in \mathcal{S}_d$  the set

$$\mathcal{K}_{sd} = \{x \in \mathbb{R}^{n_1} | \exists y \in \mathcal{Y} : W_s y = h_s - T_s x\}.$$

Similarly, for all  $d \in \mathcal{D}$ , we let  $\mathcal{K}_d = \bigcap_{s \in \mathcal{S}_d} \mathcal{K}_{sd}$ . We then generate cuts of the form (12) for solutions  $x^v \notin \mathcal{K}_{d^v}$ .

**Theorem 9.** *Let  $x^v$  be a solution to (4). Then solution  $x^v$  violates the following inequality.*

$$\sum_{i \in \mathcal{I}^v} x_i - \sum_{i \notin \mathcal{I}^v} x_i \leq |\mathcal{I}^v| - 1 \tag{12}$$

Furthermore, the inequality is satisfied by any  $x^l \neq x^v$ .

*Proof.* As  $x_i^v = 1 \forall i \in \mathcal{I}^v$  and  $x_i^v = 0 \forall i \notin \mathcal{I}^v$ , we have that

$$\sum_{i \in \mathcal{I}^v} x_i^v - \sum_{i \notin \mathcal{I}^v} x_i^v = |\mathcal{I}^v|.$$

Thus,  $x^v$  violates the proposed inequality.

Next, consider a solution  $x^l \neq x^v$ . As either  $\exists i \in \mathcal{I}^v$  such that  $x_i^l = 0$  or  $\exists i \notin \mathcal{I}^v$  such that  $x_i^l = 1$ , we always have that

$$\sum_{i \in \mathcal{I}^v} x_i^l - \sum_{i \notin \mathcal{I}^v} x_i^l < |\mathcal{I}^v|.$$

□

Observe, however, that unlike cuts (6), cuts of type (12) only render infeasible a single solution.

### 3.2.2 Distribution-specific optimality cut

We next propose an optimality cut that renders infeasible a solution  $(x^*, \mu^*)$  such that  $\mu^*$  does not correctly estimate expected second stage costs. To derive optimality cuts for this case we extend the optimality cuts for integer stochastic programs [19], which rely on assumption A4 and A5 introduced next.

**A5** There exists a constant  $U$  such that  $\infty > U \geq \max_{d \in \mathcal{D}} L_d - \min_{d \in \mathcal{D}} L_d$  where  $L_d \leq \min_{x \in \mathcal{X}} \sum_{s \in \mathcal{S}_d} \pi_{sd} Q(x, \xi_s)$

The optimality cut is as follows.

**Theorem 10.** *Let  $(x^v, \mu^v)$  be a solution to (4) for which  $\mu^v < Q_{d^v}(x^v)$ . Then,  $(x^v, \mu^v)$  violates the following inequality.*

$$\mu \geq (Q_{d^v}(x^v) - L_{d^v}) \left( \sum_{i \in \mathcal{I}^v} x_i - \sum_{i \notin \mathcal{I}^v} x_i \right) - (Q_{d^v}(x^v) - L_{d^v}) (|\mathcal{I}^v| - 1) + L_{d^v} - U (1 - \mathbb{1}_{d^v}(x)) \quad (13)$$

*Proof.* For  $x = x^v$  we have  $\mathbb{1}_{d^v}(x^v) = 1$  and the inequality reduces to

$$\mu \geq (Q_{d^v}(x^v) - L_{d^v}) \left( \sum_{i \in \mathcal{I}^v} x_i - \sum_{i \notin \mathcal{I}^v} x_i \right) - (Q_{d^v}(x^v) - L_{d^v}) (|\mathcal{I}^v| - 1) + L_{d^v}$$

Following Proposition 2 in [19] for  $x = x^v$ , the right-hand-side reduces to  $Q_{d^v}(x^v)$ . Hence, the inequality is violated given the premise that  $\mu^v < Q_{d^v}(x^v)$ .  $\square$

The following proposition shows that the inequality is not violated by solutions  $(x^l, \mu^l)$  such that  $\mu^l$  does not under-estimate expected second stage costs.

**Theorem 11.** *Let  $(x^l, \mu^l)$  be a solution to (4) for which  $\mu^l = Q_{d^l}(x^l)$ . Then,  $(x^l, \mu^l)$  satisfies optimality cut (13) generated for some  $x^v, x^v \neq x^l$ .*

*Proof.* We again consider the cases  $d^l \neq d^v$  and  $d^l = d^v$  separately.

When  $d^l \neq d^v$  we have  $\mathbb{1}_{d^v}(x^l) = 0$  and the inequality reduces to

$$\mu^l \geq (Q_{d^v}(x^v) - L_{d^v}) \left( \sum_{i \in \mathcal{I}^v} x_i^l - \sum_{i \notin \mathcal{I}^v} x_i^l \right) - (Q_{d^v}(x^v) - L_{d^v}) (|\mathcal{I}^v| - 1) + L_{d^v} - U$$

Letting  $(\mathcal{I}^v)^C$  represent the complement of  $\mathcal{I}^v$  to  $\mathcal{I}$  we have that

$$-|(\mathcal{I}^v)^C| \leq \sum_{i \in \mathcal{I}^v} x_i^l - \sum_{i \notin \mathcal{I}^v} x_i^l \leq |\mathcal{I}^v| - 1$$

Given this, the inequality reduces to  $\mu^l \geq \alpha + L_{d^v} - U$  wherein  $\alpha$  is such that

$$-(Q_{d^v}(x^v) - L_{d^v})(|\mathcal{I}| - 1) \leq \alpha \leq 0$$

By subtraction of  $U$  the inequality is always satisfied, even for  $\alpha = 0$ .

When  $d^l = d^v$ , given that  $x^v \neq x^l$  we have  $\mathbb{1}_{d^v}(x^l) = 1$  and the inequality reduces to

$$\mu \geq (Q_{d^v}(x^v) - L_{d^v}) \left( \sum_{i \in \mathcal{I}^v} x_i^l - \sum_{i \notin \mathcal{I}^v} x_i^l \right) - (Q_{d^v}(x^v) - L_{d^v}) (|\mathcal{I}^v| - 1) + L_{d^v}$$

Then, the inequality becomes  $\mu \geq \alpha + L_{d^v}$  which, given the definition of  $L_{d^v}$ , and that  $\alpha \leq 0$ , is always satisfied for distribution  $d^v$ .  $\square$

### 3.2.3 Reformulation

In this section we present a reformulation of (2) in terms of the feasibility and optimality cuts just proposed for the case of second stage problems that take the form of MILPs. We also prove that an optimal solution to this reformulation can be used to derive an optimal solution to (2) with the same objective function value. The reformulation is as follows.

$$\begin{aligned} \min_{x \in \mathcal{X}} c^\top x + \mu & \tag{14} \\ \sum_{i \in \mathcal{I}^v} x_i - \sum_{i \notin \mathcal{I}^v} x_i & \leq |\mathcal{I}^v| - 1 & \forall d \in \mathcal{D}, x^v \in \mathcal{X}_d \setminus \mathcal{K}_d, \\ \mu & \geq (Q_{d^v}(x^v) - L_{d^v}) \left( \sum_{i \in \mathcal{I}^v} x_i - \sum_{i \notin \mathcal{I}^v} x_i \right) \\ & \quad - (Q_{d^v}(x^v) - L_{d^v}) (|\mathcal{I}^v| - 1) \\ & \quad + L_{d^v} - U (1 - \mathbb{1}_{d^v}(x)) & \forall d \in \mathcal{D}, x^v \in \mathcal{X}_d \cap \mathcal{K}_d. \end{aligned}$$

Given assumption A4 that  $\mathcal{X} = \{0, 1\}^{n_1}$ , the following result holds.

**Lemma 12.** *There exists only finitely many cuts (12) and (13).*

*Proof.* We first observe that  $|\mathcal{X}| \leq 2^{n_1}$ . We next note that there can only be one cut (13) for each solution in  $\mathcal{X}$  that does not induce an infeasible second stage problem and one cut (12) for each solution in  $\mathcal{X}$  that does. Thus, there can be at most  $2^{n_1}$  cuts in total.  $\square$

The analog of Proposition 7 and Theorem 8 for the case of second stage subproblems that are MILPs can be proven in a similar manner to those results. This in turn proves that an L-Shaped method that iteratively adds cuts of the form (12) and/or (13) and does not remove them will ultimately produce an optimal solution to (2) as well as a proof of its optimality.

## 3.3 Master problem formulations

An indicator function  $\mathbb{1}_d(x)$  can be represented in (4) by extending it to include binary variables  $\delta_d$ ,  $\forall d \in \mathcal{D}$ , as well as adding constraints that ensure  $x \in \mathcal{X}_d \iff \delta_d = 1$ . This approach may be inconvenient in applications where the number of distributions  $|\mathcal{D}|$  is exponential in the size of the problem, as the ones we present in Section 4. We will now show

that, under two sets of general conditions, it is possible to formulate (4) and, consequently, the cuts without indicator functions  $\mathbb{1}_d(x)$ . The first set of conditions is when the  $x$  variables can be continuous or integer. The second set of conditions is when the  $x$  variables are binary.

