Noname manuscript No.
(will be inserted by the editor)

Optimization under Connected Uncertainty

Omid Nohadani[1] · **Kartikey Sharma**[1]

Received: date / Accepted: date / Version:

Abstract Distributionally robust and standard robust optimization methods provide a tractable way to address uncertainties. Recently, their efficacy has been extended to multistage settings, where the uncertainty at each stage is independent of the past. However, in many applications, past observations influence future uncertainties. In this paper, we leverage this dependence via connected uncertainty sets, where the set parameters at each period depend on previous realizations. To find optimal here-and-now solutions, we reformulate distributionally robust and standard robust constraints for connected uncertainty sets. We illustrate the advantages of this framework with two applications.

Keywords Distributionally Robust Optimization · Robust Optimization · Connected Sets.

1 Introduction

Robust optimization (RO) has gained considerable popularity in recent years for its ability to provide tractable solutions to problems under uncertainty [4, 6]. This is manifested in the number of applications of RO to real-world problems, such as healthcare [14], unit commitment [23], and queueing [3], amongst others. In RO, the nature of uncertainty is captured by sets that bound the uncertain parameters without further assumptions. When the uncertain component is assumed to follow distributions, distributionally robust optimization (DRO) offers a probabilistic alternative to RO with uncertainty sets replaced by ambiguity sets over distributions. These sets can be characterized by moments [39], by distance measures [25], or by hypothesis tests [8]. DRO was introduced by Scarf et al. [33] in the context of a newsvendor problem. Ever since, it has been applied to various applications, such as Markov decision processes [41], machine learning [30], and chance constrained programming [13], to name a few.

For both RO and DRO, the structure and tractability of the problem decisively depend on the size and geometry of the underlying uncertainty sets. The size controls the "magnitude" of possible uncertainties, to which the solution is immune. It also determines the probabilistic guarantee of constraint satisfaction. The geometry, on the other hand, determines the computational tractability of the formulation. For example, certain combinatorial RO problems achieve a tractable reformulation when the uncertain objective coefficients reside in a cardinality constrained set but not necessarily otherwise [9]. Similarly, Delage and Ye [17] leverage the structure of the DRO problems to obtain tractable formulations. However, many DRO problems do not have polynomial time algorithms [29].

In recent years the efficacy of optimization under uncertainty has been extended to multistage problems. The standard RO paradigm of static (here-and-now) solutions involves decisions which do not adapt to uncertainty realizations. However, this approach leads to highly conservative solutions. In order to allow for less conservative decisions and to better model reality, Ben-Tal et al. [5] introduced the notion of adjustable robust optimization (ARO) to accurately capture the *wait-and-see* nature of decisions by allowing them to adapt to the realization of the uncertainty. This allows ARO to achieve considerable improvement in solution quality over static solutions. However this comes at the cost of

O. Nohadani
E-mail: nohadani@northwestern.edu

K. Sharma
E-mail: kartikeysharma2014@u.northwestern.edu
[1] Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL 60208

higher computational complexity. Decision rules provide a smooth trade-off between complexity of ARO and and solution quality [24].

In the context of multistage DRO, solutions adapt to the realization of the uncertainty instead of adapting to the realized distribution. Non-anticipative decision rules can be leveraged to provide tractable reformulations for moment-based uncertainty sets [18]. Furthermore, adaptability has been extended to ambiguity sets defined by the Wasserstein metric with a conic reformulation for a two-stage DRO problem [19]. Lastly, in scenario-tree problems, the uncertainty sets were defined using a nested distance measure and the solution is optimized against all distributions within the set [2].

In conventional multistage problems, the uncertain parameters for each stage are modeled to reside in separate sets (for RO, see [7, 22, 28, 31] and for DRO, see [16, 35]). Such structures enable the tractability of a broad range of problems. However, in many applications, the uncertainty at a stage depends on the realization of the uncertain parameters in previous stages. Often, data exhibit autocorrelations, which cannot be modeled with independent sets. For a newsvendor model with autoregressive demand several examples were provided where the demand for commodities is autocorrelated [1]. Such correlations also occur in forecasting weather [12] and in predicting stock returns [26], amongst other phenomena.

In this paper, we seek to provide a step towards allowing uncertainty sets to depend on previous realizations for both DRO and RO problems. We consider problems with multiple time periods where all decisions are taken at the beginning, but the uncertainty at each period is influenced by the uncertainty realizations in the past. First we study DRO problems with connected uncertainty (CU) sets, i.e., sets whose parameters depend on previously realized uncertainties. We provide a portfolio optimization application that illustrates the advantages gained from leveraging this framework We then extend this approach to RO settings in order to exhibit alternative models. Here, we furbish the theoretical results with numerical experiments on a knapsack problem. Figure 1 schematically compares the uncertainty coverage of standard and connected sets. Note that at later periods, some sample paths may not reside within fixed sets.

