On the Computational Solution of Some Distributionally Ambiguous Two-Stage Stochastic Problems

Pedro Borges* Claudia Sagastizábal† Mikhail Solodov‡ Asgeir Tomasgard §

Abstract

Optimization under ambiguity sets provides a mechanism to deal with uncertainty about the distributions of the random variables involved in the formulation of stochastic programs. We consider a new model, that can be senn as a mixture of the optimistic and pessimistic paradigms in distributionally robust stochastic optimization. Tractability is shown for two-stage stochastic linear programs using Wasserstein balls as ambiguity sets. The resulting reformulation has a bilinear objective function and nonsmooth convex feasible set, with a number of constraints depending on the number of scenarios. We propose a decomposition method along scenarios that converges to a solution, provided a global optimization solver for bilinear programs with polyhedral feasible sets is available. The solution procedure is applied to a case study on expansion of energy generation that takes into account the progressive decommissioning of fossil-fueled technologies under uncertain future market conditions.

Keywords distributionally robust optimization, nonconvex nonsmooth optimization, decomposition methods.

1 Introduction

Decision-making in an uncertain environment is the realm of Stochastic Programming, an area of Mathematical Optimization particularly suitable for real-life problems, that often include unknown parameters. Stochastic programs assume that randomness realizes with a known probability, but in many decision problems the probability distribution of uncertain parameters is only indirectly observable, for example through samples. If the probability values used in the model are incorrect, the stochastic program can give suboptimal results. To address this issue, the more recent area of Distributionally Robust Stochastic Optimization (DRSO) considers the probability as an additional decision variable, to be chosen among many distributions in a certain *ambiguity set*, see for instance the works by G. Pflüg and Wozabal (2007) and Wozabal (2010).

A second major concern for the decision-maker is the type of randomness. If the parameter realizations are influenced by the decisions, uncertainty is said to be decision-dependent or of endogenous type. An exogenous probability that has an endogenous type arises when scheduling clinical trials or vaccination campaigns, in connection with the technical uncertainty in the outcome, clearly exogenous. However, as explained by Colvin and Maravelias (2009), the timing at which this uncertainty is observed is endogenous because it depends on the decision on when to perform the trial or on how to deploy the vaccination campaign Stochastic Programming models with decision-dependent probabilities were considered by Jonsbråten, Wets, and Woodruff (1998), Goel and Grossmann (2004) and, more recently, Hellemo, Barton, and Tomasgard (2018). Robust formulations of problems with decision-dependent uncertainty are shown to be NP-complete by Nohadani and Sharma (2018). Both Goh and Sim (2010) and Postek, Hertog, and Melenberg (2014) confirm that the well-known conservatism of robust optimization approaches is mitigated thanks to the consideration of ambiguity sets.

*Instituto de Matemática Pura e Aplicada, Rio de Janeiro, RJ, Brazil (pborges@impa.br)
†IMECC/UNICAMP, Campinas, São Paulo, Brazil (sagastiz@unicamp.br).
‡Instituto de Matemática Pura e Aplicada, Rio de Janeiro, RJ, Brazil (solodov@impa.br).
§Norwegian University of Science and Technology, Trondheim, Trondelag, Norway (asgeir.tomasgard@ntnu.no).

With a DRSO model, endogenous probability distributions imply a variation of the ambiguity set with the decision variable. This is the model explored in Luo and Mehrotra (2020), for different types of ambiguity sets. It is also the topic of this work, that we focus on decision-dependent distributionally robust two-stage stochastic optimization, ddDR2SO or ddDRSO for short. A dd2DRSO problem outputs the optimal value for the probability distribution together with the optimal decisions. The connection between the endogenous probability and the decision variable can be explicitly given through constraints in the ambiguity set, or indirectly, by the fact that the optimization is performed jointly on both type of variables; see (4) below.

Several different specifications for ambiguity sets have been considered in the literature. One possibility is to look for probability distributions whose first- and second-order moments are close to those of some exogenous empirical estimation. To cite a couple of works among many others, this is the approach in Bertsimas, Doan, Natarajan, and Teo (2010) and Delage and Ye (2010), where the wording "distributionally robust" was first coined. Another line of research considers probability distributions that are close in some metric to a nominal probability, taken as a reference. Typical measures of closeness of measures are the phi-divergence distance and the Wasserstein balls, respectively analyzed by Ben-Tal and Teboulle (1987) and G. Pflüg and Wozabal (2007). Phi-divergences, introduced in Calafiore (2007) for ambiguity sets, are employed in Wang, Glynn, and Ye (2015) to include distributions that make observed historical data achieve a certain level of likelihood. Bayraksan and D. K. Love (2015) extended the concept to two-stage stochastic linear programs. An indication of the interest that the Mathematical Optimization community is currently paying to the subject is given by the recent tutorial by Kuhn, Esfahani, Nguyen, and Shafieezadeh-Abadeh (2019), oriented to problems in machine learning, and the review by Leyffer, Menickelly, Munson, Vanaret, and Wild (2020), focusing on solution approaches involving mixed-integer nonlinear programming techniques.

In the spectrum of possible choices for the probability distribution that enters the optimization problem, the modeling paradigms of Stochastic Programming and DRSO are positioned at opposite extremes. The latter takes the worst case over the ambiguity set while with the former the choice shrinks to a singleton. From the point of view of decision-making, this amounts to adopting either a fully optimistic or a completely pessimistic view of how well the probability distribution fits the random nature of the uncertain parameters.

We propose an in-between paradigm, that is both optimistic and pessimistic to certain degrees. For ambiguous two-stage stochastic programs, our ddDRSO model defines a robustified expected recourse function using probabilities in a Wasserstein ball. The novelty is that, instead of taking a nominal probability, the ball center is considered variable. This additional variable is minimized in the first stage over a simple convex set, for example the convex hull of several nominal probabilities taken as a reference.

When considering discrete distributions, with a finite number of scenarios, the new model is tractable and can be reformulated as a bilinear programming problem; see Lemma 3.1 below. The structure is suitable for decomposition: bilinearity appears only in the objective function, and the feasible set has as many convex nonsmooth constraints as scenarios in the problem. Indeed, as stated in (9), the convex nonsmooth recourse functions of the original two-stage problem appear in the constraint set of our reformulation. Similarly to an L-shaped method Van Slyke and Wets (1969), the solution of scenario subproblems provides cuts that either linearize the recourse function or cut-off infeasible first-stage points. The algorithm then proceeds by incorporating feasibility and optimality cuts into a master program that still has a bilinear objective function, but now a polyhedral feasible set. If the solution method employed to solve the master programs can find global solutions, the procedure converges to a solution. In our case study, this is ensured by using the solver BARON, by Tawarmalani and Sahinidis (2005).

The interest of the decomposition in a ddDR2SO framework is that, instead of solving one large, difficult, problem with many bilinear terms (and a nonsmooth feasible set), each iteration solve a much easier problem, having less bilinear terms, less variables and a polyhedral feasible set. These features are crucial for computational efficiency. The methodology is illustrated on a case study for planning investments in energy generation, under uncertain market conditions. The simplified model considers the whole of Europe over the horizon 2020-2050, taking into account the progressive decommissioning of thermal power plants and the increasing proportions of renewable technologies that are foreseable for the power mix. Ambiguity sets are particularly suitable for this setting, because the long-term horizon complicates vastly the assignement of a unique probability to an uncertain event. Our approach gives an indication if the investment return is stable relative to parameters of the problem.

The work is organized as follows. The new paradigm, that deals with ambiguity from a perspective placed between the optimistic and the pessimistic models, is given in Section 2. Tractability for Wasserstein balls, shown in Section 3, is exploited in Section 4 to derive the decomposition method for ddrSO two-stage linear programs.

A thorough numerical assessment, validating the approach for the energy investment problem, is performed in Section 5. The paper ends with some concluding comments and remarks.

2 Partly optimistic and partly pessimistic settings

Consider a vector of decision variables $x \in \mathbb{R}^n$ with feasible set $\mathcal{X} \subset \mathbb{R}^n$. Uncertain parameters are modeled by a measurable mapping $\xi : \Omega \to \Xi \subset \mathbb{R}^m$ and, given a σ-algebra $\mathcal{F}$, the sample space is $(\Omega, \mathcal{F})$. Given a function $C_x : \mathbb{R}^n \to \mathbb{R}$ and a cost-to-go mapping $\mathfrak{Q} : \mathbb{R}^n \times \Xi \to \mathbb{R} \cup \{\pm\infty\}$, the two-stage problem attached to the probability distribution p is

$$\min_{x \in \mathcal{X}} \left\{ C_x(x) + \mathbb{E}_p \left[\mathfrak{Q}(x, \xi(\omega)) \right] \right\}. \tag{1}$$

In the objective function we make explicit the deterministic cost $C_x(x)$ to exploit two-stage structural properties favorable for decomposition methods, like those in Sagastizábal (2012), see also D. Love and Bayraksan (2013).