Consider the case in which  $\mathcal{X} = \{x \in K^{n_1} | l^i \leq a_i^\top x \leq b^i, i = 1, \dots, m\}$  with  $a_i \in \mathbb{R}^{n_1}$ ,  $l^i, b^i \in \mathbb{R}$  and  $K \in \{\mathbb{R}, \mathbb{Z}\}$ . In addition, assume that for  $i = 1, \dots, m$ , there exist constants  $l^{ij}$  and  $b^{ij}$  for  $j = 1, \dots, J$  such that  $l^{i1} = l^i$  and  $b^{iJ} = b^i$  and  $[l^{ij}, b^{ij}] \cap [l^{ij'}, b^{ij'}] = \emptyset, j, j' \in \{1, \dots, J\}, j \neq j'$ . That is, each product  $a_i^\top x$  may be in at most one out of  $J$  disjoint boxes. Let variables  $v_{ij} \in \{0, 1\}, i = 1, \dots, m, j = 1, \dots, J$  indicate in which box the product  $a_i^\top x$  falls, that is

$$l^{ij} \leq a_i^\top x \leq b^{ij} \iff v_{ij} = 1$$

Relatedly, we let the function  $j(i, d) : \{1, \dots, m\} \times \mathcal{D} \rightarrow \{1, \dots, J\}$  return the index  $j$  such that  $\mathcal{X}_d = \{l^{i,j(i,d)} \leq a_i^\top x \leq b^{i,j(i,d)}, i = 1, \dots, m\}$ . Then, given the variables  $v_{ij}$  and the function  $j(i, d)$ , the implication " $x \in \mathcal{X}_d \implies$ " can be enforced by rewriting, e.g., (8), as follows

$$\mu \geq \sum_{s \in \mathcal{S}_d} \pi_{sd} \rho_s^v (h_{sd} - T_{sd}x) - U(m - \sum_{i=1}^m v_{i,j(i,d)})$$

Here, if all the bounds implied by distribution  $d$  apply, the last term becomes zero, otherwise it reduces to an upper bound. This approach allows us to reformulate (4) with  $mJ$  binary variables  $v$  instead of the  $|\mathcal{D}| = J^m$  binary variables  $\delta_d$ . In a similar manner it is possible to rewrite cuts (6) and (13).

Consider now the case in which  $\mathcal{X} = \{0, 1\}^{n_1}$ . Then each  $\mathcal{X}_d \subseteq \mathcal{X}$  is itself a finite collection of points. Recall that the sets  $\mathcal{X}_d$  form a partition of  $\mathcal{X}$ . Let  $0 = n^{(0)} < 1 < n^{(1)} < n^{(2)} < \dots \leq n^{(T)} = n_1$  be integers that split the decision vector  $x$  into  $T \geq 1$  segments  $(x_{n^{(0)}+1}, \dots, x_{n^{(1)}}), (x_{n^{(1)}+1}, \dots, x_{n^{(2)}}), \dots, (x_{n^{(T-1)}+1}, \dots, x_{n^{(T)}})$ . Consider a set  $\mathcal{K}_t$  of exhaustive conditions that may verify in segment  $t$  (see the example in Appendix A). We have  $|\mathcal{D}| = \prod_{t=1}^T |\mathcal{K}_t|$  and we let  $k(t, d)$  be the condition that must verify on segment  $t$  for  $x$  to be a member of  $\mathcal{X}_d$ . Then  $\mathcal{X}_d = \{x \in \{0, 1\}^{n_1} | \mathbb{1}_{t,k(t,d)}(x) = 1, t = 1, \dots, T\}$  where  $\mathbb{1}_{t,k}(x)$  is an indicator function that takes value one if condition  $k$  is verified on segment  $t$ , zero otherwise. Then, let  $v_{tk} \in \{0, 1\}$  take value one if and only if condition  $k \in \mathcal{K}_t$  on segment  $t$  holds. We can rewrite (8) as follows

$$\mu \geq \sum_{s \in \mathcal{S}_d} \pi_{sd} \rho_s^v (h_{sd} - T_{sd}x) - U \left( T - \sum_{t=1}^T v_{t,k(t,d)} \right)$$

In a similar manner it is possible to rewrite cuts (6) and (13). Thus, we use  $\mathcal{O}(T \times \max_t |\mathcal{K}_t|)$  variables  $v_{tk}$  instead of the  $|\mathcal{D}| = \prod_{t=1}^T |\mathcal{K}_t|$  variables  $\delta_d$ .

### 3.4 A numerical example

We next present a numerical example to illustrate the steps performed by the proposed method. Consider the following two-stage stochastic program

$$\min x + \sum_{d=1}^2 \mathbb{1}_d(x) \mathbb{E}_{\mathbb{P}_d} Q(x, \xi)$$

$$x \geq 0$$

in which the random variable  $\xi$  follows one of two discrete probability distributions,  $P_1$  and  $P_2$ , depending on which of the sets  $\mathcal{X}_1, \mathcal{X}_2$  contains the value of  $x$ . These distributions are presented in Table 1 along with the definitions of the sets  $\mathcal{X}_1, \mathcal{X}_2$ . Each distribution contains exactly two realizations. Relatedly, we let  $\mathbb{1}_1(x), \mathbb{1}_2(x)$  denote the characteristic functions of  $\mathcal{X}_1, \mathcal{X}_2$ , respectively. (Observe, that this problem could be trivially solved by solving two ordinary stochastic programs over  $\mathcal{X}_1$  and  $\mathcal{X}_2$ , respectively. However, we believe its simplicity is useful in an illustrative example).

Table 1: Probability distributions for  $\xi$ .

| $P_1 : x \in \mathcal{X}_1 = [0.5, 3]$ |                 | $P_2 : x \in \mathcal{X}_2 = [3.5, 10]$ |                 |
|--|-----------------|---|-----------------|
| $\xi_{1,\cdot}$                        | $\pi_{1,\cdot}$ | $\xi_{2,\cdot}$                         | $\pi_{2,\cdot}$ |
| 4                                      | 0.7             | 10                                      | 0.3             |
| 12                                     | 0.3             | 18                                      | 0.7             |

Regarding the recourse function, we have

$$Q(x, \xi) = \min_{y_1, y_2 \geq 0} \{y_1 + 2y_2 | y_1 + y_2 \geq 2 + x, y_1 \geq \xi - x\}$$

In terms of the general notation used in (1b) we have  $h = [2 \ \xi]^\top, T = [-1 \ 1]^\top, W = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$  and  $q = [1 \ 2]^\top$ . One can observe that  $Q(x, \xi)$  is convex and piecewise linear. It also admits closed form solution. Namely,

$$Q(x, \xi) = \begin{cases} \xi - x, & \text{if } x \leq (\xi - 2)/2 \\ x + 2, & \text{if } x > (\xi - 2)/2 \end{cases}$$

The recourse function  $Q(x) = \sum_{d=1}^2 \mathbb{1}_d(x) \mathbb{E}_{P_d} Q(x, \xi)$  is depicted in Figure 1a.

One can also observe that  $Q(x, \xi)$  is a linear program and thus the proposed method will generate valid inequalities of the form  $\mu \geq \sum_{s \in \mathcal{S}_{d^v}} \pi_{sd^v} (\rho_s^v)^\top (h_{sd^v} - T_{sd^v} x) - U(1 - \mathbb{1}_{d^v}(x))$  that were presented in Section 3.1. The data elements  $\rho_s^v$  represent extreme points associated with the feasible region of the dual of the second stage problem for scenario  $s$  of distribution  $d^v$ , which is as below.

$$\begin{aligned} Q(x, \xi) = \max \quad & (2 + x)\rho_1 + (\xi - x)\rho_2 \\ & \rho_1 + \rho_2 \leq 1 \\ & \rho_2 \leq 2 \\ & \rho_1, \rho_2 \geq 0. \end{aligned}$$

The value  $U$  in these valid inequalities is an upper bound on the greatest difference between the largest and smallest values of the recourse function across the potential distributions. In our case  $U = 12.5$  is a safe value.



To formulate the RMP we let  $\delta_i$  denote whether or not  $x \in \mathcal{X}_i$ . We observe that the implication of the domain of  $x$  implied by the  $\delta_i$  variables can be modeled with the constraint  $0.5\delta_1 + 3.5\delta_2 \leq x \leq 3\delta_1 + 10\delta_2$ . We can thus formulate the initial RMP as follows

$$\min x + \mu \tag{15}$$

$$0.5\delta_1 + 3.5\delta_2 \leq x \leq 3\delta_1 + 10\delta_2, \tag{16}$$

$$\delta_1 + \delta_2 = 1, \tag{17}$$

$$\delta_1, \delta_2 \in \{0, 1\}, \tag{18}$$

$$\mu \geq 0. \tag{19}$$

**Iteration 1** At this iteration, the solution to RMP is  $(x, \mu, \delta_1, \delta_2) = (0.5, 0, 1, 0)$ . As  $x \in \mathcal{X}_1$ , the method solves subproblems associated with distribution  $\mathbf{P}_1$ . We observe that  $Q(.5, \xi_{1,1})$  and  $Q(.5, \xi_{1,2})$  have the same optimal solution to their duals, which is  $(\rho_1, \rho_2) = (0, 1)$ . The cut generated is

$$\mu \geq 6.4 - x - 12.5(1 - \delta_1) \tag{20}$$

which is depicted in Figure 1b.

**Iteration 2:** At this iteration the method solves the RMP (15)-(19) along with the constraint (20) to find the optimal solution  $(x, \mu, \delta_1, \delta_2) = (3.5, 0, 0, 1)$ . As  $x \in \mathcal{X}_2$ , the method solves subproblems associated with distribution  $\mathbf{P}_2$ . We observe that  $Q(3.5, \xi_{2,1})$  and  $Q(3.5, \xi_{2,2})$  have the same optimal solution to their duals, which is  $(\rho_1, \rho_2) = (0, 1)$ . The optimality cut generated is

$$\mu \geq 15.6 - x - 12.5(1 - \delta_2) \tag{21}$$

which is depicted in Figure 1c.