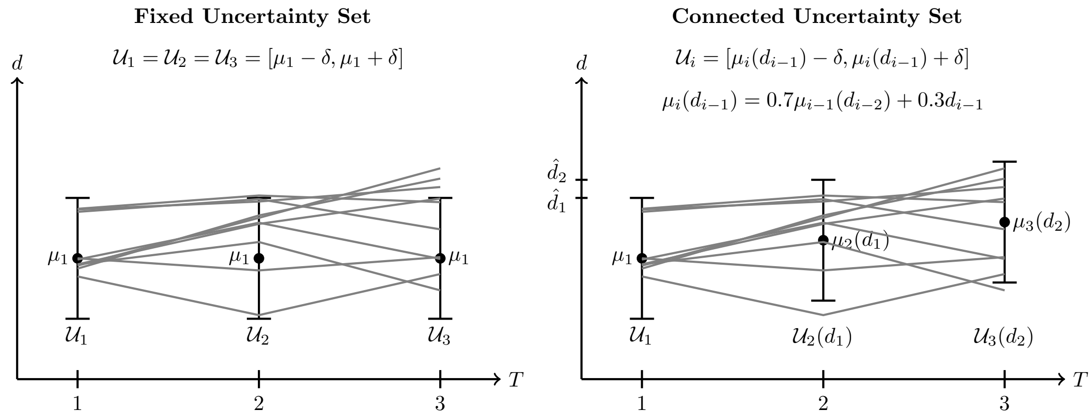

Fig. 1 Illustration of conventional (left) and connected uncertainty sets (right) for different sample paths of d which follow $d_i = d_{i-1} + \epsilon$. The connected sets are updated for specific values $d_i = \hat{d}_i$.

To illustrate the objective of this paper, we start with two examples that exhibit CU sets. The first example is of a practical nature, whereas the second one displays the mathematical framework of these class of problems.

Example 1: A frequent traveler on a road network seeks to minimize travel time. On each trip, the knowledge acquired from previous trips can contribute towards better planning. Since each road segment is subject to uncertain traffic, the aggregate experience can improve forecasts and lead to better estimates of uncertainty and hence improve decisions. This aggregate experience can be captured using CU sets. This and many other examples have in common that, when portions of the problem are reused, knowledge from the past can be leveraged towards improved decisions.

Example 2: Consider the following RO problem which aims to give intuition:

$$\begin{aligned} \min_{\mathbf{x}_1,\mathbf{x}_2} \; & \mathbf{c}_1^\top \mathbf{x}_1 + \mathbf{c}_2^\top \mathbf{x}_2 \\ \text{s.t. } & \mathbf{d}_1^\top \mathbf{x}_1 + \mathbf{d}_2^\top \mathbf{x}_2 \le B \;\; \forall \mathbf{d}_2 \in \mathcal{U}_2(\mathbf{d}_1) \;\; \forall \mathbf{d}_1 \in \mathcal{U}_1 \\ & \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X} \\ & \mathbf{x}_1, \mathbf{x}_2 \ge \mathbf{0}, \end{aligned} \tag{1}$$

where $\mathbf{x}_1, \mathbf{x}_2$ are first and second period decision variables in $\mathcal{X} \subseteq \mathbb{R}^n$, $\mathbf{c}_1, \mathbf{c}_2 \in \mathbb{R}^n$ are known, $\mathbf{d}_1, \mathbf{d}_2 \in \mathbb{R}^n$ are uncertain coefficients, and B is the RHS (right-hand side) coefficient. The robust counterpart of the uncertain constraint can be written as

$$\max_{\mathbf{d}_1 \in \mathcal{U}_1} \{\mathbf{d}_1^\top \mathbf{x}_1 + \max_{\mathbf{d}_2 \in \mathcal{U}_2(\mathbf{d}_1)} \mathbf{d}_2^\top \mathbf{x}_2\} \leq B. \tag{2}$$

For problem (1), consider the uncertainty sets

$$\mathcal{U}_1 = \{\mathbf{d}_1 \mid \mathbf{A}_1 \mathbf{d}_1 = \mathbf{b}_1, \mathbf{d}_1 \geq 0\}, \quad \mathcal{U}_2(\mathbf{d}_1) = \{\mathbf{d}_2 \mid \mathbf{A}_2 \mathbf{d}_2 = \mathbf{b}_2 + \boldsymbol{\Delta} \mathbf{d}_1, \mathbf{d}_2 \geq 0\},$$

where $\mathbf{A}_1 \in \mathbb{R}^{m_1 \times n}, \mathbf{A}_2 \in \mathbb{R}^{m_2 \times n}$, and $\boldsymbol{\Delta} \in \mathbb{R}^{m_2 \times n}$ is a coefficient matrix that determines how $\mathcal{U}_2$ depends on $\mathbf{d}_1$. In this setting, the RHS coefficients of $\mathcal{U}_2$ linearly depend on the realization of the uncertain parameter $\mathbf{d}_1$. A corresponding distributionally robust version of constraint (2) is

$$\sup_{P_1 \in \widetilde{\mathcal{U}}_1} \mathbb{E}_{\mathbf{d}_1 \sim P_1} \left[\mathbf{d}_1^\top \mathbf{x}_1 + \sup_{P_2 \in \widetilde{\mathcal{U}}_2(\mathbf{d}_1)} \mathbb{E}_{\mathbf{d}_2 \sim P_2} \left[\mathbf{d}_2^\top \mathbf{x}_2 \right] \right] \leq B. \tag{3}$$