In order to formalize our approach, a nonempty set of probability distributions $\mathcal{P}$ is defined on the set $\mathcal{M}$ of measures on the sample space. If selecting the distribution $p \in \mathcal{P}$, among all the possible probability distributions, entails a cost $C_p : \mathcal{M} \to \mathbb{R}$, our first model of distributionally ambiguous problem has the form

$$\min_{x \in \mathcal{X}} \left\{ C_x(x) + \sup_{p \in \mathcal{P}} \Big\{ C_p(p) + \mathbb{E}_p \left[\mathfrak{Q}(x, \xi(\omega)) \right] \Big\} \right\}. \tag{2}$$

The format (2) also encompasses the robust optimization pradigm by Ben-Tal, Ghaoui, and Nemirovski (2009), taking as ambiguity set the full space of measures. Continuing with connections with other models, notice that the DRSO problem (2) resolves the ambiguity present in the probability distribution by adopting a pessimistic view, as in Wiesemann, Tsoukalas, Kleniati, and Rustem (2013) and Kuhn, Esfahani, Nguyen, and Shafieezadeh-Abadeh (2019, Section 2.1). Specifically, introducing the worst-case risk functional,

$$\Theta_{\text{pess}}(x) := \sup_{p \in \mathcal{P}} \Big\{ C_p(p) + \mathbb{E}_p \left[\mathfrak{Q}(x, \xi(\omega)) \right] \Big\},$$

gives the following problem, equivalent to (2):

$$\min_{x \in \mathcal{X}} \Big\{ C_x(x) + \Theta_{\text{pess}}(x) \Big\}.$$

Another option is to choose the most favorable output, plugging in the optimization problem the risk functional

$$\Theta_{\text{opt}}(x) := \inf_{p \in \mathcal{P}} \Big\{ C_p(p) + \mathbb{E}_p \left[\mathfrak{Q}(x, \xi(\omega)) \right] \Big\},$$

that Dempe, Dutta, and Mordukhovich (2007) called optimistic solution in bilevel programming.

Our proposal, set amid these two alternatives, is to consider problems of the form

$$\min_{x \in \mathcal{X}} \Big\{ C_x(x) + \Theta_\kappa(x) \Big\} \quad \text{for } \Theta_\kappa(x) := \inf_{p \in \mathcal{P}} \Big\{ C_p(p) + \sup_{q \in \mathbb{B}_{\kappa(p)}} \mathbb{E}_q \left[\mathfrak{Q}(x, \xi(\omega)) \right] \Big\}, \tag{3}$$

where $\mathbb{B}_{\kappa(p)} \subset \mathcal{M}$ is a ball about p of radius κ. Tractability of the new approach depends on the choice of these balls, an issue considered in details in Section 3. For now we just mention that with the Wasserstein metric the balls are defined by a system of affine inequalities, a convenient feature when it comes to computational implementation. In particular, proceeding in a manner similar to the presentation by Noyan, Rudolf, and Lejeune (2018), we use duality arguments and the Wasserstein balls, in Lemma 3.1 to rewrite the supremum defining the risk functional as a minimum, therefore avoiding technical issues that arise with the pessimistic formulation.

Since in (3), letting $\kappa = 0$ and $\mathbb{B}_{\kappa(p)} = \mathcal{P}$ respectively gives the optimistic and pessimistic approaches, problem (3) can be thought of being cast in a *robustified optimistic* setting. The parameter κ, called the robustification ratio, determines to which extent the risk functional in (3) bends towards an optimistic or a pessimistic view. The resulting problem takes into account some ambiguity, while hedging against estimation errors on the probability values. Additionally, because the optimization process outputs the "nominal" probability p together with the decision variable x, our model (3) is also suitable for problems with decision-dependent probabilities.

When the sampling space is discrete, say given by scenarios in the set $\lfloor S \rfloor := \{1, \ldots, S\}$, only discrete measures are considered and $\mathcal{P} \subset \mathbb{R}^S$ is assumed to be a bounded set. In this context, problem (3) writes down as follows:

$$\min_{x \in \mathcal{X}} \left\{ C_x(x) + \left[\min_{p \in \mathcal{P} \subset \mathbb{R}^S} \left(C_p(p) + \max_{q \in \mathbb{B}_\kappa(p) \subset \mathbb{R}^S} \sum_{s=1}^{S} q_s \mathfrak{Q}_s(x) \right) \right] \right\} , \tag{4}$$

where we use the shorter notation $\mathfrak{Q}_s(\cdot) := \mathfrak{Q}(\cdot, \xi_s(\omega))$ for the recourse functions.

We now show that when the set $\mathbb{B}_{\kappa(p)}$ is a Wasserstein-type ball of nearby probabilities, a dualization formula inspired from Noyan, Rudolf, and Lejeune (2018) gives an explicit expression that makes our ddDRSO problem tractable. To this end, it is convenient to consider separately the q-decision variable in the maximum, and write the problem in the following form:

$$\min_{x \in \mathcal{X}, p \in \mathcal{P}} \Big\{ C_x(x) + C_p(p) + \mathbb{E}_{\mathbb{B}_\kappa(p)} \left[\mathfrak{Q}(x) \right] \Big\}, \quad \text{where} \quad \mathbb{E}_{\mathbb{B}_\kappa(p)} \left[\mathfrak{Q}(x) \right] := \max_{q \in \mathbb{B}_\kappa(p)} \sum_{s=1}^{S} q_s \mathfrak{Q}_s(x) \,. \tag{5}$$

3 Wasserstein Balls and Likelihood Robustification

Robustification operations are based on defining vicinities for the quantities being robustified, and then taking the supremum over that vicinity. The underlying quantity can be an expectation or a variance. The idea can be applied in general with varying degrees of practical computational use, the procedure depends on the specific form of the balls considered in (5). Each specific form of vicinity gives rise to a different robustification strategy.

We focus on Wasserstein-distance based sets because their polyhedral structure makes them particularly suitable for algorithmic developments. The basic workhorse to arrive to tractable reformulations is Fenchel duality. Tractability of ambiguous problems with Wasserstein balls is discussed in Postek, Hertog, and Melenberg (2014) for discrete measures and in Hanasusanto and Kuhn (2018) for continuous distributions. The latter work, in particular, derives a copositive reformulation for two-stage robust and DRSO linear programs from ℓ_2-Wasserstein balls. The same authors show that when the ambiguity set is centered at a discrete distribution and there are no support constraints, the ℓ_1-Wasserstein ball gives linear programming problems. More recently, complexity bounds for two-stage DRSO problems with the ℓ_∞-norm were given in Xie (2020).

Back to our ddDRSO problem (5), a first step is to specify the concept of closeness of the discrete probability $(q_1, \ldots, q_S)$ to the ball center, $(p_1, \ldots, p_S)$. As explained in Noyan, Rudolf, and Lejeune (2018), this could be done by considering the distance between random vectors $\xi(\omega_r)$ and $\xi(\omega_s)$, or directly considering the values p_r and p_s. The first approach, adopted by G. C. Pflüg, Pichler, and Wozabal (2012) and Esfahani and Kuhn (2017) to deal with continuous measures, is what Noyan, Rudolf, and Lejeune (2018) called a continuous robustification. The specific stage structure in our risk functional in (5) is more suitable for the second option, that directly compares the vectors p and q through the recourse functions. We see this second approach as being a *robustification of the likelihood* (the model is called discrete robustification in Noyan, Rudolf, and Lejeune (2018)).

For discrete measures in the set

$$\mathcal{M}^S := \left\{ p \in \mathbb{R}^S : p \geq 0 \,, \sum_{r=1}^{S} p_r = 1 \right\} ,$$

the notion of Wasserstein distance depends on certain weighed costs $\delta_{rs}^{\mathbb{D}} > 0$ for $r, s \in \lfloor S \rfloor$, related to transporting the probability mass p_r to q_s, using a general distance function $\mathbb{D}$. For our two-stage setting, the cost $\delta^{\mathbb{D}}$ measures similarity between recourse functions using the ℓ_1-norm. Examples on how to compute such values from problem data with $\mathbb{D}(\cdot) = \| \cdot \|_1$ are given below, after problem (9).