**Iteration 3:** At this iteration the method solves the RMP (15)-(20) along with (21) to find the optimal solution  $(x, \mu, \delta_1, \delta_2) = (3, 3.4, 1, 0)$ . Given that  $x \in \mathcal{X}_1$ , the method solves the subproblems  $Q(3, \xi_{1,1})$  and  $Q(3, \xi_{1,2})$  associated with distribution  $\mathbf{P}_1$ . The optimal dual solutions are  $(\rho_1, \rho_2) = (1, 0)$  for the first scenario and  $(\rho_1, \rho_2) = (0, 1)$  for the second. The optimality cut is

$$\mu \geq 5 + .4x - 12.5(1 - \delta_1), \tag{22}$$

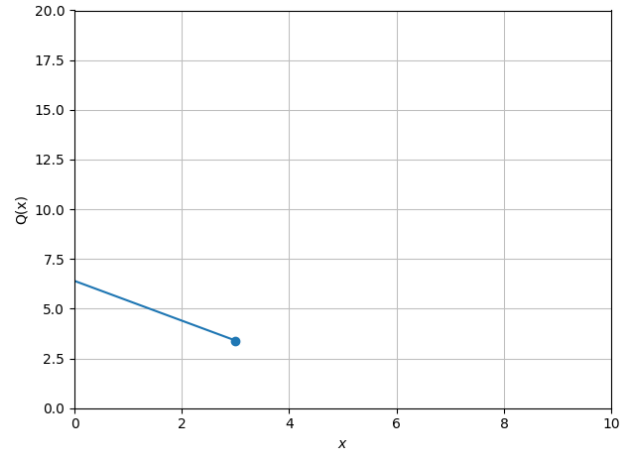
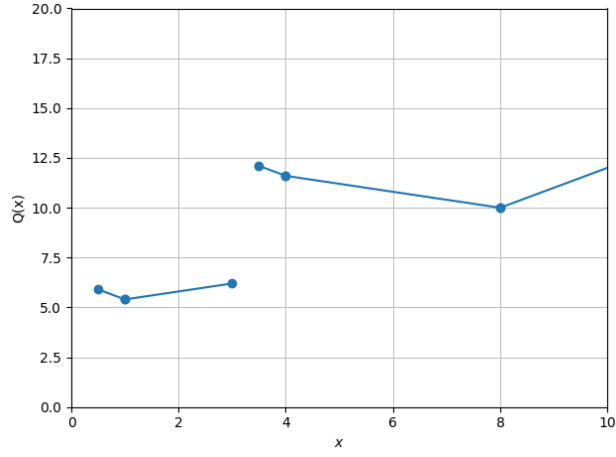
which is depicted in Figure 1d.

**Iteration 4:** At this iteration the method solves the RMP (15)-(21) along with (22) to find the optimal solution  $(x, \mu, \delta_1, \delta_2) = (1, 5.4, 1, 0)$ . By evaluating the subproblems associated with distribution  $\mathbf{P}_1$  the method can conclude that this solution is optimal.

In this example, the method converged having generated only one inequality to approximate  $\mathbb{E}_{\mathbf{P}_2} Q(x, \xi)$ . In other words, with only one optimality cut the method was able to rule out the possibility of  $\mathcal{X}_2$  containing the optimal solution. As we will see in Section 5, this behavior is not unique to this example. Namely, we will see that the method often generates fewer than two inequalities per distribution.

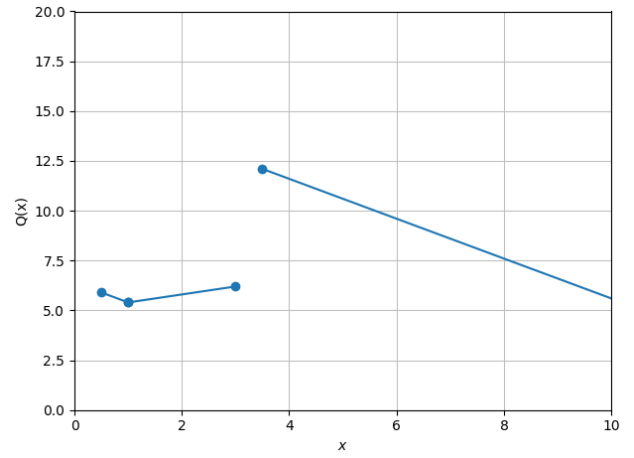
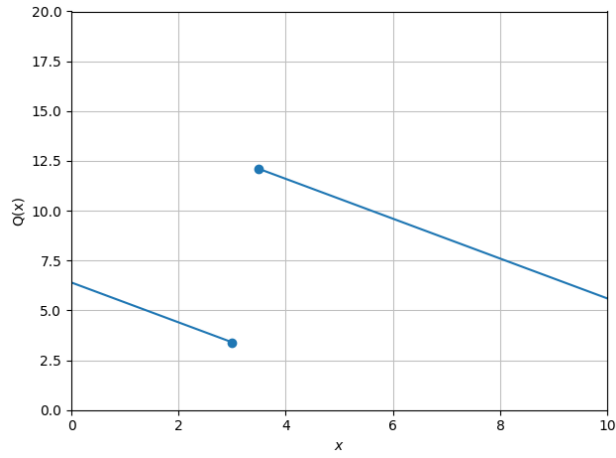
### 3.5 Distribution-independent valid inequalities

The proposed extension of the L-Shaped method requires adding to the master problem distribution-specific inequalities. We next present a class of distribution-independent valid



(a)  $Q(x) = \sum_{d=1}^2 \mathbb{1}_d(x) \mathbb{E}_{P_d} Q(x, \xi)$

(b) Optimality cuts after first iteration



(c) Optimality cuts after second iteration

(d) Optimality cuts after third iteration

Figure 1: Recourse function and cuts added.

inequalities. These are, in general, not tight and do not guarantee convergence when used alone. However, they may provide non-trivial lower bounds that facilitate convergence.

Recall that the recourse function of (2) is defined as

$$Q(x) = \sum_{d \in \mathcal{D}} \mathbb{1}_d(x) \mathbb{E}_{\mathbf{P}_d} Q(x, \xi_{sd}).$$

The following proposition paves the way to a set of distribution-independent valid inequalities.

**Proposition 13.** *For all  $x \in \mathcal{X}$  it holds that*

$$Q(x) \geq \sum_{d \in \mathcal{D}} \sum_{s \in \mathcal{S}_d} \pi_{sd} L^d(x, \xi_{sd})$$

where, for given  $d$  and  $s \in \mathcal{S}_d$ ,  $L^d(x, \xi_{sd})$  is the linear programming relaxation to the following integer programming problem

$$\begin{aligned} Q^d(x, \xi_{sd}) = & \min_{y \in \mathcal{Y}, \tau \geq 0, \delta \in \{0,1\}} q_{sd}^\top \tau \\ & W_{sd}y = h_{sd} - T_{sd}x \\ & \delta = 1 \iff x \in \mathcal{X}_d \\ & \tau - y \leq 0 \\ & \tau - Y\delta \leq 0 \\ & \tau - y - Y\delta \geq -Y \end{aligned}$$

where  $Y \in \mathbb{R}^{n_2}$  is such that  $Y \geq y$  (element-wise) for all  $y \in \mathcal{Y}$ .

*Proof.* Under assumption A1 we have

$$Q(x) = \sum_{d \in \mathcal{D}} \mathbb{1}_d(x) \sum_{s \in \mathcal{S}_d} \pi_{sd} Q(x, \xi_{sd}) = \sum_{d \in \mathcal{D}} \mathbb{1}_d(x) \sum_{s \in \mathcal{S}_d} \pi_{sd} \min_{y \in \mathcal{Y}} \{q_{sd}^\top y \mid W_{sd}y = h_{sd} - T_{sd}x\}$$

Having  $\mathbb{1}_d(x) \geq 0$  for all  $x$  and  $d \in \mathcal{D}$ , by homogeneity we can move it inside the minimization problem obtaining

$$Q(x) = \sum_{d \in \mathcal{D}} \sum_{s \in \mathcal{S}_d} \pi_{sd} \min_{y \in \mathcal{Y}} \{q_{sd}^\top y \mathbb{1}_d(x) \mid W_{sd}y = h_{sd} - T_{sd}x\} \quad (23)$$

Let us refer to the new minimization problem as  $Q^d(x, \xi_{sd})$ . Inside each  $Q^d(x, \xi_{sd})$  the  $\mathbb{1}_d(x)$  can be implemented by a binary variable  $\delta$  such that  $\delta = 1 \iff x \in \mathcal{X}_d$ . We obtain

$$Q^d(x, \xi_{sd}) = \min_{y \in \mathcal{Y}, \delta \in \{0,1\}} \{q_{sd}^\top y \delta \mid W_{sd}y = h_{sd} - T_{sd}x, \delta = 1 \iff x \in \mathcal{X}_d\} \quad (24)$$

Observe that  $\delta$  is determined by  $x$ .

The bilinear terms in the objective can be replaced by their McCormick envelopes [22]. We obtain

$$Q^d(x, \xi_{sd}) = \min_{y \in \mathcal{Y}, \tau \geq 0, \delta \in \{0,1\}} q_{sd}^\top \tau \quad (25)$$

$$W_{sd}y = h_{sd} - T_{sd}x \quad (26)$$

$$\delta = 1 \iff x \in \mathcal{X}_d \quad (27)$$

$$\tau - y \leq 0 \quad (28)$$

$$\tau - Y\delta \leq 0 \quad (29)$$

$$\tau - y - Y\delta \geq -Y \quad (30)$$

It follows that, for all  $x \in \mathcal{X}$

$$Q(x) = \sum_{d \in \mathcal{D}} \sum_{s \in \mathcal{S}_d} \pi_{sd} Q^d(x, \xi_{sd}) \geq \sum_{d \in \mathcal{D}} \sum_{s \in \mathcal{S}_d} \pi_{sd} L^d(x, \xi_{sd})$$

where the inequality holds because  $Q^d(x, \xi_{sd}) \geq L^d(x, \xi_{sd})$  for all  $x, d$  and  $s$ .  $\square$

Next, let  $\nabla_{sd}^v$  be a subgradient to  $L^d(x, \xi_{sd})$  at  $x^v$ . Proposition 13 and convexity of  $L^d(x, \xi_{sd})$  in  $x$  suggest adding cuts

$$\mu \geq \sum_{d \in \mathcal{D}} \sum_{s \in \mathcal{S}_d} \pi_{sd} L^d(x^v, \xi_{sd}) + (\nabla_{sd}^v)^\top (x - x^v) \quad (31)$$

generated from  $L^d(\cdot, \cdot)$  upon visiting solutions  $x^v$  to (4). These cuts support the recourse function from below and create a convex lower envelop to  $Q(x)$ .