Note that $\widetilde{\mathcal{U}}_1$ and $\widetilde{\mathcal{U}}_2(\mathbf{d}_1)$ represent sets of distributions as opposed to sets of vectors in $\mathcal{U}_1$ and $\mathcal{U}_2(\mathbf{d}_1)$. A detailed description of constraint (3) is provided in Sect. 2. This example illustrates the structure of problems that encounter CU sets.

The concept of interstage dependence in multistage problems is well studied in the context of stochastic optimization and is naturally embedded in scenario trees. The probability of future scenarios may depend on the current branch [37]. However, common stochastic models usually assume independence which allows for simplified decomposition algorithms and sharing of generated cuts among other scenarios in the same stage. The L-shaped method is a prominent example [21]. In general, sharing cuts is not permitted when the scenarios are dependent. However, if the underlying dependency of the uncertain coefficients follows simple time series models, these cuts can be modified and shared across the scenarios [15, 21].

In RO, a typical uncertainty model involves using a single set for all periods. Depending on the set geometry, different periods can be connected or not. The recent popularity of budgeted uncertainty sets has led to the introduction of sets which have budgets on the uncertainty across multiple periods [11, 23, 32, 43]. This couples the uncertainty realizations across periods. Building on this concept, the notion of dynamic uncertainty sets was introduced where the set at each period depended on the uncertainty realizations in all previous periods [27]. The primary motivation was the use of a time series model to capture fluctuating demand. In the context of DRO, the scenario tree formulation incorporated such connections across uncertainty sets [2]. Such distributional uncertainty sets have also been used in an inventory control problem [40].

Notation. Throughout this paper, we use bold lowercase and uppercase letters to denote vectors and matrices. Scalars are marked in regular font. All vectors are column vectors. $\mathbf{A} \cdot \mathbf{B}$ denotes the matrix dot product $\sum_{ij} A_{ij} B_{ij}$. In various locations the type of connection amongst the sets is referred to as "uncertainty dependent." To streamline the exposition, we use "uncertainty set" for both the RO and DRO settings. The former is over parameters and the latter over distributions and is also known as an ambiguity set. All problems in this paper yield *here-and-now* decisions.

2 DRO with CU sets

In this section, we formalize the concept of CU sets in the distributional context. In the general DRO setting, uncertainty sets are sets of all possible distributions with specified parameters. In the case of CU sets for multiperiod settings, these conditions may depend on previous realizations. In the introductory Example 2, the CU sets can be given by

$$\widetilde{\mathcal{U}}_1 = \left\{ P_1 \in \mathcal{M} \middle| P_1(\mathbf{d}_1 \in \Xi_1) = 1,\ \underline{\boldsymbol{\mu}}_1 \leq \mathbb{E}_{P_1}[\mathbf{d}_1] \leq \overline{\boldsymbol{\mu}}_1,\ \mathbb{E}_{P_1}[(\mathbf{d}_1 - \boldsymbol{\mu}_1)(\mathbf{d}_1 - \boldsymbol{\mu}_1)^\top] \preceq \boldsymbol{\Sigma}_1 \right\},$$

$$\widetilde{\mathcal{U}}_2(\mathbf{d}_1) = \left\{ P_{2|1} \in \mathcal{M} \middle| P_{2|1}(\mathbf{d}_t \in \Xi_2) = 1,\ \underline{\boldsymbol{\mu}}_2(\mathbf{d}_1) \leq \mathbb{E}_{P_{2|1}}[\mathbf{d}_2] \leq \overline{\boldsymbol{\mu}}_2(\mathbf{d}_1), \right.$$
$$\left. \mathbb{E}_{P_{2|1}}[(\mathbf{d}_2 - \boldsymbol{\mu}_2(\mathbf{d}_1))(\mathbf{d}_2 - \boldsymbol{\mu}_2(\mathbf{d}_1))^\top] \preceq \boldsymbol{\Sigma}_2(\mathbf{d}_1) \right\},$$