The Wasserstein distance between p and $q \in \mathcal{M}^S$, defined below, is denoted by $\Delta(p, q)$, where we drop dependence on the distance $\mathbb{D}$ to alleviate notation. Its value is given by the optimal transport plan

$$\Delta(p,q) := \min_z \left\{ \sum_{s,r=1}^{S} \delta^{\mathbb{D}}_{rs} z_{rs} \quad : \quad \sum_{s=1}^{S} z_{rs} = p_r, \quad \sum_{r=1}^{S} z_{rs} = q_s, \quad z_{rs} \geq 0\,, \text{for } r\,, s \in \lfloor S \rfloor \right\}, \tag{6}$$

where z_{rs} represents the amount of mass of p_r that is moved to q_s. The function $\Delta(p,q)$ is convex in both $p\,, q \in \mathcal{M}^S$. The notion is a distance (and not a mere semi-distance) if the transportation cost has the properties $r \neq s \Rightarrow \delta^{\mathbb{D}}_{rs} > 0$ and $\delta^{\mathbb{D}}_{ss} = 0$ (which implies that $\Delta(p,q) = 0$ if and only if $p = q$). When the transportation cost is defined using the ℓ_1-norm, the Kantorovich-Rubinstein theorem stated in G. C. Pflüg, Pichler, and Wozabal (2012, Section 1.3) gives an equivalent dual expression for (6).

For the reformulation of our ddDRSO problem (5), the following technical result, similar to a statement in Noyan, Rudolf, and Lejeune (2018, Section 5), is useful. The direct proof given below is based on Lagrangian relaxation.

Lemma 3.1 (Support function of Wasserstein balls)**.** *Given the distance defined in* (6)*, consider the associated ℓ_1-Wasserstein ball*

$$\mathbb{B}_{\kappa(p)} := \left\{ q \in \mathcal{M}^S : \Delta(p,q) \leq \kappa \right\} .$$

Then, its support function, defined as $\sigma_{\mathbb{B}_\kappa(p)}(d) := \max\left\{ d^\top q : q \in \mathbb{B}_{\kappa(p)} \right\}$ for any $d \in \mathbb{R}^S$, has the equivalent expression

$$\sigma_{\mathbb{B}_\kappa(p)}(d) = \begin{cases} \min & \tau\kappa + \displaystyle\sum_{s=1}^{S} p_s v_s \\ \text{s.t.} & \tau \in \mathbb{R}\,, \tau \geq 0, \\ & v \in \mathbb{R}^S, \\ & v_s \geq d_r - \tau \delta^{\mathbb{D}}_{sr} \quad \textit{for } r\,, s \in \lfloor S \rfloor\,. \end{cases}$$

Proof. Introducing the S-dimensional vector $\mathbf{1}$ with all components equal to 1 for $s \in \lfloor S \rfloor$, write the maximum as a minimization problem:

$$-\sigma_{\mathbb{B}_\kappa(p)}(d) = \min\left\{ -d^\top q : \mathbf{1}^\top q = 1\,, q \geq 0\,, q \in \mathbb{B}_{\kappa(p)} \right\} .$$

Also, for notational convenience, we write the relations in (6) considering $\delta^{\mathbb{D}}$ and z vectors in $\mathbb{R}^{S^2}$, and introduce $S \times S^2$ matrices M_p and M_q so that

$$\Delta(p,q) = \min\left\{ z^\top \delta^{\mathbb{D}} : M_p z = p\,, M_q z = q\,, z \geq 0 \right\} .$$

Then, in the resulting problem

$$-\sigma_{\mathbb{B}_\kappa(p)}(d) = \begin{cases} \min & -d^\top q \\ \text{s.t.} & \mathbf{1}^\top q = 1, \\ & q \geq 0, \\ & z \geq 0, \\ & z^\top \delta^{\mathbb{D}} \leq \kappa, \\ & M_p z = p, \\ & M_q z = q\,. \end{cases} \quad \text{eliminate the variable } q: \quad \begin{cases} \min & -d^\top M_q z \\ \text{s.t.} & \mathbf{1}^\top M_q z = 1, \\ & M_q z \geq 0, \\ & z \geq 0, \\ & z^\top \delta^{\mathbb{D}} \leq \kappa, \\ & M_p z = p\,. \end{cases}$$

This is a linear program, and there is no duality gap. Next, introduce Lagrange multipliers $\eta \in \mathbb{R}$ for the first constraint, $0 \leq \mu \in \mathbb{R}^S$ for the second constraint, $\tau \geq 0$ for the constraint $z^\top \delta^{\mathbb{D}} \leq \kappa$, and $\lambda \in \mathbb{R}^S$ for the last constraint. Since the corresponding Lagrangian

$$L(z, \eta, \mu, \tau, \lambda) = z^\top \left(M_q^\top(-d + \mathbf{1}\eta - \mu) + \tau\delta^{\mathbb{D}} + M_p^\top \lambda \right) - \eta - \tau\kappa - p^\top \lambda$$

is separable, minimizing each component z_{rs} over the set $z_{rs} \geq 0$ gives a solution $z^*_{rs} = 0$ as long as

$$M_q^\top(-d + \mathbf{1}\eta - \mu) + \tau\delta^{\mathbb{D}} + M_p^\top \lambda \geq 0 \iff -d_r + \eta - \mu_r + \tau\delta^{\mathbb{D}}_{rs} + \lambda_s \geq 0\,.$$

This yields the dual formulation

$$-\sigma_{\mathbb{B}_\kappa(p)}(d) = \begin{cases} \max & -\eta - \tau\kappa - p^\top \lambda \\ \text{s.t.} & \eta\,, \tau \in \mathbb{R}\,, \mu\,, \lambda \in \mathbb{R}^S, \\ & \tau \geq 0\,, \mu \geq 0, \\ & \eta + \lambda_s \geq \mu_r + d_r - \tau\delta^{\mathbb{D}}_{sr} \text{ for } r\,, s \in \lfloor S \rfloor\,. \end{cases}$$

Taking $v := \eta\mathbf{1} + \lambda$, discarding the redundant variable μ, and reformulating the problem in minimization form concludes the proof. □

When applied to (5) written with ambiguity set equal to the the ℓ_1-Wasserstein ball, this result implies that

$$\mathbb{E}_{\mathbb{B}_\kappa(p)}\left[\mathfrak{Q}(x)\right] = \max_{q\geq 0}\left\{\sum_{s=1}^S q_s \mathfrak{Q}_s(x) : \sum_{s=1}^S q_s = 1\,, \Delta(p,q) \leq \kappa \text{ for } \Delta(p,q) \text{ from (6)}\right\}.$$

In terms of support functions, this boils down to the identity

$$\mathbb{E}_{\mathbb{B}_\kappa(p)}\left[\mathfrak{Q}(x)\right] = \sigma_{\mathbb{B}_\kappa(p)}(\mathfrak{Q}(x))\,.$$

Tractability of (5) then results from a direct application of Lemma 3.1, since

$$\mathbb{E}_{\mathbb{B}_\kappa(p)}\left[\mathfrak{Q}(x)\right] = \begin{cases} \min & \tau\kappa + \sum_{r=1}^S p_r v_r \\ \text{s.t.} & \tau \in \mathbb{R}\,, \tau \geq 0, \\ & v \in \mathbb{R}^S, \\ & v_r \geq \mathfrak{Q}_s(x) - \tau\delta^{\mathbb{D}}_{rs} \quad \text{for } r\,, s \in \lfloor S \rfloor\,. \end{cases} \tag{7}$$

We note, in passing, that it would be possible to include a cost for selecting a given probability in the robustification (the only change would be to add the vector C_p to the recourse function $\mathfrak{Q}_s(x)$ in the constraints).

We now give an interpretation of the object (7) regarding the original two-stage stochastic problem that justifies our naming, of likelihood robustification. Specifically, in (1), for each fixed $x \in \mathcal{X}$, and on the sampling space Ω^S, the random variable $\mathfrak{Q}(\xi(\omega), x)$ has realizations $\mathfrak{Q}_s(x)$. The usual expected recourse function therein is

$$\mathbb{E}_p\left[\mathfrak{Q}(x)\right] = \sum_{s=1}^S p_s \mathfrak{Q}_s(x)\,,$$

the best expected cost that can be computed when the probability distribution is known exactly. In comparison, the value in (7) gives the best possible value for the expectation when the worst possible error on p is done, considering all the distributions that are at distance $\kappa > 0$ of the probability p, taken as reference. Having this interpretation in mind, the value in (7) represents a likelihood robustification of such expectation, considering a ball with radius $\kappa \geq 0$.