Observe that cuts (31), while general, require solving one linear program for each  $d$  and for each  $s$ . Under specific assumptions, we can generate distribution-independent inequalities by solving one subproblem per distribution.

**Proposition 14.** *Assume  $Q(x, \xi)$ , hence  $L^d(x, \xi)$ , are convex in  $\xi$ . Then, for all  $x \in \mathcal{X}$*

$$Q(x) \geq \sum_{d \in \mathcal{D}} L^d(x, \mathbb{E}_{\mathbf{P}_d}[\xi])$$

*Proof.* The result follows immediately from Jensen's inequality.  $\square$

This result allows us to solve only one subproblem per distribution instead of  $|\mathcal{S}_d|$ . Hence, (31) reduces to

$$\mu \geq \sum_{d \in \mathcal{D}} L^d(x^v, \mathbb{E}_{\mathbf{P}_d}[\xi]) + (\nabla_d^v)^\top (x - x^v) \quad (32)$$

where  $\nabla_d^v$  is a subgradient to  $L^d(x, \mathbb{E}_{\mathbf{P}_d}[\xi])$  at  $x^v$ .

Another opportunity for generating distribution-independent cuts involves examining the scenarios themselves to see if they meet certain criteria.

**Proposition 15.** *Assume there exists  $\hat{\xi} = (\hat{q}, \hat{W}, \hat{T}, \hat{h})$  such that  $Q(x, \hat{\xi}) \leq \mathbb{E}_{\mathbf{P}_d} Q(x, \xi)$  for all  $x \in \mathcal{X}$  and  $d \in \mathcal{D}$ , where  $Q(x, \xi)$  is a linear program. Let  $\rho^v$  be an optimal dual solution to  $Q(x^v, \hat{\xi})$  for some  $x^v \in \mathcal{X}$ . Then*

$$\mathbb{E}_{\mathbf{P}_{d^l}}[Q(x^l, \xi)] \geq (\rho^v)^\top (\hat{h} - \hat{T}x^l)$$

for all  $x^l \in \mathcal{X}$ .

*Proof.* Consider  $x^l \in \mathcal{X}$  which yields  $d^l$ .

$$\begin{aligned}\mathbb{E}_{\mathbf{P}_{d^l}}[Q(x^l, \xi)] &= \sum_{s \in \mathcal{S}_{d^l}} \pi_{sd^l} Q(x^l, \xi_{sd^l}) \\ &\geq Q(x^l, \hat{\xi}) = (\rho^l)^\top (\hat{h} - \hat{T}x^l) \geq (\rho^v)^\top (\hat{h} - \hat{T}x^l)\end{aligned}$$

where  $\rho^l$  and  $\rho^v$  are, respectively, optimal and feasible to the dual of  $Q(x^l, \hat{\xi})$ .  $\square$

Proposition 15 justifies adding valid inequality

$$\mu \geq (\rho^v)^\top (\hat{h} - \hat{T}x) \quad (33)$$

from  $x^v$ , if the optimality test fails. Compared to Proposition 13, Proposition 15 requires solving only one additional linear program at every iteration. However, it requires the existence and identification of a scenario  $\hat{\xi}$ , which might not be trivial. Next, we describe a special case where this task is simplified.

**Definition 1.** For all  $x \in \mathcal{X}$  the function  $Q(x, \xi) : \mathbb{R}^N \rightarrow \mathbb{R}$  is said to be monotone non-decreasing/non-increasing if for all  $i = 1, \dots, N$ , given and fixed  $\xi^j = \hat{\xi}^j \in \mathbb{R}$  for  $j \in \{1, \dots, N\} \setminus \{i\}$ , the functions  $Q(x, (\hat{\xi}^1, \dots, \hat{\xi}^{i-1}, \xi^i, \hat{\xi}^{i+1}, \dots, \hat{\xi}^N)) : \mathbb{R} \rightarrow \mathbb{R}$  are monotone non-decreasing/non-increasing.

**Proposition 16.** Assume  $Q(x, \xi)$  is convex and monotone non-decreasing in  $\xi$  for all  $x \in \mathcal{X}$ . Consider  $\hat{\xi} = \min_{d \in \mathcal{D}} \mathbb{E}_{\mathbf{P}_d}[\xi]$  where the minimum is taken component-wise. Then  $\hat{\xi}$  satisfies the definition in Proposition 15.

*Proof.* Let  $x^v \in \mathcal{X}$  be arbitrary and  $d^v$  the corresponding distribution. Then

$$\mathbb{E}_{\mathbf{P}_d}[Q(x^v, \xi)] \geq Q(x^v, \mathbb{E}_{\mathbf{P}_d}[\xi]) \geq Q(x^v, \hat{\xi})$$

Where the first inequality holds by convexity and Jensen's inequality and the second by monotonicity and by definition of  $\hat{\xi}$ .  $\square$

## 4 Application of proposed method to production planning

To assess the computational effectiveness of the proposed method, we apply it to three variants of a production planning problem under both endogenous and exogeneous uncertainty. While all three variants possess the properties required by the proposed framework, each presents a different computational challenge with respect to the solution of the second stage subproblem. We note that many of the concepts and notation are common across the three variants. We introduce these when introducing the first variant. When presenting the second and third variants we only discuss how they differ from the first.

## 4.1 Production Planning Problem - Variant 1

A manufacturer can produce a product at one or more facilities  $\mathcal{F}$ . When there is demand for a unit, it can be sold at price  $P$ . When there is not, it can be sold at price  $O < P$ . The manufacturer seeks to determine how much to make at each facility  $f \in \mathcal{F}$ , given a per unit cost of production,  $C_f$ .

For each facility  $f$ , the manufacturer chooses a production level  $l$ , from a set  $\mathcal{L}_f$ , that defines a production volume range  $[L_{fl}, U_{fl}]$ . We presume that the set  $\mathcal{L}_f$  includes a level  $l$  such that  $L_{fl} = 0$ . Furthermore, we assume that for all  $f \in \mathcal{F}$  we have  $[L_{f,l_1}, U_{f,l_1}] \cap [L_{f,l_2}, U_{f,l_2}] = \emptyset$  for all  $l_1, l_2 \in \mathcal{L}_f$ .

When planning production, the yield at each facility, and the cumulative demand of finished goods for the product are uncertain. We let  $\boldsymbol{\xi} = (\mathbf{Y}_{f_1}, \dots, \mathbf{Y}_{f_{|\mathcal{F}|}}, \mathbf{D})$  denote an  $\mathbb{R}^{|\mathcal{F}|+1}$  random vector representing the uncertain yield at the different facilities,  $\mathbf{Y}_f$ , and the cumulative demand,  $\mathbf{D}$ . The yield at a given facility can also depend on the production levels chosen.

We let  $\mathcal{D}$  denote the set of joint distributions of yields and demand. However, as uncertainty in demand is exogenous, its marginal distribution is identical for all distributions in  $\mathcal{D}$ . For each facility  $f$ , and distribution  $\mathcal{D}$ , there is a map  $l(f, d) : \mathcal{F} \times \mathcal{D} \rightarrow \cup_{f \in \mathcal{F}} \mathcal{L}_f$  that indicates the production level at facility  $f$  required for distribution  $d \in \mathcal{D}$ . It is necessary that the production level indicated by the map  $l(f, d)$  is chosen for each facility  $f$  for probability distribution  $d \in \mathcal{D}$  to accurately represent uncertainty in yields. Observe that the resulting number of distributions is  $|\mathcal{D}| = \prod_{f \in \mathcal{F}} |\mathcal{L}_f|$ .

We further assume that each probability distribution is either discrete or can be adequately represented by a finite number of scenarios. We let  $\mathcal{S}_d$  represent the set of scenarios in distribution  $d \in \mathcal{D}$ . Associated with scenario  $s \in \mathcal{S}_d$  for distribution  $d \in \mathcal{D}$  we let  $\pi_{sd}$  denote the probability that it occurs. Regarding yields at facilities, we let  $Y_{f,sd}$  represent the realization of the yield at facility  $f$  under scenario  $s \in \mathcal{S}_{pd}$  of distribution  $d \in \mathcal{D}$ . We let  $D_{sd}$  represent the realization of the demand under scenario  $s \in \mathcal{S}_d$  of distribution  $d \in \mathcal{D}$ .

To model the problem under consideration, we let  $x_f \geq 0$  represent the amount of raw materials, quoted in terms of finished goods, allocated to facility  $f \in \mathcal{F}$ . We let binary variable  $y_{fl}$  equal 1 if production level  $l \in \mathcal{L}_{fp}$  is chosen at facility  $f \in \mathcal{F}$  and 0 otherwise. The binary decision variable  $\delta_d$  represents whether distribution  $d \in \mathcal{D}$  should be used to describe yield uncertainties. Decision variables  $x := (x_f)_{f \in \mathcal{F}}$ ,  $y := (y_{fl})_{f \in \mathcal{F}, l \in \mathcal{L}_f}$  and  $\delta := (\delta_d)_{d \in \mathcal{D}}$  are first stage decision variables.

$$\min \sum_{f \in \mathcal{F}} C_f x_f - \sum_{d \in \mathcal{D}} \delta_d \left[ \sum_{s \in \mathcal{S}_d} \pi_{sd} Q(x, \xi_{sd}) \right] \quad (34a)$$

$$\text{s.t.} \quad \sum_{l \in \mathcal{L}_f} y_{fl} = 1 \quad \forall f \in \mathcal{F}, \quad (34b)$$

$$\sum_{l \in \mathcal{L}_f} L_{fl} y_{fl} \leq x_f \leq \sum_{l \in \mathcal{L}_f} U_{fl} y_{fl} \quad \forall f \in \mathcal{F}, \quad (34c)$$

$$\sum_{f \in \mathcal{F}} y_{f, l(f, d)} \geq |\mathcal{F}| \delta_d \quad \forall d \in \mathcal{D}, \quad (34d)$$

$$x_f \geq 0 \quad \forall f \in \mathcal{F}, \quad (34e)$$

$$y_{fl} \in \{0, 1\} \quad \forall f \in \mathcal{F} l \in L_f, \quad (34f)$$

$$\delta_d \in \{0, 1\} \quad \forall d \in \mathcal{D}. \quad (34g)$$

The objective (34a) seeks to maximize expected profit. Here,  $Q(x, \xi_{sd})$  represents the revenue earned by selling inventory under realization  $\xi_{sd}$  of yield and demand. Constraints (34b) ensure that a production level is chosen for each product at each facility. Constraints (34c) ensure that the amount allocated to a facility falls within the bounds defined by the chosen production level. Constraints (34d) ensure that the choice of production levels enforces the appropriate distribution.