where $P_{2|1}(\mathbf{d}_2)$ denotes the conditional distribution of $\mathbf{d}_2$ given $\mathbf{d}_1$. Here, Ξ_1 and Ξ_2 are the compact support sets of $\mathbf{d}_1$ and $\mathbf{d}_2$ and $\mathcal{M}$ denotes the set of all positive measures. The parameters $\underline{\boldsymbol{\mu}}_1 < \boldsymbol{\mu}_1 < \overline{\boldsymbol{\mu}}_1$ and $\boldsymbol{\Sigma_1}$ are constants and $\underline{\boldsymbol{\mu}}_2(\mathbf{d}_1) < \boldsymbol{\mu}_2(\mathbf{d}_1) < \overline{\boldsymbol{\mu}}_2(\mathbf{d}_1)$ and $\boldsymbol{\Sigma_2}(\mathbf{d}_1)$ are functions of the realizations of $\mathbf{d}_1$. For example, these could be marginal updates of the moments. As illustrated in $\widetilde{\mathcal{U}}_1$ and $\widetilde{\mathcal{U}}_2(\mathbf{d}_1)$, we consider distributional uncertainty sets to be *connected*, when the mean and moment of all distributions in the uncertainty set of a given period depend on the realizations from the previous period.

The aim of this section is to reformulate the constraint

$$\mathbb{E}_P[\sum_{t=1}^{T} h_t(\mathbf{x}_t, \mathbf{d}_t)] \leq B, \tag{C-DRO}$$

which combines T periods. The expectation $\mathbb{E}[\cdot]$ is taken with respect to the joint distribution P of all $\mathbf{d}_t$. Furthermore, $h_t(\mathbf{x}_t, \mathbf{d}_t)$ is a function of the decision variable $\mathbf{x}_t \in \mathbb{R}^{n_t}$ and the uncertain parameter $\mathbf{d}_t \in \mathbb{R}^m$. The dimension of $\mathbf{d}_t$ shall be constant for the clarity of exposition. Unless specified, we do not make any assumptions on the structure of $h_t(\cdot,\cdot)$ beyond regularity conditions required for the existence of integrals. Each $\mathbf{d}_t$ has a distribution that lies in a different uncertainty set and each set depends on the previous realization $\mathbf{d}_{t-1}$.

2.1 Reformulation

The first part of the reformulation transforms the uncertain constraint (C-DRO) into its robust counterpart over CU sets. The second part focuses on reformulating the counterpart into an infinite dimensional optimization problem (IOP), in which the variables are functions of uncertainty realizations (analogous to the adjustable problem). We then show that the complexity of the IOP formulation can be reduced by using static solutions.

Consider the following CU sets

$$\begin{aligned}\widetilde{\mathcal{U}}_t(\mathbf{d}_{t-1}) = \Big\{ P_{t|t-1} \in \mathcal{M} \Big| P_{t|t-1}(\mathbf{d}_t \in \Xi_t) = 1,\ \underline{\boldsymbol{\mu}}_t(\mathbf{d}_{t-1}) \leq \mathbb{E}_{P_{t|t-1}}[\mathbf{d}_t] \leq \overline{\boldsymbol{\mu}}_t(\mathbf{d}_{t-1}),\\ \mathbb{E}_{P_{t|t-1}}[(\mathbf{d}_t - \boldsymbol{\mu}_t(\mathbf{d}_{t-1}))(\mathbf{d}_t - \boldsymbol{\mu}_t(\mathbf{d}_{t-1}))^\top] \preceq \boldsymbol{\Sigma}_t(\mathbf{d}_{t-1}) \Big\},\end{aligned} \tag{D}$$

where $\widetilde{\mathcal{U}}_t(\mathbf{d}_{t-1})$ denotes the set of conditional distributions for $\mathbf{d}_t$ (conditioned on $\mathbf{d}_{t-1}$). For $t = 1$, all parameters are known, as in $\widetilde{\mathcal{U}}_1$. Note that the parameters depend on the realization of only the previous period. Observe that the last constraint in $\widetilde{\mathcal{U}}_t(\mathbf{d}_{t-1})$ is equivalent to

$$\mathbb{E}_{P_{t|t-1}}[\mathbf{d}_t\mathbf{d}_t^\top] \preceq \boldsymbol{\Sigma}_t(\mathbf{d}_{t-1}) + \boldsymbol{\mu}_t(\mathbf{d}_{t-1})\boldsymbol{\mu}_t(\mathbf{d}_{t-1})^\top.$$

For clarity, we define the RHS $\boldsymbol{\Sigma}_t^{'}(\mathbf{d}_{t-1}) \equiv \boldsymbol{\Sigma}_t(\mathbf{d}_{t-1}) + \boldsymbol{\mu}_t(\mathbf{d}_{t-1})\boldsymbol{\mu}_t(\mathbf{d}_{t-1})^\top$, which will be used later.