Noyan, Rudolf, and Lejeune (2018) analyze numerous reformulations of DRSO problems. They also show an asymptotic result explaining why the continuous robustification may not be suitable for discrete sampling spaces. More precisely, with the discrete robustification (our setting), when the radius κ grows, Section 5 in that work shows that the likelihood robustified expectation (7) approaches the value of the worst case scenario. This property, confirmed numerically in our case study, follows from observing that, for κ sufficiently large, any probability is feasible for the max-operation in (4). This means, in particular, that the total probability of the worst scenario can be taken. The continuous robustification, by contrast, can far exceed the worst case scenario; when the radius κ grows the considered ball would include any distribution for the cost realizations.

Lemma 3.1, inspired by a discussion in Noyan, Rudolf, and Lejeune (2018), is the key to develop our new decomposition method, presented below.

4 Decomposition Method for Two-Stage Stochastic Linear Programs

When problem (1) is a two-stage stochastic linear program, its nonsmooth convex recourse functions

$$\mathfrak{Q}_s(x) = \inf\{d_s^\top y : W_s y = h_s - T_s x, \quad y \geq 0\}, \tag{8}$$

can take the values $\pm\infty$ (givne vectors d_s , h_s and matrices W_s , T_s, of appropriate dimensions).

Suppose, in addition, the first-stage feasible set is the polyhedron

$$\mathcal{X} := \{x \geq 0 : Ax = b\},$$

and the costs for both x and p are linear. In this setting, given a likelihood robustification ratio $\kappa \geq 0$ and using (7), problem (5) can be expressed as follows:

$$\begin{cases} \min\limits_{x,p,\tau,v} & C_x^\top x + C_p^\top p + \sum\limits_{r=1}^{S} p_r v_r + \kappa\tau \\ \text{s.t.} & x \geq 0, \quad Ax = b, \quad p \in \mathcal{P}, \\ & v_r \geq \mathfrak{Q}_s(x) - \tau\delta_{rs}^{\mathbb{D}} \quad s, r \in \lfloor S \rfloor, \\ & \tau \geq 0. \end{cases} \tag{9}$$

When $\mathcal{P} = \{p\}$, this is a well-known convex two-stage stochastic programming problem, for which useful solution procedures based on scenario decomposition are available, including the ones in Fábián and Szőke (2006) and Oliveira, Sagastizábal, and Scheimberg (2011).

When the set $\mathcal{P}$ is not a singleton and p becomes a variable, the bilinear terms $p_r v_r$ in the objective function lead to clear computational difficulties. The format of $\mathcal{P}$ is left unspecified for the moment. The only requirement is that it should be a simple convex set, for instance the convex hull of certain probability vectors taken as reference, so that the feasible set in (9) remains convex.

Before proceeding further, as announced before (6), we discuss possible choices for the transportation costs. A first option is to take

$$\delta_{rs}^{\mathbb{D}} = 1 \text{ if } r = s \text{ and } \delta_{rs}^{\mathbb{D}} = 0 \text{ otherwise.} \tag{10}$$

The corresponding measure $\Delta(p, q)$ is the the total variation distance between the probabilities p and q. This is the distance used in our case study. For other options, denote all the uncertain data of the problem by

$$\xi_s := (d_s, W_s, h_s, T_s),$$

and notice that this random variable determes the value for $\mathfrak{Q}_s(x)$, via the solution of the s-th second-stage. Then

$$\delta_{rs}^{\mathbb{D}} = \|\xi_s - \xi_r\|_p^p \text{ for } p \geq 1, \text{ and } \delta_{rs}^{\mathbb{D}}(x) = |\mathfrak{Q}_s(x) - \mathfrak{Q}_r(x)|.$$

Notwithstanding, as in (7) the transportation cost appears in the constraints, the latter choice is not suitable in practice, because it would yield a ddDRSO problem with nonlinearities that are hard to deal with computationally.

Our solution procedure replaces the convex functions $\mathfrak{Q}_s(x)$ by lower-bounding cutting-plane models, and uses global bilinear optimization software to solve a sequence of approximating master problems. The optimal values of these approximating problems estimate from below the true optimal value of the nonsmooth and nonconvex master problem (9).

For efficiency, it is crucial to compute optimality and feasibility cuts for the recourse functions $\mathfrak{Q}_s(x)$, as introduced by Van Slyke and Wets (1969). In what follows, we explain how to compute these cuts. For a fixed first-stage decision $x_k \in \mathcal{X}$ we can solve the problem defining $\mathfrak{Q}_s(x_k)$, concluding either that $\mathfrak{Q}_s(x_k) \in \{\pm\infty\}$ or obtaining a solution together with an associated Lagrange multiplier for the equality constraints. Each case is explained below.

Algorithm 1: CUT GENERATION FOR s-TH PROBLEM (8)

Input: For $s \in \lfloor S \rfloor$, and given x_k, solve (8).

Output: Either an optimality cut (Ocut), or a feasibility cut (Fcut), or an indefinite situation (Indef):

(Ocut) If $\mathfrak{Q}_s(x_k)$ is finite, it is well known that

$$\mathfrak{Q}_s(x) \geq \mathfrak{Q}_s(x_k) - \lambda_{sk} T_s^\top (x - x_k) \quad \forall x \in \mathbb{R}^n. \tag{11}$$

(Fcut) If $\mathfrak{Q}_s(x_k) = \infty$ we have to remove the point $x_k \in \mathcal{X}$ from the feasible set of (9) as follows. First define

$$U_s(x) = \inf\{\|s^+ - s^-\|_1 : W_s y = h_s - T_s x + s^+ - s^-, \quad y, s^+, s^- \geq 0\}.$$

Then, compute the respective primal solutions y_{sk}, s^+_{sk}, s^-_{sk} and an associated Lagrange multiplier λ_{sk} for the equality constraints. Now note that the inequality

$$0 \geq U_s(x_k) - \lambda_{sk} T_s^\top (x - x_k)$$

is not satisfied by $x_k \in \mathcal{X}$ because $U_s(x_k) > 0$, and it is satisfied for all points $x \in \mathbb{R}^n$ such that $\mathfrak{Q}_s(x) < +\infty$ by the gradient inequality applied to U_s and because we have $U_s(x) = 0$.

(Indef) If $\mathfrak{Q}_s(x_k) = -\infty$ we have to analyze the other scenarios to understand whether the master problem is unbounded from below, or whether $\mathfrak{Q}_r(x_k) = \infty$ for some $r \neq s$, in which case we would add a feasibility cut for scenario r.

Having computed optimality and feasibility cuts for the functions $\mathfrak{Q}_s(x)$, we are in position to define our sequence of lower-bounding approximating master problems to (9).

We assume given disjoint sets of optimality indices O_{sk} and feasibility indices F_{sk} such that

$$O_{sk} \cup F_{sk} = \{1, \ldots, k\},$$

as well as the history of first-stage iterates and associated Lagrange multipliers of the second-stage problems,

$$\{x_1, \ldots, x_k\} \text{ and } \{\lambda_{si} = \lambda_s(x_i) : s \in \lfloor S \rfloor, \quad i \in \lfloor k \rfloor\}.$$

The sets O_{sk} and F_{sk} are used to organize the information for the generated cuts, noting that knowing if λ_{sk} refers to a feasibility or optimality cut depends on the value of $\mathfrak{Q}_s(x_k)$. The k-th approximate master problem is:

$$\begin{cases} \min\limits_{x,p,\tau,v} & c^\top x + \sum\limits_{r=1}^{S} p_r v_r + \kappa\tau \\ \text{s.t.} & x \geq 0, \quad Ax = b, \quad p \in \mathcal{P}, \quad \tau \geq 0, \\ & v_r \geq \mathfrak{Q}_s(x_i) - \lambda_{si} T_s^\top (x - x_i) + d_s - \tau \delta^{\mathbb{D}}_{rs} \quad \forall s, r \in \lfloor S \rfloor \quad i \in O_{sk}, \\ & 0 \geq U_s(x_i) - \lambda_{si} T_s^\top (x - x_i) \quad \forall s \in \lfloor S \rfloor \quad \forall i \in F_{sk}, \\ & \|x\|_\infty, \|p\|_\infty, \|\tau\|_\infty, \|v\|_\infty \leq M. \end{cases} \tag{12}$$

As stated, the nonconvexity of the model comes from the terms $p_r v_r$. Additional box constraints are commonly used in cutting-plane methods with a sufficiently large constant $M > 0$. With those constraints the approximate master problem (12) always has a solution (which is not guaranteed using only optimality and feasibility cuts, especially at the initial iterations when the cuts for a given k represent poorly the true recourse functions $\mathfrak{Q}_s(x)$).