Once yield and demand information have materialized through a scenario  $s \in \mathcal{S}_d$  for the distribution  $d \in \mathcal{D}$  indicated by  $\delta_d = 1$ , the decision maker makes second stage decisions concerning sales. The continuous variable  $w$  defines the amount of product that is sold at “full” price  $P$  while the continuous variable  $o$  represents the amount of inventory that is sold at the discounted price,  $O$ . For  $d \in \mathcal{D}$  and  $s \in \mathcal{S}_d$  the second stage decision problem is as follows.

$$Q(x, \xi_{sd}) = \max_{w, o \geq 0} \left\{ Pw + Oo \mid w + o = \sum_{f \in \mathcal{F}} Y_{fsd} x_f, w \leq D_{sd} \right\} \quad (35)$$

The constraints limit the amount of inventory that is sold by how much is available given the yields. Furthermore, the amount of inventory that is sold at full price is limited by the demand realization.

## 4.2 Production Planning Problem with Per-unit Transportation Costs - Variant 2

This problem is like the one just discussed, except that demand for finished goods occurs in different locations. Thus, for this variant, the first stage decisions are the same as in problem (34). However, the second stage problem is different as we must determine to which customer locations finished goods should be transported to meet demand.

We let  $\mathcal{Q}$  denote the set of locations wherein demand can occur. Hence,  $\xi$  is now an  $\mathbb{R}^{|\mathcal{F}|+|\mathcal{Q}|}$  random vector. We let  $D_{qsd}$  represent the demand realization for finished goods in location  $q \in \mathcal{Q}$  given distribution  $d \in \mathcal{D}$  and scenario  $s \in \mathcal{S}_d$ . We also let  $H_{fq}, f \in \mathcal{F}, q \in \mathcal{Q}$  represent the per-unit cost associated with transporting units from facility  $f$  to demand location  $q$ .

We let the continuous decision variable  $w_{fq}$  denote the amount of finished goods transported from facility  $f \in \mathcal{F}$  to demand location  $q \in \mathcal{Q}$  and  $o_f$  the amount of inventory left at facility  $f$ . The second stage problem is as follows.

$$Q(x, \xi_{sd}) = \max \sum_{f \in \mathcal{F}} \sum_{q \in \mathcal{Q}} (P - H_{fq}) w_{fq} + \sum_{f \in \mathcal{F}} O o_f \quad (36a)$$

$$\sum_{q \in \mathcal{Q}} w_{fq} + o_f = Y_{fsd} x_f \quad \forall f \in \mathcal{F}, \quad (36b)$$

$$\sum_{f \in \mathcal{F}} w_{fq} \leq D_{qsd} \quad \forall q \in \mathcal{Q}, \quad (36c)$$

$$o_f \geq 0 \quad \forall f \in \mathcal{F}, \quad (36d)$$

$$w_{fq} \geq 0 \quad \forall f \in \mathcal{F}, q \in \mathcal{Q}. \quad (36e)$$

### 4.3 Production Planning Problem with Batches and Per-vehicle Transportation Costs - Variant 3

Finally, we consider a variant of the problem that differs in two ways from those previously presented. First, production decisions in the first stage are made in terms of batches as opposed to continuous amounts. Second, while we again presume demand for finished goods occurs in different locations, we now presume that transportation costs are paid on a per-vehicle basis.

We presume a set of potential production amounts,  $\mathcal{B}_f$  for facility  $f \in \mathcal{F}$ , at most one of which must be chosen. Associated with batch  $b \in \mathcal{B}_f$  is a batch size  $Q_{bf}$ . We let the binary variable  $x_{fb} \in \{0, 1\}$  denote whether batch size  $b \in \mathcal{B}_f$  is chosen for facility  $f$ . As in the previous variants, we assume probability distributions are determined by production levels.

$$\min \sum_{f \in \mathcal{F}} \sum_{b \in \mathcal{B}_f} C_f Q_{bf} x_{fb} - \sum_{d \in \mathcal{D}} \delta_d \left[ \sum_{s \in \mathcal{S}_d} \pi_{sd} Q(x, \xi_{sd}) \right] \quad (37a)$$

$$\text{s.t.} \sum_{b \in \mathcal{B}_f} x_{fb} \leq 1 \quad \forall f \in \mathcal{F}, \quad (37b)$$

$$\sum_{l \in \mathcal{L}_f} y_{fl} = 1 \quad \forall f \in \mathcal{F}, \quad (37c)$$

$$\sum_{l \in \mathcal{L}_f} L_{fl} y_{fl} \leq \sum_{b \in \mathcal{B}_f} Q_{bf} x_{fb} \leq \sum_{l \in \mathcal{L}_f} U_{fl} y_{fl} \quad \forall f \in \mathcal{F} \quad (37d)$$

$$\sum_{d \in \mathcal{D}} \delta_d = 1 \quad (37e)$$

$$\sum_{f \in \mathcal{F}} y_{f,l(f,d)} \geq |\mathcal{F}| \delta_d \quad \forall d \in \mathcal{D} \quad (37f)$$

$$x_{fb} \in \{0, 1\} \quad \forall f \in \mathcal{F}, b \in \mathcal{B}_f \quad (37g)$$

$$y_{fl} \in \{0, 1\} \quad \forall f \in \mathcal{F}, l \in \mathcal{L}_f \quad (37h)$$

$$\delta_d \in \{0, 1\} \quad \forall d \in \mathcal{D} \quad (37i)$$

Let  $K$  be the capacity of the vehicles and  $G_{fq}$  the cost of a vehicle traveling from facility  $f \in \mathcal{F}$  to customer location  $q \in \mathcal{Q}$ . Thus, we define integer decision variables  $v_{fq}$  that model the number of vehicles that travel from facility  $f \in \mathcal{F}$  to location  $q \in \mathcal{Q}$ . The second stage problem is as follows.

$$Q(x, \xi_{sd}) = \max \sum_{f \in \mathcal{F}} \left( \sum_{q \in \mathcal{Q}} (P w_{fq} - G_{fq} v_{fq}) + O_o f \right) \quad (38a)$$



$$\text{s.t. } \sum_{q \in \mathcal{Q}} w_{fq} + o_f = Y_{fsd} \sum_{b \in \mathcal{B}_f} Q_{bf} x_{fb} \quad \forall f \in \mathcal{F}, \quad (38b)$$

$$\sum_{f \in \mathcal{F}} w_{fq} \leq D_{qsd} \quad \forall q \in \mathcal{Q}, \quad (38c)$$

$$w_{fq} \leq K v_{fq} \quad \forall f \in \mathcal{F}, q \in \mathcal{Q}, \quad (38d)$$

$$o_f \geq 0 \quad \forall f \in \mathcal{F}, \quad (38e)$$

$$w_{fq} \geq 0 \quad \forall f \in \mathcal{F}, q \in \mathcal{Q}, \quad (38f)$$

$$v_{fq} \in \mathbb{Z}^+ \quad \forall f \in \mathcal{F}, q \in \mathcal{Q}. \quad (38g)$$

## 5 Computational study

To assess the effectiveness of the proposed, general, method we use it to solve instances of each variant of the production planning problem presented in Section 4. We compare the performance of the method with an off-the-shelf optimization solver solving instances of the extensive linearized form of the three variants of the problem. The linearized versions are obtained by linearizing the bilinear terms in the objective functions using [22, 2]. In Appendix B we present the linearized extensive form of the first variant. Linearized extensive forms of the other variants are similar.

We begin this section presenting the instances used in the experiments that form the basis of our computational analyses and the computational environment in which the experiments were run. We finish the section by discussing and analyzing the results of those experiments.

### 5.1 Instances

All instances underlying our computational study are randomly generated, albeit with a process inspired by a motivating industrial application. Thus, we will next describe the parameter values used when generating those instances. We note that the parameter values defining an instance of the second or third variants include the parameter values required by the first. We also note that all instances are available upon request.

We recall that there are two primary parameters,  $|\mathcal{F}|$  and  $|\mathcal{L}_f|$ , that (partially) define a set of instances for all three variants. Regarding the first parameter, for the first variant we consider instances that contain between two and seven facilities. For the second and third variants we consider instances that contain between two and five facilities. Regarding the second parameter, each facility has the same number of production levels, which ranges from two to five. For the second and third variants there is another parameter, which is the number of customer locations,  $|\mathcal{Q}|$ . In all instances of variants two or three we considered five such locations. Finally, for the third variant another parameter is the number of production batches,  $|\mathcal{B}|$ . In all instances of variant three we considered five batches.