Proposition 1 *Given the sets* $\widetilde{\mathcal{U}}_1, \ldots, \widetilde{\mathcal{U}}_T(\mathbf{d}_{T-1})$*, the robust counterpart of constraint* (C-DRO) *is*

$$\sup_{P_1 \in \widetilde{\mathcal{U}}_1} \mathbb{E}_{P_1}\left[h_1(\mathbf{x}_1, \mathbf{d}_1) + \sup_{P_{2|1} \in \widetilde{\mathcal{U}}_2(\mathbf{d}_1)} \left\{ \mathbb{E}_{P_{2|1}}\left[h_2(\mathbf{x}_2, \mathbf{d}_2) + \cdots + \sup_{P_{T|T-1} \in \widetilde{\mathcal{U}}_T(\mathbf{d}_{T-1})} \left\{ \mathbb{E}_{P_{T|T-1}}\left[h_T(\mathbf{x}_T, \mathbf{d}_T) \right] \right\} \right] \right\} \right] \leq B. \tag{4}$$

Proof Let $P_1 \in \widetilde{\mathcal{U}}_1$ be the distribution of $\mathbf{d}_1$ and for each $\mathbf{d}_{t-1}$ let $P_{t|t-1} \in \widetilde{\mathcal{U}}_t(\mathbf{d}_{t-1})$ denote the distribution of $\mathbf{d}_t$ conditioned on $\mathbf{d}_{t-1}$ $\forall t = 2, \ldots, T$. Define P to be the joint distribution of all $\mathbf{d}_1, \ldots, \mathbf{d}_T$. Note that by construction, each distribution $P_{t|t-1}$ only depends on a given $\mathbf{d}_{t-1}$ and not further back. Thus, given $P_1, \ldots, P_{T|T-1}$, the joint distribution can be defined by $P = P_1 \times P_{2|1} \times \cdots \times P_{T|T-1}$. Then P lies in a joint uncertainty set $\widetilde{\mathcal{U}}$ constructed using $\widetilde{\mathcal{U}}_t(\mathbf{d}_{t-1})$ for each $\mathbf{d}_{t-1} \in \Xi_{t-1}$ as

$$\widetilde{\mathcal{U}} = \Big\{ P \mid P = P_1 \times P_{2|1} \times \cdots \times P_{T|T-1},\ P_1 \in \widetilde{\mathcal{U}}_1,\ P_{t|t-1} \in \widetilde{\mathcal{U}}_t(\mathbf{d}_{t-1})\ \forall \mathbf{d}_{t-1} \in \Xi_{t-1}\ \forall t = 2, \ldots, T \Big\}.$$

In other words, $\widetilde{\mathcal{U}}$ is the set of all distributions P with the marginals lying in the specified uncertainty sets. Observe that for any distribution P, which has corresponding marginal distribution P_1, and for each $\mathbf{d}_{t-1}$ with the conditional distribution $P_{t|t-1}$, it holds that

$$\mathbb{E}_{P_1}\left[h_1(\mathbf{x}_1, \mathbf{d}_1) + \left\{ \mathbb{E}_{P_{2|1}}\left[h_2(\mathbf{x}_2, \mathbf{d}_2) + \cdots + \left\{ \mathbb{E}_{P_{T|T-1}}\left[h_T(\mathbf{x}_T, \mathbf{d}_T) \right] \right\} \right] \right\} \right] = \mathbb{E}_P\left[\sum_{t=1}^{T} h_t(\mathbf{x}_t, \mathbf{d}_t) \right]. \tag{5}$$

With this, the robust counterpart of constraint (C-DRO) can be written as

$$\sup_{P \in \widetilde{\mathcal{U}}} \mathbb{E}_P \left[\sum_{t=1}^{T} h_t(\mathbf{x}_t, \mathbf{d}_t) \right] \leq B. \tag{6}$$

For a small $\epsilon > 0$, let $P^* \in \widetilde{\mathcal{U}}$ be such that

$$\mathbb{E}_{P^*} \left[\sum_{t=1}^{T} h_t(\mathbf{x}_t, \mathbf{d}_t) \right] \geq \sup_{P \in \widetilde{\mathcal{U}}} \mathbb{E}_P \left[\sum_{t=1}^{T} h_t(\mathbf{x}_t, \mathbf{d}_t) \right] - \epsilon. \tag{7}$$

That means the LHS of (7) is ϵ-optimal to the LHS of (6). Since $P^* \in \widetilde{\mathcal{U}}$, the marginal distribution of $\mathbf{d}_1$, P_1^* lies in $\widetilde{\mathcal{U}}_1$ and the conditional distribution of $\mathbf{d}_t$, $P^*_{t|t-1}$ lies in $\widetilde{\mathcal{U}}(\mathbf{d}_{t-1})$, where $\mathbf{d}_{t-1} \in \Xi_{t-1}$. This holds true for all $t = 2, \ldots, T$. Using (5), this means