The solution procedure itself consists in defining a stopping test and rules to manage the sets O_{sk} and F_{sk}, and to compute the sequence of iterates converging to a solution of (9). Thanks to the well-known lower-bounding property of (12), which amounts to satisfaction of the subgradient inequality (11) for convex functions, we shall be able to measure an estimate for the gap, as long as the solver computes a global solution of (12) using bilinear global optimization. Putting all this information together, using standard arguments, it is easy to show that the following algorithm approximates arbitrarily well a global solution to our ddDRSO problem (9).

Algorithm 2: Solution procedure for problem (9).

Initialization: k $\longleftarrow 0$, $\mathbb{UB} \longleftarrow +\infty$, $\mathbb{LB} \longleftarrow -\infty$; $O_{s0} = F_{s0} \longleftarrow \emptyset$ for all $s \in \lfloor S \rceil$;
while $\mathbb{UB}$- $\mathbb{LB}>$ *tolerance* **do**
  compute a global solution $z_{k+1} = (x_{k+1}, p_{k+1}, \tau_{k+1}, v_{k+1})$ of (12) and store the optimal value at $\mathbb{LB}$;
  for $s \in \lfloor S \rceil$ **do**
    try to compute $\mathfrak{Q}_s(x_{k+1})$ and $\lambda_{s,k+1}$ as in Algorithm 1(**Ocut**);
    if $\mathfrak{Q}_s(x_{k+1}) = \infty$ **then**
      compute $U_s(x_{k+1})$ and $\lambda_{s,k+1}$ as in Algorithm 1(**Fcut**);
      $F_{s,k+1} \longleftarrow F_{sk} \cup \{k+1\}$;
    else if $\mathfrak{Q}_s(x_{k+1}) > -\infty$ **then**
      $O_{s,k+1} \longleftarrow O_{sk} \cup \{k+1\}$;
  end
  if $\exists r$ *s.t.* $\mathfrak{Q}_r(x_{k+1}) = -\infty$ *and* $\mathfrak{Q}_s(x_{k+1}) < \infty$ *for* $r \neq s$ **then**
    declare the problem unbounded from below and stop the loop;
  else if $\mathfrak{Q}_s(x_k) < \infty \quad \forall s \in \lfloor S \rceil$ **then**
    update $\mathbb{UB}$ as the optimal value of (9) with $x = x_k$;
  $k \longleftarrow k+1$;
end

From a computational burden standpoint, the algorithm proposed also has some advantages if we compare the number of bilinear terms present in (12) with the number of bilinear terms on the deterministic equivalent associated with (9) if we drop the robustification terms and set $v_r = \mathfrak{Q}_r(x)$ dirctly in the objective. In this case, our algorithm solves many problems with a smaller amount of bilinear terms, while the deterministic equivalent would solve one large problem with many bilinear terms.

Problem (12) can be solved with the package BARON by Tawarmalani and Sahinidis (2005), provided some attention is given to details. General hints on how this was achieved in our case study are listed below:

1. it is important to warm-start the calculations using previous iterates;
2. constraints in (12) are convex, a fact that should be informed to BARON;
3. the branching priorities experimentally. For instance, there is no need to branch in $x \in \mathcal{X}$.
4. We obtained better performance relaxing the default stopping tolerances. The default absolute and relative tolerances of BARON can be too tight for the application at hand, which might make the solver branch over "numerical trash".

5 Case Study

Our study refers to expansion of electricity generation under uncertain future market conditions. These market conditions can be, for example, demand for electricity, cost of production of electricity and cost of equipment. The original problem is taken from R. Birge and Louveaux (1997, Section 1.3). For our study we also want to consider ambiguity of probabilities. For this reason, we adapt a bit the original problem to highlight this issue.

Regarding invesments, they are planned for time steps $t \in \lfloor T \rceil$, deciding how much to invest for each type of technology indexed by $i \in \lfloor I \rceil$. The new amount of technology i, made available at time t, is denoted by x_i^t. The accumulated capacity of technology i at time t is denoted by w_i^t. The cost to install x_i^t is $c_i^t x_i^t$, and the maintenance cost of the accumulated production capacity w_i^t is $\eta_i^t w_i^t$. A technology i decided at time t takes B_i^t years to build and has a lifetime of L_i^t.

Regarding generation, costs are uncertain because they depend on uncertain market conditions. The cost of generating electricity given the installed capacity w at scenario $s \in \lfloor S \rceil$ is denoted by $\mathfrak{Q}_s(w)$. In the original version of R. Birge and Louveaux (1997), each scenario s is endowed with an exogenous probability p_s.

The decision problem is stated below, where $w_i^0 = 0$ and $x_i^t = 0$ for $t < 0$:

$$\begin{cases} \min_{x,w} & \sum_{t=1}^{T}\sum_{i=1}^{I} c_i^t x_i^t + \sum_{t=1}^{T}\sum_{i=1}^{I} \eta_i^t w_i^t + \sum_{s=1}^{S} p_s \mathfrak{Q}_s(w) \\ \text{s.t.} & w_i^t = w_i^{t-1} + x_i^{t-B_i^t} - x_i^{t-B_i^t-L_i^t}, \quad x, w \geq 0. \end{cases} \tag{13}$$

To understand the contribution of ambiguity sets for the probabilities, note first that, because making decisions in real-life in the long term is hard, the investor cannot assign a single distribution of probability to the uncertain events in the problem. Rather, the investor prefers to compute what would be the best possible cost, taking into account that the plausible probability distributions lie in some convex region $\mathcal{P}$. The goal of the DRSO problem is to check if in the best case the investment return is stable relative to the probabilities chosen. This is a necessary condition when investing in risky assets, especially in the long term.

After careful consideration of the political and technological situations, the decision maker defines some reference probability vectors for the outcomes of operational costs:

$$\left\{p^1, \ldots, p^L\right\} \subset (0,1)^{\rfloor S}$$

(each p^l is a different probability distribution in $\mathbb{R}^S$). For this finite number of probabilities, the investor knows that it is enough to solve L different instances of problem (13) and compute the minimum. However, for a more comprehensive analysis, it can be preferable to take decisions considering an infinite number of probabilities, for instance by setting

$$\mathcal{P} = \text{conv}\left\{p^1, \ldots, p^L\right\} \subset (0,1)^{\rfloor S}. \tag{14}$$

Introducing the simplicial variables α_l for $l \in \lfloor L \rfloor$, this gives the following nonsmooth and nonconvex problem, with decision-dependent probabilities:

$$\begin{cases} \min_{x,w,\alpha} & \sum_{t=1}^{T}\sum_{i=1}^{I} c_i^t x_i^t + \sum_{t=1}^{T}\sum_{i=1}^{I} \eta_i^t w_i^t + \sum_{s=1}^{S} \left\{\sum_{l=1}^{L} \alpha_l p_s^l\right\} \mathfrak{Q}_s(w) \\ \text{s.t.} & w_i^t = w_i^{t-1} + x_i^{t-B_i^t} - x_i^{t-B_i^t-L_i^t}, \quad \sum_{l=1}^{L} \alpha_l = 1, \quad x, w, \alpha \geq 0. \end{cases}$$

We next explain how the likelihood robustification radius $\kappa \geq 0$ comes into play. Having the mentioned stability goal in mind, our investor wants also to find out how the cost would change if instead of best case scenarios, also a worst case scenario were to be considered. However, not to exaggerate on the conservatism, this should be done in a continuous manner, for different degrees of pessimism, depending on the value of the parameter κ. When null, $\kappa = 0$ represents no pessimism while $\kappa = \infty$ represents total pessimism. Ultimately, because the plausible probabilities taken as reference may be slightly inaccurate, the investor wants to know how stable the decisions are with respect to the probabilities assigned to the events. For this case study, selecting probabilities comes at no cost, so $C_p \equiv 0$ and the ddDRSO formulation (9) for problem (13) is given by

$$\begin{cases} \min_{x,w,\alpha,v,\tau} & \sum_{t=1}^{T}\sum_{i=1}^{I} c_i^t x_i^t + \sum_{t=1}^{T}\sum_{i=1}^{I} \eta_i^t w_i^t + \sum_{s=1}^{S} \left\{\sum_{l=1}^{L} \alpha_l p_s^l\right\} v_s + \kappa\tau \\ \text{s.t.} & w_i^t = w_i^{t-1} + x_i^{t-B_i^t} - x_i^{t-B_i^t-L_i^t}, \quad \sum_{l=1}^{L} \alpha_l = 1, \quad x, w, \alpha, \tau \geq 0, \\ & v_s \geq \mathfrak{Q}_r(w) - \tau\delta_{rs}^{\mathbb{D}} \quad s, r \in \lfloor S \rfloor. \end{cases}$$