Regarding the economics underlying the instances, the cost  $C_f$  of manufacturing in facility  $f \in \mathcal{F}$  was randomly drawn from the range  $[60, 80]$ . The price  $P$  at which product is sold was randomly drawn from the range  $[125, 185]$ . As such, the expected gross margin on products sold is 54%. Finally, the revenue  $O$  recovered from left-over inventory was randomly drawn from the range  $[15, 40]$ . Note this ensures that in each instance there is a loss on products

Table 2: Average sizes of the linearized extensive forms by variant for the smallest and largest number of distributions.

| Variant | $ \mathcal{D} $ | # variables |             | # int. variables |             | # constraints |               |
|---------|-----------------|-------------|-------------|------------------|-------------|---------------|---------------|
|         |                 | min         | max         | min              | max         | min           | max           |
| V1      | 4               | 410         | 40,010      | 8                | 8           | 813           | 80,013        |
| V1      | 78,125          | 78,90,667   | 781,328,167 | 78,160           | 78,160      | 15,703,154    | 1,562,578,154 |
| V2      | 4               | 2,414       | 240,014     | 12               | 12          | 4,313         | 430,013       |
| V2      | 3,125           | 4,703,155   | 468,765,655 | 15,650           | 15,650      | 7,815,646     | 781,253,146   |
| V3      | 4               | 4,422       | 440,022     | 2,022            | 200,022     | 8,313         | 830,013       |
| V3      | 3,125           | 8,609,425   | 859,390,675 | 3,921,925        | 390,640,675 | 15,628,146    | 1,562,503,146 |

manufactured but not sold at full price. Turning to transportation costs, we randomly generated distances in miles between each facility and each customer location based on a uniform distribution over the interval  $[75, 300]$ . Per-unit costs,  $H_{fq}, f \in \mathcal{F}, q \in \mathcal{Q}$ , were computed based on those distances and a cost of \$0.10 per mile. Per-vehicle costs,  $G_{fq}, f \in \mathcal{F}, q \in \mathcal{Q}$ , were computed similarly albeit based on a per-mile cost of \$1.55.

We next discuss the scenario generation process. We first recall that associated with each instance of any of the three variants there are  $|\mathcal{D}| = \prod_{f \in \mathcal{F}} |\mathcal{L}_f|$  distributions. The number of distributions in our instances for variant 1 ranged from  $2^2 = 4$  to  $5^7 = 78,125$ . For variants 2 and 3, as those instances consist of at most five facilities, the number of distributions ranged from  $2^2 = 4$  to  $5^5 = 3,125$ .

For the uncertainty in production yields, we consider a single joint distribution for each combination of production levels at different facilities. This joint distribution is composed of marginal normal distributions truncated at 0.25 and 1.0, with different parameters at different facilities as well as at different production levels within the same facility. Generally speaking, we considered individual yield distributions such that for a given mean, the larger the production level the smaller the standard deviation in production yield. Demands were also presumed to follow a normal distribution. In the case of variants 2 and 3, in which demands occur at customer locations, we presumed the demand distributions at different customer locations were independent. However, the mean of the distribution varies across customer locations.

Scenarios were generated by random sampling from the given joint distribution. We considered  $S = 5, 10, 15, \dots, 45, 50$  scenarios per distribution. As a result, an instance consists of  $S \prod_{f \in \mathcal{F}} |\mathcal{L}_f|$  scenarios. Table 2 reports statistics on the size of the extensive linearized formulations of the three variants for the smallest and largest number of distributions considered.

For all three variants we define a class of instances by the triplet  $(|\mathcal{F}|, |\mathcal{L}|, S)$ , wherein  $|\mathcal{L}_f| = |\mathcal{L}| \ \forall f \in \mathcal{F}$ . We note that for each variant we generate five instances for each class. In total, we consider 880 instances of the first variant, 800 instances of the second, and 800 instances of the third.

## 5.2 Computational environment

All code supporting both the proposed method presented in Section 3 as well as solving the extensive linearized versions is written in Java. All optimization models solved by that code were done so using CPLEX 20.1 [17]. The proposed method was implemented in a branch-and-benders-cut-type fashion. Namely, the relaxed master problem is solved by a branch-and-bound-based method and inequalities generated at nodes of the tree at which integer solutions are found. We also note that when solving instances with integer second stage problems (e.g. (11)) we also solve the linear programming relaxation of the subproblem in order to generate duality-based cuts (8). We note that the three problem variants considered have relatively complete recourse and thus feasibility cuts need not be generated.

All tests were performed on computing nodes equipped with 40 CPUs and 188 GB memory. However, no parallelization techniques were used when executing our L-Shaped algorithm for any of the three variants. Each of those experiments was run using a single thread. However, when solving extensive linearized forms, CPLEX was configured to use its default deterministic parallel search strategy for the branch-and-bound method, which used up to 32 threads. In all experiments we used an optimality tolerance of 0.0001 and a time limit of 1,800 seconds.

## 5.3 Performance of the L-shaped method

We compare the performance of two methods. The first is the proposed extension of the L-shaped method, which we label **LS** in the subsequent tables. We note that we initially report results based on experiments in which the distribution-independent valid inequalities presented in Section 3.5 were not used. We end this section with a study of their impact. The second method involves solving the extensive linearized forms with CPLEX, which we label **CPLEX** in the subsequent tables. To compare the computational effectiveness of the two methods we focus on four statistics. The first is the percentage of instances that method could solve to the desired tolerance and within the given time limit. The second is the average time, in seconds, that method required to solve those instances. The third is the average optimality gap reported by the method at termination for the instances it could not solve, but for which it completed without exceeding the memory limit. The fourth is the percentage of instances for which the method terminated because it exceeded the memory allotted. We report these statistics, by variant, in Table 3.

We see that while solving the deterministic equivalent with CPLEX was fairly effective for the first variant, it solved less than half the instances of the second and third variants. Further, CPLEX exceeded the allotted memory when solving many instances. Finally, when CPLEX did not exceed the allotted memory, it reported optimality gaps over 30% for the instances it did not solve.

The proposed method was able to solve all instances of variants 1 and 3 and 96.25% of instances of variant 2. For the instances the method could solve, it required, on average, less than six minutes to do so. We further note that all the instances of variant 2 that the proposed method did not solve were instances involving 3,125 distributions. Unlike CPLEX, the proposed method rarely required more than 1 megabyte of memory during its execution. We conclude from these results that the proposed method is superior to the benchmark, at

Table 3: Comparison of methods on three variants. Optimality gaps are computed as  $100 \cdot (|\text{best\_primal} - \text{best\_dual}|)/|\text{best\_dual}|$ .

| Variant 1    | % of instances solved | Average time to solve (sec.) | % unsolved exceeded memory | % unsolved did not exceed memory | Avg. gap unsolved |
|--------------|-----------------------|------------------------------|----------------------------|----------------------------------|-------------------|
| <b>LS</b>    | 100.00%               | 0.04                         | N/A                        | N/A                              | N/A               |
| <b>CPLEX</b> | 90.57%                | 12.48                        | N/A                        | N/A                              | 9.43%             |
| Variant 2    | % of instances solved | Average time to solve (sec.) | % unsolved exceeded memory | % unsolved did not exceed memory | Avg. gap unsolved |
| <b>LS</b>    | 96.25%                | 95.86                        | N/A                        | 3.75%                            | 95.62%            |
| <b>CPLEX</b> | 31.75%                | 462.56                       | 33.00%                     | 35.25%                           | 98.59%            |
| Variant 3    | % of instances solved | Average time to solve (sec.) | % unsolved exceeded memory | % unsolved did not exceed memory | Avg. gap unsolved |
| <b>LS</b>    | 100.00%               | 355.81                       | N/A                        | N/A                              | N/A               |
| <b>CPLEX</b> | 49.88%                | 332.67                       | 21.50%                     | 28.63%                           | 30.69%            |

least for this class of problem. Thus, we next focus on analyzing the performance of the proposed method.

We next analyze three more statistics regarding the performance of the method. The first is the number of nodes in the branch-and-bound tree the method explored. The second is the number of optimality cuts the method generated. The third is the time the method spent generating cuts. We report averages of these statistics, by variant, in Table 4. We note we only report these statistics for instances the method was able to solve to optimality.

Table 4: Performance statistics by variant.

| Variant | Branch and bound nodes | Optimality cuts | Time generating cuts (sec.) |
|---------|------------------------|-----------------|-----------------------------|
| 1       | 0.12                   | 6.22            | 0.02                        |
| 2       | 446.38                 | 254.91          | 95.49                       |
| 3       | 3.66                   | 4.17            | 355.79                      |

Comparing Table 3 and Table 4 we observe that for all three variants the method spent the vast majority of its time generating cuts. The method was often able to converge at the root node, and generate few cuts, when solving instances of the first variant. However, the more complicated second stage subproblem associated with the second variant required it to perform a more in-depth search of the branch-and-bound tree as well as generate far more cuts during that search. Comparing variants two and three we observe that the method did not need to generate many optimality cuts to converge. However, the time required to

generate that smaller set of cuts was much greater.

Given that the proposed method solved instances of the first variant nearly instantaneously, we next report on the time it required to solve the other two variants in more detail. We first note that the method was able to solve the majority (73.12%) of instances of the second variant within 50 seconds. While instances of the third variant took more time, the method was able to solve 81.23% within ten minutes. On the other hand, the method needed more than 1,000 seconds to solve 9.14% of the instances of the third variant.

We next seek to understand what instance parameters impact the solution time of the proposed method when applied to instances of variant two or three. To that effect, we first consider the average solve time, averaged over instances that share the same number of distributions ( $|\mathcal{D}|$ , Table 5). We see that while the solution time for instances of variant two increases as the number of distributions increases, the same trend is not present for instances of variant three. Given that few branch and bound nodes were explored when solving instances of variant three (Table 4), this is likely due to few master problem solutions needing to be evaluated before a provably optimal solution was found.

Table 5: Average time to solve by  $|\mathcal{D}|$ .

(a)  $|\mathcal{D}| = 4 - 64$ .

| Variant | $ \mathcal{D} $ |        |        |        |        |        |        |        |
|---------|-----------------|--------|--------|--------|--------|--------|--------|--------|
|         | 4               | 8      | 9      | 16     | 25     | 27     | 32     | 64     |
| 2       | 2.45            | 4.02   | 4.11   | 7.21   | 9.71   | 10.34  | 15.43  | 23.11  |
| 3       | 267.24          | 342.81 | 507.83 | 410.81 | 540.42 | 325.39 | 370.93 | 340.16 |

(b)  $|\mathcal{D}| = 81 - 3125$ .

| Variant | $ \mathcal{D} $ |        |        |        |        |        |        |
|---------|-----------------|--------|--------|--------|--------|--------|--------|
|         | 81              | 125    | 243    | 256    | 625    | 1024   | 3125   |
| 2       | 31.20           | 41.59  | 95.31  | 93.97  | 272.15 | 589.66 | 672.06 |
| 3       | 277.60          | 303.63 | 380.61 | 292.54 | 272.31 | 341.59 | 308.33 |

To further understand the behavior of the method when applied to variant two we present in Table 6 the average number of optimality cuts generated when solving instances involving the same number of distributions,  $|\mathcal{D}|$ . We see that the method generates more cuts for instances that involve more distributions. However, the method regularly generates less than two optimality cuts per distribution. This suggests that the method is often able to generate a single optimality cut that provides sufficient evidence that first stage variable values from a given subset  $\mathcal{X}_d$  will lead to worse objective values.