$$\begin{aligned}
&\sup_{P_1 \in \widetilde{\mathcal{U}}_1} \mathbb{E}_{P_1} \left[h_1(\mathbf{x}_1, \mathbf{d}_1) + \sup_{P_{2|1} \in \widetilde{\mathcal{U}}_2(\mathbf{d}_1)} \left\{ \mathbb{E}_{P_{2|1}} \left[h_2(\mathbf{x}_2, \mathbf{d}_2) + \cdots + \sup_{P_{T|T-1} \in \widetilde{\mathcal{U}}_T(\mathbf{d}_{T-1})} \left\{ \mathbb{E}_{P_{T|T-1}} \left[h_T(\mathbf{x}_T, \mathbf{d}_T) \right] \right\} \right] \right\} \right] \\
&\geq \mathbb{E}_{P_1^*} \left[h_1(\mathbf{x}_1, \mathbf{d}_1) + \left\{ \mathbb{E}_{P^*_{2|1}} \left[h_2(\mathbf{x}_2, \mathbf{d}_2) + \cdots + \left\{ \mathbb{E}_{P^*_{T|T-1}} \left[h_T(\mathbf{x}_T, \mathbf{d}_T) \right] \right\} \right] \right\} \right] = \mathbb{E}_{P^*}[\sum_{t=1}^{T} h_t(\mathbf{x}_t, \mathbf{d}_t)] \\
&\geq \sup_{P \in \widetilde{\mathcal{U}}} \mathbb{E}_P[\sum_{t=1}^{T} h_t(\mathbf{x}_t, \mathbf{d}_t)] - \epsilon.
\end{aligned}$$

Now for the opposite side of the inequality, let $P_1^*, P_{2|1}^*$ up to $P^*_{T|T-1}$ be ϵ-optimal to (4). The constraints of the problem ensure that $P_1^* \in \widetilde{\mathcal{U}}_1$ and $P^*_{t|t-1} \in \widetilde{\mathcal{U}}_t(\mathbf{d}_{t-1})$. However the uncertainty set $\widetilde{\mathcal{U}}$ is the set of all joint distributions with these specified marginals. Therefore, for the joint distribution P^* with these marginals, the equation (5) will hold. However since P_1^* and $P^*_{t|t-1}$ are ϵ-optimal, this means that

$$\begin{aligned}
&\sup_{P_1 \in \widetilde{\mathcal{U}}_1} \mathbb{E}_{P_1} \left[h_1(\mathbf{x}_1, \mathbf{d}_1) + \sup_{P_{2|1} \in \widetilde{\mathcal{U}}_2(\mathbf{d}_1)} \left\{ \mathbb{E}_{P_{2|1}} \left[h_2(\mathbf{x}_2, \mathbf{d}_2) + \cdots + \sup_{P_{T|T-1} \in \widetilde{\mathcal{U}}_T(\mathbf{d}_{T-1})} \left\{ \mathbb{E}_{P_{T|T-1}} \left[h_T(\mathbf{x}_T, \mathbf{d}_T) \right] \right\} \right] \right\} \right] \\
&\leq \mathbb{E}_{P_1^*} \left[h_1(\mathbf{x}_1, \mathbf{d}_1) + \left\{ \mathbb{E}_{P^*_{2|1}} \left[h_2(\mathbf{x}_2, \mathbf{d}_2) + \cdots + \left\{ \mathbb{E}_{P^*_{T|T-1}} \left[h_T(\mathbf{x}_T, \mathbf{d}_T) \right] \right\} \right] \right\} \right] + \epsilon = \mathbb{E}_{P^*} \left[\sum_{t=1}^{T} h_t(\mathbf{x}_t, \mathbf{d}_t) \right] + \epsilon \\
&\leq \sup_{P \in \widetilde{\mathcal{U}}} \mathbb{E}_P[\sum_{t=1}^{T} h_t(\mathbf{x}_t, \mathbf{d}_t)] + \epsilon.
\end{aligned}$$

This gives the opposite inequality, and the result follows by letting ϵ shrink towards zero. □

Another way to prove Proposition 1 is to leverage the rectangularity of set $\widetilde{\mathcal{U}}$, similar to the construction in [36]. We introduce the following function Q and S_t to more compactly express the constraints:

$$\begin{aligned}
Q(p_k, \overline{\mathbf{q}}_k, \underline{\mathbf{q}}_k, \mathbf{R}_k; \mathbf{d}_k, \mathbf{d}_{k-1}) &= p_k(\mathbf{d}_{k-1}) + (\overline{\mathbf{q}}_k(\mathbf{d}_{k-1})^\top \mathbf{d}_k - \underline{\mathbf{q}}_k(\mathbf{d}_{k-1}))^\top \mathbf{d}_k + \mathbf{d}_k^\top \mathbf{R}_k(\mathbf{d}_{k-1}) \mathbf{d}_k, \\
S_{t+1}(p_{k+1}, \overline{\mathbf{q}}_{k+1}, \underline{\mathbf{q}}_{k+1}, \mathbf{R}_{k+1}; \mathbf{d}_k) &= p_{k+1}(\mathbf{d}_k) + \overline{\mathbf{q}}_{k+1}(\mathbf{d}_k)^\top \overline{\boldsymbol{\mu}}_{t+1}(\mathbf{d}_t) \\
&\quad - \underline{\mathbf{q}}_{k+1}(\mathbf{d}_k)^\top \underline{\boldsymbol{\mu}}_{t+1}(\mathbf{d}_t) + \mathbf{R}_{k+1}(\mathbf{d}_k) \cdot \boldsymbol{\Sigma}'_{t+1}(\mathbf{d}_t).
\end{aligned}$$