The recourse functions defining the values $\mathfrak{Q}_s(w)$ correspond to second-stage generation problems, for the accumulated capacity w, given the market condition s. Specifically, a time index $t \in \lfloor T \rfloor$ is divided into m modes of operation. A mode of operation at time t is characterized by a duration τ_j^t and a demand d_{sj}^t for electricity. The cost of generating energy with technology i and time t is d_{si}^t. For technology i at time t, there are deterministic values of existing capacity g_i^t and decommissioned amounts u_i^t. In scenario s, the capacity g_i^t is affected by a stochastic availability factor A_{sij}^t. The demand D_{sj}^t of electricity is satisfied with the generation y_{sij}^t, produced by technology i, time t, scenario s and mode of operation j. Altogether, this gives the recourse function below:

$$\mathfrak{Q}_s(w) := \begin{cases} \min\limits_{y} & \sum\limits_{t=1}^{T}\sum\limits_{i=1}^{I}\sum\limits_{j=1}^{m} d_{si}^t \tau_j^t y_{sij}^t \\ \text{s.t.} & \sum\limits_{i=1}^{I} y_{sij}^t = D_{sj}^t, \quad y_{sij}^t \leq A_{sij}^t (g_i^t - u_i^t + w_i^t), \quad y \geq 0. \end{cases}$$

Let us now explain the practical matters of the problem. We use the model above to make a simplified investment planning for Europe with yearly decisions from 2020 to 2050 using data from Granado, Skar, Doukas, and Trachanas (2018) for costs, existing installed capacity, yearly demand, lifetimes and building times. The technologies considered are coal-based, solar PV and wind onshore. The parameters for each technology, as well as their projected demand, are shown in Tables 1-5.

c_i^t	2020	2030	2040	2050
Coal-based	1500.0	1500.0	1500.0	1500.0
Solar PV	826.0	687.0	548.0	409.0
Wind Onshore	1033.0	972.6	911.6	851.0

Table 1: Investment costs (Euro/KWh).

η_i^t	2020	2030	2040	2050
Coal-based	31.1	31.1	31.1	31.1
Solar PV	18.6	15.7	12.86	10.0
Wind Onshore	52.6	50.8	49.0	47.3

Table 2: Maintenance costs (Euro/KWh).

d_{si}^t	2020	2030	2040	2050
Coal-based	0.5	0.5	0.5	0.5
Solar PV	0.0	0.0	0.0	0.0
Wind Onshore	0.0	0.0	0.0	0.0

Table 3: Variable generation costs (Euro/KWh).

B_i^t	2020	2030	2040	2050
Coal-based	3	3	3	3
Solar PV	3	2	2	1
Wind Onshore	3	2	2	1

Table 4: Building time (Years).

2020	2030	2040	2050
0.49	0.53	0.57	0.62

Table 5: Aggregate demand (TW/h).

In general, stochastic optimization is used to achieve the diversification effect for the portfolio of electricity generation equipment. However, due to our simplified modeling this diversification does not appear naturally. For this purpose, the complementarity between wind and solar generation is built into the model using m different modes of generation. More precisely, each time period has two modes, one only for wind (60%) and another for solar (40%). The mode duration is taken proportionally. The availability factor for wind has mean 0.35 and for solar has mean 0.5. Both have variance equal to 50% of the mean. The mean for the demand is reported in Table 5 and its variance is 30% of the mean. The scenarios are generated sampling the respective quantities using a normal distribution, $\mathcal{N}$, with the corresponding mean and variances.

Also because of the simplified modeling, in the master problem we have to add a linear constraint limiting the investment at each time period to at most 12% of total installed capacity. The lifetime for coal-based generation is 30 years, while for wind onshore it is 20 years, and for solar PV 25 years. The decommissioning rate is assumed to be linear so that at the end of the period all thermal power plants are decommissioned. The initial installed capacity of coal-based power is 0.92 TW/h and the other ones are assumed to be zero.

As mentioned, we use the total variation distance for the transportation costs, as in (10). All experiments are run on a notebook with Intel i7 1.90GHz processor, running under Ubuntu 18.04.3 LTS, and using CPLEX 12.10 and Julia 1.1.1; see Lubin and Dunning (2015).

Below, results are reported only until 2040 because, as often in this setting, the model exhibits abnormal when approaching the end of the horizon.

As explained as the end of Section 3, the likelihood robustified expectation (7) approaches the value of the worst case scenario for sufficiently large κ. For this reason, we computed the min-max expansion for $S = 30$ scenarios, shown in Figure 1. This output, known to be cost-consertive is used as a reference for comparison. We expect the (less conservative) DDRSO variants to make more investments.

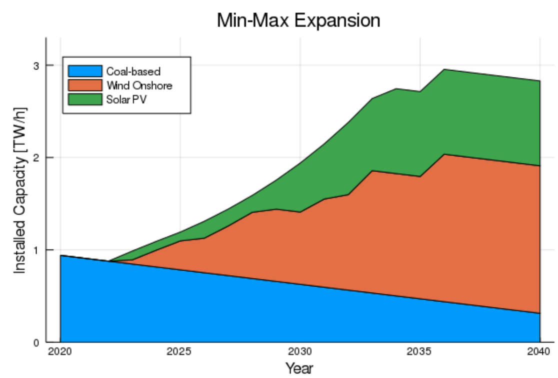

Figure 1: Resulting mix with the min-max expansion.

The comparison is done with the expansions obtained with the following two sets $\mathcal{P}$ as in (14), with

$$\begin{array}{llll} p^l & \equiv & \frac{1}{S} & l = 1, \dots, 10 \quad \text{denoted by Simple} \\ p^l & \sim & \mathcal{N}(\frac{1}{S}, 0.20) & l = 1, \dots, 10 \quad \text{denoted by Opt-robust}. \end{array}$$

The resulting mixes, obtained by varying the likelihood robustification radius $\kappa \in \{0, 0.1, 0.2\}$, are given in Figure 2, 3, and 4, respectively. All experiments take less than one minute to finish using the decomposition method.

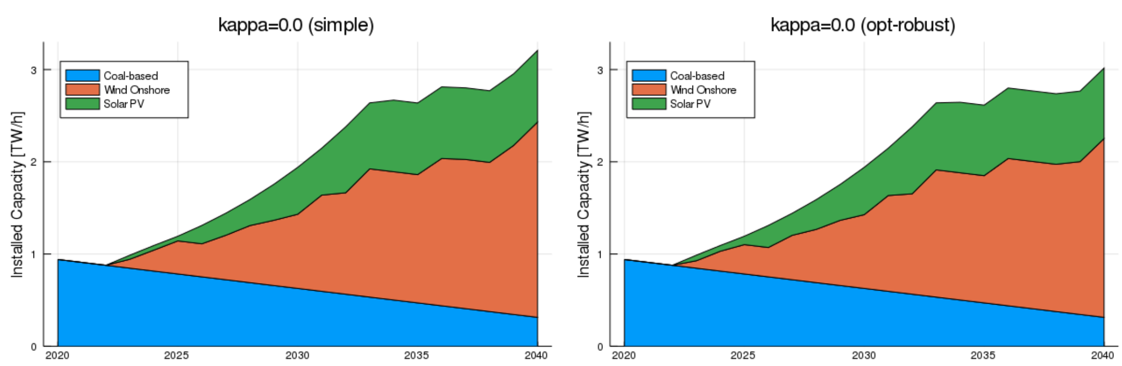

Figure 2: Mix with the likelihood robustification radius $\kappa = 0$.

As is seen in Figure 2, the strategy opt-robust makes less investments than the simple, but more investments than the min-max. This is expected from the definitions of the strategies themselves. Notice also that, in all the runs, the amount of coal-based generation decreases linearly due to decommissioning, and because it is not cost-effective.