This pattern appears to be broken for the instances with 3125 distributions. Recall that the results in Table 6 are obtained using instances of variant two. In Table 3 it is evident 3.75% of these instances were not solved to optimality. These are precisely some of the instances with 3125 distributions. For these instances the method did not manage to add all the cuts necessary for convergence within the time limit.

Next, we report in Table 7 the performance of the method on instances with the same number of scenarios per distribution ( $S$ ). We see that for both variants the solution time increases as the number of scenarios per distribution increases.

We next turn our attention to the impact of the distribution-independent inequalities

Table 6: Average number of optimality cuts in variant two by  $|\mathcal{D}|$ .

(a)  $|\mathcal{D}| = 4 - 64$

| $ \mathcal{D} $ |       |       |       |       |       |       |       |
|-----------------|-------|-------|-------|-------|-------|-------|-------|
| 4               | 8     | 9     | 16    | 25    | 27    | 32    | 64    |
| 6.74            | 11.52 | 11.82 | 20.98 | 30.40 | 32.00 | 41.16 | 72.44 |

(b)  $|\mathcal{D}| = 81 - 3125$

| $ \mathcal{D} $ |        |        |        |        |         |         |
|-----------------|--------|--------|--------|--------|---------|---------|
| 81              | 125    | 243    | 256    | 625    | 1024    | 3125    |
| 89.14           | 134.06 | 254.08 | 267.98 | 636.74 | 1039.58 | 2060.34 |

Table 7: Average time to solve by  $S$ .

| Variant | $S$  |       |       |        |        |        |        |        |        |        |
|---------|------|-------|-------|--------|--------|--------|--------|--------|--------|--------|
|         | 5    | 10    | 15    | 20     | 25     | 30     | 35     | 40     | 45     | 50     |
| 2       | 8.03 | 31.00 | 76.82 | 113.38 | 55.11  | 84.47  | 100.66 | 132.75 | 163.32 | 203.38 |
| 3       | 8.52 | 39.54 | 84.94 | 142.76 | 197.78 | 292.01 | 463.82 | 608.11 | 879.34 | 841.32 |

presented in Section 3.5. To do so, we executed the proposed L-Shaped method on variants two and three again, albeit with the additional valid inequalities (33) that are based on Proposition 16. We note that given that the subproblems are maximization problems, and concave, the inequality was derived by replacing  $\min_{d \in \mathcal{D}} \mathbb{E}_{\mathbf{P}_d}[\xi]$  with  $\max_{d \in \mathcal{D}} \mathbb{E}_{\mathbf{P}_d}[\xi]$ . We further note that for variant three Proposition 16 was applied to the LP relaxation of the integer second stage problem. Also, for both variants, distribution-independent inequalities were added only at the root node of the branch-and-bound tree.

We first note that the use of the distribution-independent inequalities did not lead to a significant improvement in the performance of the method when applied to instances of variant two. Considering variant three, we recall that the method executed without valid inequalities (33) solved every instance, see Table 3. We observe that the method exhibited the same performance with the use of these additional inequalities. Further, we report in Table 8 the average time the method required with and without these inequalities, for instances with large numbers of distributions. We see that, for instances involving large number of distributions, the use of these inequalities can significantly decrease the time required by the method to converge to an optimal solution.

Table 8: Average time for **LS** to solve the largest instances of variant three with and without valid inequality (33).

| $ \mathcal{D} $ | With Dist-ind. VI (sec.) | Without Dist-ind. VI (sec.) | $\Delta$ (%) |
|-----------------|--------------------------|-----------------------------|--------------|
| 243             | 346.03                   | 380.61                      | -9.99        |
| 256             | 255.06                   | 292.53                      | -14.68       |
| 625             | 241.21                   | 272.30                      | -12.89       |
| 1024            | 309.43                   | 341.58                      | -10.39       |
| 3125            | 263.13                   | 308.32                      | -17.17       |

## 6 Conclusions and future work

In this paper, we considered two-stage stochastic programs in which uncertainty is endogenous. Namely, the probability distributions of the data of the second stage problem depend on first stage decisions. We proposed an adaptation of the L-Shaped method for solving two-stage stochastic programs to this class of problem. Critical to this adaptation is the concept of *distribution-specific* cuts. We presented a mechanism for adding such cuts to a master problem solved in the context of executing an L-Shaped method. We also demonstrated how such cuts can be derived for two important classes of sub-problem. The first is sub-problems that can be formulated as a linear program. The second is those that can be formulated as a mixed integer linear program. We also proposed optimality cuts that are distribution-independent. We adapted the general method to three variants of a production planning problem under endogenous and exogenous uncertainty. With an extensive computational study involving over 2,000 instances we demonstrated the superior computational performance of the proposed adaptation of the L-Shaped method.

We see multiple avenues for future work. First, we recall that the validity of the cuts presented for the case of second stage subproblems that are mixed integer linear programs requires an additional assumption regarding the domains of first stage decision variables. Thus, deriving cuts that do not require this assumption would leave a method that is applicable to a broad class of problems. The mechanism we provide for creating distribution-specific cuts can be easily adapted to cuts applicable to different types of first stage variables. Second, the method involves generating distribution-specific inequalities. While we also propose inequalities that are distribution-independent, one can envision enhancing the method with techniques that adapt the inequalities generated for one distribution to be valid for another. Third, there has been a tremendous amount of research recently on techniques for speeding up the L-Shaped method when solving stochastic programs in which uncertainty is exogenous. Adapting such techniques to the proposed method in this paper is a promising line of research. One example of such a technique is a multi-cut version of the proposed method in which cuts are generated for each distribution and/or each of its scenarios. Fourth, there are many practical applications in which uncertainty is endogenous. Thus, another line of future work is to adapt the proposed method to problems other than the production planning problem considered in this paper. Tailoring the method to specific problems may enable further enhancements as problem structure may be exploited. Fifth, preprocessing techniques for identifying when a partition of the feasible region can not contain an optimal solution will likely enable the method to converge much more quickly. The avenues just outlined are primarily methodological. We also believe there is value in extending the well-known concepts of *Value of the Stochastic Solution* and *Expected Value of Perfect Information* to problems with decision-dependent uncertainty.

## References

- [1] S. Ahmed. *Strategic planning under uncertainty: Stochastic integer programming approaches*. Phd, University of Illinois at Urbana-Champaign, 2000.

- [2] F. A. Al-Khayyal and J. E. Falk. Jointly constrained biconvex programming. *Mathematics of Operations Research*, 8(2):273–286, 1983.
- [3] R. M. Apap and I. E. Grossmann. Models and computational strategies for multistage stochastic programming under endogenous and exogenous uncertainties. *Computers & Chemical Engineering*, 103:233–274, 2016.
- [4] E. Balas. *Disjunctive programming*. Springer, 2018.
- [5] M. Colvin and C. T. Maravelias. A stochastic programming approach for clinical trial planning in new drug development. *Computers & Chemical Engineering*, 32(11):2626–2642, 2008.
- [6] M. Colvin and C. T. Maravelias. Modeling methods and a branch and cut algorithm for pharmaceutical clinical trial planning using stochastic programming. *European Journal of Operational Research*, 203(1):205–215, 2010.
- [7] B. da Costa Flach. *Stochastic Programming with Endogenous Uncertainty: An Application in Humanitarian Logistics*. PhD thesis, PUC-Rio, 2010.
- [8] L. F. Escudero, M. A. Garín, J. F. Monge, and A. Unzueta. On preparedness resource allocation planning for natural disaster relief under endogenous uncertainty with time-consistent risk-averse management. *Computers & Operations Research*, 98:84–102, 2018.
- [9] V. Goel and I. E. Grossmann. A Class of stochastic programs with decision dependent uncertainty. *Mathematical Programming*, 108:355–394, 2006.
- [10] V. Gupta and I. E. Grossmann. Solution strategies for multistage stochastic programming with endogenous uncertainties. *Computers & Chemical Engineering*, 35:2235–2247, 2011.
- [11] H. Held and D. L. Woodruff. Heuristics for Multi-Stage Interdiction of Stochastic Networks. *Journal of Heuristics*, 11(5-6):483–500, 2005.
- [12] L. Hellemo. *Managing Uncertainty in Design and Operation of Natural Gas Infrastructure*. Phd, Norwegian University of Science and Technology, 2016.
- [13] L. Hellemo, P. I. Barton, and A. Tomasgard. Decision-dependent probabilities in stochastic programs with recourse. *Computational Management Science*, 15(3):369–395, 2018.
- [14] M. Hewitt and G. Pantuso. Production planning under demand and endogenous supply uncertainty. *INFORMS Journal on Computing*, 0(0):0, 2025.
- [15] F. Hooshmand and S. A. MirHassani. Efficient constraint reduction in multistage stochastic programming problems with endogenous uncertainty. *Optimization Methods and Software*, 31:359–376, 2016.