Note that Q is fully defined by its arguments, whereas S_{t+1} depends on the parameter t via $\overline{\boldsymbol{\mu}}_{t+1}$, $\underline{\boldsymbol{\mu}}_{t+1}$, and $\boldsymbol{\Sigma}'_{t+1}$. Furthermore, $p_k(\mathbf{d}_{k-1})$ and other similar terms denote variables which are functions of the realization $\mathbf{d}_{k-1}$. For example, if $\mathbf{d}_{k-1} = \mathbf{d}_0$ in function $Q(p_k, \overline{\mathbf{q}}_k, \underline{\mathbf{q}}_k, \mathbf{R}_k; \mathbf{d}_k, \mathbf{d}_{k-1})$, and if $\mathbf{d}_k = \mathbf{d}_0$ in the function $S_{t+1}(p_{k+1}, \overline{\mathbf{q}}_{k+1}, \underline{\mathbf{q}}_{k+1}, \mathbf{R}_{k+1}; \mathbf{d}_k)$, then the variable arguments in Q and S_{t+1} are constants, i.e., independent of the input $\mathbf{d}_{k-1}$ and $\mathbf{d}_k$, respectively. This is because $\mathbf{d}_0$ is assumed to be known. To obtain the complete reformulation, the following theorem casts (4) as an infinite dimensional problem.

Theorem 1 *If the set $\widetilde{\mathcal{U}}_1$ and, for each $\mathbf{d}_{t-1}$, the sets $\widetilde{\mathcal{U}}_t(\mathbf{d}_{t-1})$ have a nonempty relative interior, using*

$$g_{t+1}(\mathbf{x}_{[t+1,T]}, \mathbf{d}_t) = \sup_{P_{t+1|t} \in \widetilde{\mathcal{U}}_{t+1}(\mathbf{d}_t)} \mathbb{E}_{P_{t+1|t}} \left[h_{t+1}(\mathbf{x}_{t+1}, \mathbf{d}_{t+1}) + \cdots + \sup_{P_{T|T-1} \in \widetilde{\mathcal{U}}_T(\mathbf{d}_{T-1})} \left\{ \mathbb{E}_{P_{T|T-1}} [h_T(\mathbf{x}_T, \mathbf{d}_T)] \right\} \right],$$

constraint (4) can be reformulated for any $t = 2, \ldots, T$ as the following collection of constraints:

$$
\begin{aligned}
S_1(p_1, \overline{\mathbf{q}}_1, \underline{\mathbf{q}}_1, \mathbf{R}_1; \mathbf{d}_0) &\leq B \\
Q(p_1, \overline{\mathbf{q}}_1, \underline{\mathbf{q}}_1, \mathbf{R}_1; \mathbf{d}_1, \mathbf{d}_0) &\geq h_1(\mathbf{x}_1, \mathbf{d}_1) + S_2(p_2, \overline{\mathbf{q}}_2, \underline{\mathbf{q}}_2, \mathbf{R}_2; \mathbf{d}_1) && \forall \mathbf{d}_1 \in \Xi_1 \\
Q(p_k, \overline{\mathbf{q}}_k, \underline{\mathbf{q}}_k, \mathbf{R}_k; \mathbf{d}_k, \mathbf{d}_{k-1}) &\geq h_k(\mathbf{x}_k, \mathbf{d}_k) + S_{k+1}(p_{k+1}, \overline{\mathbf{q}}_{k+1}, \underline{\mathbf{q}}_{k+1}, \mathbf{R}_{k+1}; \mathbf{d}_k) && \forall \mathbf{d}_{k-1}, \mathbf{d}_k \in \Xi_{k-1} \times \Xi_k \\
Q(p_t, \overline{\mathbf{q}}_t, \underline{\mathbf{q}}_t, \mathbf{R}_t; \mathbf{d}_t, \mathbf{d}_{t-1}) &\geq h_t(\mathbf{x}_t, \mathbf{d}_t) + g_{t+1}(\mathbf{x}_{[t+1,T]}, \mathbf{d}_t) && \forall \mathbf{d}_{t-1}, \mathbf{d}_t \in \Xi_{t-1} \times \Xi_t \quad (8) \\
\underline{\mathbf{q}}_k(\mathbf{d}_{k-1}), \overline{\mathbf{q}}_k(\mathbf{d}_{k-1}) &\geq 0 && \forall \mathbf{d}_{k-1} \in \Xi_{k-1} \\
\mathbf{R}_1 \succeq \mathbf{0},\ \mathbf{R}_k(\mathbf{d}_{k-1}) &\succeq \mathbf{0} && \forall \mathbf{d}_{k-1} \in \Xi_{k-1} \\
\underline{\mathbf{q}}_1, \overline{\mathbf{q}}_1 &\geq 0,
\end{aligned}
$$

where $k = 2, \ldots, t-1$ and variables $p_k(\mathbf{d}_{k-1}), \overline{\mathbf{q}}_k(\mathbf{d}_{k-1}), \underline{\mathbf{q}}_k(\mathbf{d}_{k-1})$ and $\mathbf{R}_k(\mathbf{d}_{k-1})$ are functions of $\mathbf{d}_{k-1}$ (same holds for variables with index t).