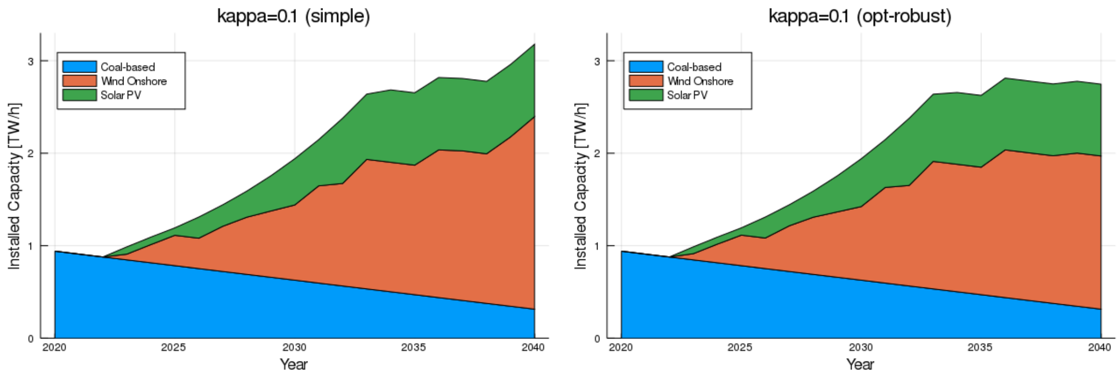

Figure 3: Mix with the likelihood robustification radius $\kappa = 0.1$.

As the ratio κ increases, we observe that the alternative strategies get closer to the min-max, shown in Figure 1. This is clear for the opt-robust variant in Figure 3, with $\kappa = 0.1$, even though the investment with opt-robust remains higher than with min-max. The simple strategy did not change with respect to $\kappa = 0$.

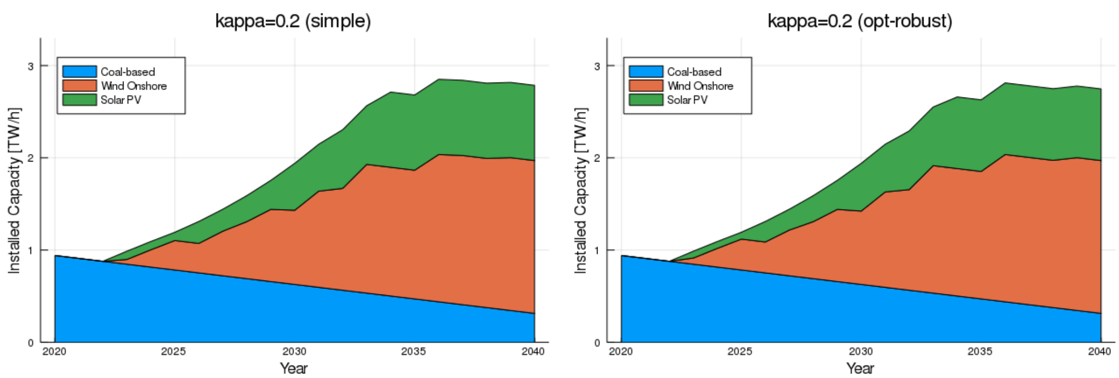

Figure 4: Mix with the likelihood robustification radius $\kappa = 0.2$.

For $\kappa = 0.2$ both simple and opt-robust strategies approach the min-max solution in Figure 3. We believe that the good agreement with the min-max strategy, clear already for small κ, is due to the choice of ambiguity set $\mathcal{P}$, which is narrow both for the simple and opt-robust variants. For other choices of $\mathcal{P}$ the behavior is likely to be different.

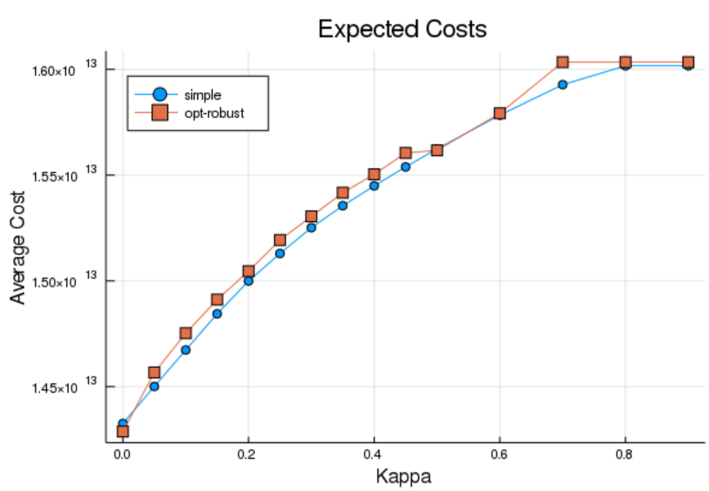

Figure 5: The evolution of costs as a function of κ.

Figure 5, with the evolution of costs as a function of κ, reveals the *opt-robust* option as the most expensive one, except for $\kappa = 0$. The ambiguity set just affects the probabilities for small κ. For large κ, the event with the worst outcome receives total probability, as shown by Noyan, Rudolf, and Lejeune (2018).

Additionally, notice in Figure 5 the smooth behaviour of the cost when the robustification ratio κ varies. This confirms the expectation that our approach brings some stability to the decision-making process (with respect to variations in the probability distribution defining the problem).

To conclude, we examine if our decomposition algorithm improves solution times. The advantage of decomposition is that the sequence of problems solved has either a much smaller number of bilinear terms or a much smaller number of variables. As illustrated by Ackooij, Frangioni, and Oliveira (2016), decomposition methods, that typically scale well, have a strong advantage when solution times of combinatorial or nonconvex problems increase dramatically with the dimension of the problem. This is confirmed by our results in the tables below. Another remark is that, since BARON uses a nonlinear programming solver in the local search procedure, issues related to precision, robustness and time tend to affect the deterministic equivalent of the problem more strongly as the number of variables grows. In Tables 6 and 7, we show statistics for the solution times varying the number of scenarios with ten probabilities ($L = 10$) and some values of κ. Each experiment is repeated four times. The structure of the problem is the same for all values of κ, and therefore, the decomposition method works just the same because the solution times of the nonconvex master problem dominates the execution.

S	Dec. (Avg.)	Dec. (Std.)	Equi. (Avg.)	Equi. (Std.)
10	11.1	1.8	27.2	0.1
20	13.6	0.5	320.0	1.3
30	37.4	0.7	293.2	0.9
40	46.1	1.8	543.4	2.2
50	60.3	3.3	-	-

Table 6: Comparison of solution times (in seconds) of the decomposition method and the full deterministic equivalent of the problem, for $L = 10$ and $\kappa = 0.0$.

S	Dec. (Avg.)	Dec. (Std.)	Equi. (Avg.)	Equi. (Std.)
10	12.5	2.1	23.9	0.1
20	12.3	0.5	246.4	0.9
30	42.6	0.6	220.2	0.8
40	52.0	1.5	450.3	2.1
50	54.3	3.4	-	-

Table 7: Comparison of solution times (in seconds) of the decomposition method and the full deterministic equivalent of the problem, for $L = 10$ and $\kappa = 0.2$

In summary, our likelihood robustified optimistic problem provides more conceptual flexibility than the min-max and min-max regret approaches to try to obtain better out-of-sample operational performance on a given set of scenarios after the expansion plan is decided. In practice, the value of κ can be selected having these remarks in mind, besides also being useful to check for the stability of the current decisions. Specifically in our case, the more technology is built, the smaller the out-of-sample cost to meet the demand, because less slack generation needs to be activated, and the dispatch follows the order of merit. In other words, the belief on ambiguity sets of probabilities determines how much technology should be built, and the more technology built the less the out-of-sample operational cost.

6 Final Comments

In this paper, we explore another paradigm for decisions under probability ambiguity, placed between the optimistic and the pessimistic ones. We also show how to use global bilinear optimization and scenario-wise decomposition to solve the resulting nonsmooth and nonconvex problems to global optimality. We illustrate our approach on a case study in energy management.

Sensitivity with respect to probabilities is not important if a large volume of data is available. However, if the underlying planning task is of long-term nature, like capacity and transmission expansion in energy systems, it becomes relatively more important to look at qualitative changes of the decisions suggested by planning models. With this perspective, trying to identify nearby problem data that would change the qualitative aspect of the decisions is appealing. For instance, one can ask what is the change in the probabilities that drives a profitable activity to change to a non-profitable one. This type of analysis can be done with our proposed framework.

The min-max and min-max regret strategies in Granado, Skar, Doukas, and Trachanas (2018) are often used in the capacity and transmission expansion planning of power systems. Neither of these measures consider robustness with respect to probability distributions. The probability of ocurrence of events does not enter the formulation of these approaches, that are well-known for their conservatism. The same strategy introduced in this work for likelihood robustification can be used to robustify important risk-measures like the CVaR.