- [16] F. Hooshmand and S. A. MirHassani. Reduction of nonanticipativity constraints in multistage stochastic programming problems with endogenous and exogenous uncertainty. *Mathematical Methods of Operations Research*, 87(1):1–18, 2018.
- [17] IBM. User’s Manual for CPLEX. Technical report, IBM Corporation, 2020.
- [18] T. W. Jonsbråten, R. J.-B. Wets, and D. L. Woodruff. A class of stochastic programs with decision dependent random elements. *Annals of Operations Research*, 82:83–106, 1998.
- [19] G. Laporte and F. V. Louveaux. The integer L-shaped method for stochastic integer programs with complete recourse. *Operations Research Letters*, 13(3):133–142, 1993.
- [20] M. Laumanns, S. Prestwich, and B. Kawas. Distribution shaping and scenario bundling for stochastic programs with endogenous uncertainty. In J. L. Higle, W. Römisch, and S. Sen, editors, *Stochastic Programming E-Print Series*, Stochastic Programming E-Print Series. Institut für Mathematik, 2014.
- [21] M. A. Lejeune, P. Dehghanian, and W. Ma. Profit-based unit commitment models with price-responsive decision-dependent uncertainty. *European Journal of Operational Research*, 314(3):1052–1064, 2024.
- [22] G. P. McCormick. Computability of global solutions to factorable nonconvex programs: Part I—Convex underestimating problems. *Mathematical programming*, 10(1):147–175, 1976.
- [23] L. Mercier and P. Van Hentenryck. An anytime multistep anticipatory algorithm for on-line stochastic combinatorial optimization. *Annals of Operations Research*, 184(1):233–271, 2011.
- [24] G. Pantuso. A node formulation for multistage stochastic programs with endogenous uncertainty. *Computational Management Science*, 18(3):325–354, 2021.
- [25] S. Peeta, F. S. Salman, D. Gunneer, and K. Viswanath. Pre-disaster investment decisions for strengthening a highway network. *Computers & Operations Research*, 37:1708–1719, 2010.
- [26] G. C. Pflug. On-line optimization of simulated markovian processes. *Mathematics of Operations Research*, 15(3):381–395, 1990.
- [27] B. Tarhan and I. E. Grossmann. A multistage stochastic programming approach with strategies for uncertainty reduction in the synthesis of process networks with uncertain yields. *Computers & Chemical Engineering*, 32:766–788, 2008.
- [28] B. Tarhan, I. E. Grossmann, and V. Goel. Stochastic Programming Approach for the Planning of Offshore Oil or Gas Field Infrastructure under Decision-Dependent Uncertainty. *Industrial & Engineering Chemistry Research*, 48:3078–3097, 2009.

- [29] B. Tarhan, I. E. Grossmann, and V. Goel. Computational strategies for non-convex multistage MINLP models with decision-dependent uncertainty and gradual uncertainty resolution. *Annals of Operations Research*, 203(1):141–166, 2013.
- [30] K. Tong, Y. Feng, and G. Rong. Planning under Demand and Yield Uncertainties in an Oil Supply Chain. *Industrial & Engineering Chemistry Research*, 51(2):814–834, 2012.
- [31] K. Viswanath, P. Srinivas, and S. F. Sibel. Investing in the Links of a Stochastic Network to Minimize Expected Shortest Path Length. 2004.
- [32] D. W. Walkup and R. J.-B. Wets. Stochastic programs with recourse. *SIAM Journal on Applied Mathematics*, 15(5):1299–1314, 1967.
- [33] C. A. Yano and H. L. Lee. Lot sizing with random yields: A review. *Operations research*, 43(2):311–334, 1995.

## 7 Acknowledgments

This research was partly supported by the *Novo Nordisk Fonden* grant NNF24OC0089770.

## Appendix A Example reformulation with binary variables

Let  $\mathcal{X} = \{0, 1\}^{n_1}$  with  $n_1 = 5$ . Let  $n^{(1)} = 2$ ,  $n^{(2)} = 4$ ,  $n^{(3)} = 5$ , so that each  $x$  vector is split into  $T = 3$  segments. Assume there are exactly two exhaustive conditions, say  $k_1$  and  $k_2$ , on segment  $t$  given by, respectively,  $\sum_{i=n^{(t-1)}+1}^{n^{(t)}} x_i < 1$  and  $\sum_{i=n^{(t-1)}+1}^{n^{(t)}} x_i \geq 1$ . Hence,  $\mathcal{K}_t = \{k_1, k_2\}$  for  $t = 1, \dots, T$ . We obtain  $|\mathcal{D}| = \prod_{t=1}^3 |\mathcal{K}_t| = 8$ . Let the function  $k(t, d)$  be specified as follows:  $k(1, d_1) = k_1$ ,  $k(2, d_1) = k_1$ ,  $k(3, d_1) = k_1$ ,  $k(1, d_2) = k_1$ ,  $k(2, d_2) = k_1$ ,  $k(3, d_2) = k_2$ ,  $k(1, d_3) = k_1$ ,  $k(2, d_3) = k_2$ ,  $k(3, d_3) = k_1$ ,  $k(1, d_4) = k_1$ ,  $k(2, d_4) = k_2$ ,  $k(3, d_4) = k_2$ ,  $k(1, d_5) = k_2$ ,  $k(2, d_5) = k_1$ ,  $k(3, d_5) = k_1$ ,  $k(1, d_6) = k_2$ ,  $k(2, d_6) = k_1$ ,  $k(3, d_6) = k_2$ ,  $k(1, d_7) = k_2$ ,  $k(2, d_7) = k_2$ ,  $k(3, d_7) = k_1$ ,  $k(1, d_8) = k_2$ ,  $k(2, d_8) = k_2$ ,  $k(3, d_8) = k_2$ . The sets  $\mathcal{X}_d$  are then specified as follows:  $\mathcal{X}_{d_1} = \{(0, 0, 0, 0, 0)\}$ ,  $\mathcal{X}_{d_2} = \{(0, 0, 0, 0, 1)\}$

$$\mathcal{X}_{d_3} = \{(0, 0, 0, 1, 0), (0, 0, 1, 0, 0), (0, 0, 1, 1, 0)\}$$

$$\mathcal{X}_{d_4} = \{(0, 0, 0, 1, 1), (0, 0, 1, 0, 1), (0, 0, 1, 1, 1)\}$$

$$\mathcal{X}_{d_5} = \{(1, 0, 0, 0, 0), (0, 1, 0, 0, 0), (1, 1, 0, 0, 0)\}$$

$$\mathcal{X}_{d_6} = \{(1, 0, 0, 0, 1), (0, 1, 0, 0, 1), (1, 1, 0, 0, 1)\}$$

$$\mathcal{X}_{d_7} = \{x \in \{0, 1\}^5 \mid \sum_{i=1}^2 x_i \geq 1, \sum_{i=3}^4 x_i \geq 1, x_5 = 0\}$$

$$\mathcal{X}_{d_8} = \{x \in \{0, 1\}^5 \mid \sum_{i=1}^2 x_i \geq 1, \sum_{i=3}^4 x_i \geq 1, x_5 = 1\}$$

The reformulation proposed in Section 3.3 entails creating six binary variables  $v_{tk}$  instead of eight binary variables  $\delta_d$ .

## Appendix B Extensive linearized form of problem variant 1

The extensive linearized form of (34) is as follows.

$$\max \quad - \sum_{f \in \mathcal{F}} C_f x_f + \sum_{d \in \mathcal{D}} \sum_{s \in \mathcal{S}_d} \pi_{sd} \left( P \mu_{sd} + O \rho_{sd} \right) \quad (39a)$$

$$\text{s.t.} \quad \sum_{l \in \mathcal{L}_f} y_{fl} = 1 \quad \forall f \in \mathcal{F} \quad (39b)$$

$$\sum_{l \in \mathcal{L}_f} L_{fl} y_{fl} \leq x_f \leq \sum_{l \in \mathcal{L}_f} U_{fl} y_{fl} \quad \forall f \in \mathcal{F} \quad (39c)$$

$$\sum_{d \in \mathcal{D}} \delta_d = 1 \quad , \quad (39d)$$

$$\sum_{f \in \mathcal{F}} y_{f,l(f,d)} \geq |\mathcal{F}| \delta_d \quad \forall d \in \mathcal{D} \quad (39e)$$

$$w_{sd} + o_{sd} = \sum_{f \in \mathcal{F}} Y_{fsd} x_f \quad \forall d \in \mathcal{D}, s \in \mathcal{S}_d, \quad (39f)$$

$$w_{sd} \leq D_{sd} \quad \forall d \in \mathcal{D}, s \in \mathcal{S}_d \quad (39g)$$

$$x_f \geq 0 \quad \forall f \in \mathcal{F}, \quad (39h)$$

$$y_{fl} \in \{0, 1\} \quad \forall f \in \mathcal{F}, l \in L_{pf}, \quad (39i)$$

$$\delta_d \in \{0, 1\} \quad \forall d \in \mathcal{D}, \quad (39j)$$

$$w_{sd} \geq 0, o_{sd} \geq 0 \quad \forall d \in \mathcal{D}, s \in \mathcal{S}_d. \quad (39k)$$

$$\mu_{sd} \leq w_{sd} \quad \forall d \in \mathcal{D}, s \in \mathcal{S}_d \quad (39l)$$

$$\mu_{sd} \leq D_{sd} \delta_d \quad \forall d \in \mathcal{D}, s \in \mathcal{S}_d \quad (39m)$$

$$\mu_{sd} \geq w_{sd} - D_{sd}(1 - \delta_d) \quad \forall d \in \mathcal{D}, s \in \mathcal{S}_d \quad (39n)$$

$$\rho_{sd} \leq o_{sd} \quad \forall d \in \mathcal{D}, s \in \mathcal{S}_d \quad (39o)$$

$$\rho_{sd} \leq N_{sd} \delta_d \quad \forall d \in \mathcal{D}, s \in \mathcal{S}_d \quad (39p)$$

$$\rho_{sd} \geq o_{sd} - N_{sd}(1 - \delta_d) \quad \forall d \in \mathcal{D}, s \in \mathcal{S}_d. \quad (39q)$$

Here, variables  $\mu_{sd}$  and  $\rho_{sd}$  linearize the products  $\delta_{sd} w_{sd}$  and  $\delta_{sd} o_{sd}$ , respectively. Constants  $M_{sd}$  and  $N_{sd}$  represent upper bounds on sales and oversupply, respectively.