The use of convex combinations of probability distributions was employed in Hellemo, Barton, and Tomasgard (2018) for models where the parameters of probability distributions can be influenced by investment decisions. Models with endogenous uncertainty are usually nonconvex and their computational solution might be time-consuming or unreliable. Both the ddDRSO model (9) and the decomposition Algorithm 2 remain applicable if the robustification parameter or the probability set depend on the first-stage variable, as long as the latter remains a convex set (in (9) $\kappa = \kappa(x)$ or in (14) $\mathcal{P} = \mathcal{P}(x)$).

When compared to the settings discussed by Noyan, Rudolf, and Lejeune (2018) and Kuhn, Esfahani, Nguyen, and Shafieezadeh-Abadeh (2019), our main contribution is in exploiting the specific structure of the two-stage problem to design an efficient decomposition algorithm. Finally note that, while most of the numerical schemes for distributionally robust optimization entail the solution of convex models, our approach is nonconvex but reamins computationally tractable, thanks to decomposition.

Acknowledgments This work was supported by the TACEMM project (Trans-Atlantic Cooperation on Energy Market Models). Research of the second author is partly funded by CNPq Grant 306089/2019-0, CEPID CeMEAI, and FAPERJ in Brazil. The third author is supported by CNPq Grant 303913/2019-3, by FAPERJ Grant E-26/202.540/2019, and by PRONEX–Optimization.

References

Ackooij, W. van, A. Frangioni, and W. de Oliveira (May 2016). “Inexact stabilized Benders’ decomposition approaches with application to chance-constrained problems with finite support”. *Computational Optimization and Applications* 65.3, pp. 637–669.

Bayraksan, G. and D. K. Love (Sept. 2015). “Data-Driven Stochastic Programming Using Phi-Divergences”. In: *The Operations Research Revolution*. INFORMS, pp. 1–19.

Ben-Tal, A. and M. Teboulle (May 1987). “Penalty Functions and Duality in Stochastic Programming Via ϕ-Divergence Functionals”. *Mathematics of Operations Research* 12.2, pp. 224–240.

Ben-Tal, A., L. E. Ghaoui, and A. Nemirovski (Aug. 2009). *Robust Optimization*. Princeton University Press.

Bertsimas, D., X. V. Doan, K. Natarajan, and C.-P. Teo (Aug. 2010). “Models for Minimax Stochastic Linear Optimization Problems with Risk Aversion”. *Mathematics of Operations Research* 35.3, pp. 580–602.

Calafiore, G. C. (Jan. 2007). “Ambiguous Risk Measures and Optimal Robust Portfolios”. *SIAM Journal on Optimization* 18.3, pp. 853–877.

Colvin, M. and C. T. Maravelias (2009). “A Branch and Cut Framework for Multi-Stage Stochastic Programming Problems Under Endogenous Uncertainty”. In: *Computer Aided Chemical Engineering*. Elsevier, pp. 255–260.

Delage, E. and Y. Ye (June 2010). “Distributionally Robust Optimization Under Moment Uncertainty with Application to Data-Driven Problems”. *Operations Research* 58.3, pp. 595–612.

Dempe, S., J. Dutta, and B. S. Mordukhovich (Oct. 2007). “New necessary optimality conditions in optimistic bilevel programming”. *Optimization* 56.5-6, pp. 577–604.

Esfahani, P. M. and D. Kuhn (July 2017). “Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations”. *Mathematical Programming* 171.1-2, pp. 115–166.

Fábián, C. I. and Z. Szőke (Aug. 2006). "Solving two-stage stochastic programming problems with level decomposition". *Computational Management Science* 4.4, pp. 313–353.
Goel, V. and I. E. Grossmann (July 2004). "A stochastic programming approach to planning of offshore gas field developments under uncertainty in reserves". *Computers & Chemical Engineering* 28.8, pp. 1409–1429.
Goh, J. and M. Sim (Aug. 2010). "Distributionally Robust Optimization and Its Tractable Approximations". *Operations Research* 58.4-part-1, pp. 902–917.
Granado, P. C. del, C. Skar, H. Doukas, and G. P. Trachanas (Dec. 2018). "Investments in the EU Power System: A Stress Test Analysis on the Effectiveness of Decarbonisation Policies", pp. 97–122.
Hanasusanto, G. A. and D. Kuhn (June 2018). "Conic Programming Reformulations of Two-Stage Distributionally Robust Linear Programs over Wasserstein Balls". *Operations Research* 66.3, pp. 849–869.
Hellemo, L., P. I. Barton, and A. Tomasgard (Aug. 2018). "Decision-dependent probabilities in stochastic programs with recourse". *Computational Management Science* 15.3-4, pp. 369–395.
Jonsbråten, T. W., R. Wets, and D. L. Woodruff (Aug. 1998). "A class of stochastic programs with decision dependent random elements". *Annals of Operations Research* 82.0, pp. 83–106.
Kuhn, D., P. M. Esfahani, V. A. Nguyen, and S. Shafieezadeh-Abadeh (Oct. 2019). "Wasserstein Distributionally Robust Optimization: Theory and Applications in Machine Learning". In: *Operations Research & Management Science in the Age of Analytics*. INFORMS, pp. 130–166.
Leyffer, S., M. Menickelly, T. Munson, C. Vanaret, and S. M. Wild (2020). "A survey of nonlinear robust optimization". *INFOR: Information Systems and Operational Research* 58.2, pp. 342–373.
Love, D. and G. Bayraksan (Dec. 2013). "Two-stage likelihood robust linear program with application to water allocation under uncertainty". In: *2013 Winter Simulations Conference (WSC)*. IEEE.
Lubin, M. and I. Dunning (2015). "Computing in Operations Research Using Julia". *INFORMS Journal on Computing* 27.2, pp. 238–248.
Luo, F. and S. Mehrotra (Apr. 2020). "Distributionally robust optimization with decision dependent ambiguity sets". *Optimization Letters*.
Nohadani, O. and K. Sharma (Jan. 2018). "Optimization under Decision-Dependent Uncertainty". *SIAM Journal on Optimization* 28.2, pp. 1773–1795.
Noyan, N., G. Rudolf, and M. Lejeune (2018). *Distributionally Robust Optimization with Decision-Dependent Ambiguity Set*. Tech. rep. Optimization Online.
Oliveira, W., C. Sagastizábal, and S. Scheimberg (Apr. 2011). "Inexact Bundle Methods for Two-Stage Stochastic Programming". *SIAM Journal on Optimization* 21.2, pp. 517–544.
Pflüg, G. C., A. Pichler, and D. Wozabal (Feb. 2012). "The 1/N investment strategy is optimal under high model ambiguity". *Journal of Banking & Finance* 36.2, pp. 410–417.
Pflüg, G. and D. Wozabal (Aug. 2007). "Ambiguity in portfolio selection". In: vol. 7. 4. Informa UK Limited, pp. 435–442.
Postek, K., D. den Hertog, and B. Melenberg (2014). "Tractable Counterparts of Distributionally Robust Constraints on Risk Measures". *SSRN Electronic Journal*.
R. Birge, J. and F. Louveaux (1997). *Introduction to Stochastic Programming*. Springer-Verlag.
Sagastizábal, C. (June 2012). "Divide to conquer: decomposition methods for energy optimization". *Mathematical Programming* 134.1, pp. 187–222.
Tawarmalani, M. and N. V. Sahinidis (May 2005). "A polyhedral branch-and-cut approach to global optimization". *Mathematical Programming* 103.2, pp. 225–249.
Van Slyke, R. M. and R. Wets (July 1969). "L-Shaped Linear Programs with Applications to Optimal Control and Stochastic Programming". *SIAM Journal on Applied Mathematics* 17.4, pp. 638–663.
Wang, Z., P. W. Glynn, and Y. Ye (Sept. 2015). "Likelihood robust optimization for data-driven problems". *Computational Management Science* 13.2, pp. 241–261.
Wiesemann, W., A. Tsoukalas, P.-M. Kleniati, and B. Rustem (Jan. 2013). "Pessimistic Bilevel Optimization". *SIAM Journal on Optimization* 23.1, pp. 353–380.
Wozabal, D. (Nov. 2010). "A framework for optimization under ambiguity". *Annals of Operations Research* 193.1, pp. 21–47.
Xie, W. (July 2020). "Tractable reformulations of two-stage distributionally robust linear programs over the type-∞ Wasserstein ball". *Operations Research Letters* 48.4, pp. 513–523.