Proof The proof proceeds by induction.

Base Case ($t = 1$): Let $g_2(\mathbf{x}_{[2,T]}, \mathbf{d}_1) = \sup_{P_{2|1} \in \widetilde{\mathcal{U}}_2(\mathbf{d}_1)} \mathbb{E}_{P_{2|1}} \left[h_2(\mathbf{x}_2, \mathbf{d}_2) + \cdots + \sup_{P_{T|T-1} \in \widetilde{\mathcal{U}}_T(\mathbf{d}_{T-1})} \left\{ \mathbb{E}_{P_{T|T-1}}[h_T(\mathbf{x}_T, \mathbf{d}_T)] \right\} \right]$. With this, constraint (4) can be written as

$$
\sup_{P_1 \in \widetilde{\mathcal{U}}_1} \mathbb{E}_{P_1} \left[h_1(\mathbf{x}_1, \mathbf{d}_1) + g_2(\mathbf{x}_{[2,T]}, \mathbf{d}_1) \right] \leq B,
$$

where the uncertainty set is given by $\widetilde{\mathcal{U}}_1$. Since this set has a nonempty relative interior, strong duality holds for this optimization problem [34] . This yields the following semi-infinite constraint

$$
\begin{aligned}
& p_1 + \overline{\mathbf{q}}_1^\top \overline{\boldsymbol{\mu}}_1 - \underline{\mathbf{q}}_1^\top \underline{\boldsymbol{\mu}}_1 + \mathbf{R}_1 \cdot \boldsymbol{\Sigma}_1' \leq B \\
& p_1 + \overline{\mathbf{q}}_1^\top \mathbf{d}_1 - \underline{\mathbf{q}}_1^\top \mathbf{d}_1 + \mathbf{d}_1^\top \mathbf{R}_1 \mathbf{d}_1 \geq h_1(\mathbf{x}_1, \mathbf{d}_1) + g_2(\mathbf{x}_{[2,T]}, \mathbf{d}_1) \quad \forall \mathbf{d}_1 \in \Xi_1 \\
& \overline{\mathbf{q}}_1, \underline{\mathbf{q}}_1 \geq 0 \\
& \mathbf{R}_1 \succeq \mathbf{0},
\end{aligned}
$$

which is the desired result. This can be compactly expressed as

$$
\begin{aligned}
& S(p_1, \overline{\mathbf{q}}_1, \underline{\mathbf{q}}_1, \mathbf{R}_1; \mathbf{d}_0) \leq B \\
& Q_1(p_1, \overline{\mathbf{q}}_1, \underline{\mathbf{q}}_1, \mathbf{R}_1; \mathbf{d}_1, \mathbf{d}_0) \geq h_1(\mathbf{x}_1, \mathbf{d}_1) + g_2(\mathbf{x}_{[2,T]}, \mathbf{d}_1) \quad \forall \mathbf{d}_1 \in \Xi_1 \\
& \overline{\mathbf{q}}_1, \underline{\mathbf{q}}_1 \geq 0 \\
& \mathbf{R}_1 \succeq \mathbf{0},
\end{aligned}
$$

Nominal case: Assume that the constraints (8) are the reformulation of (4) with respect to the uncertain distributions $P_1, \ldots, P_{t|t-1}$ lying in sets $\widetilde{\mathcal{U}}_1, \ldots, \widetilde{\mathcal{U}}_t(\mathbf{d}_{t-1})$, respectively. Using the definition of $g_{t+2}(\mathbf{x}_{[t+2,T]}, \mathbf{d}_{t+1})$, we can write

$$
g_{t+1}(\mathbf{x}_{[t+1,T]}, \mathbf{d}_t) = \sup_{P_{t+1|t} \in \widetilde{\mathcal{U}}_{t+1}(\mathbf{d}_t)} \mathbb{E}_{P_{t+1|t}}[h_{t+1}(\mathbf{x}_{t+1}, \mathbf{d}_{t+1}) + g_{t+2}(\mathbf{x}_{[t+2,T]}, \mathbf{d}_{t+1})], \tag{9}
$$

where the set $\widetilde{\mathcal{U}}_{t+1}(\mathbf{d}_t)$ is given by

$$
\begin{aligned}
\widetilde{\mathcal{U}}_{t+1}(\mathbf{d}_t) = \Big\{ P_{t+1|t} \in \mathcal{M} \Big| & P_{t+1|t}(\mathbf{d}_{t+1} \in \Xi_{t+1}) = 1,\ \ \underline{\boldsymbol{\mu}}_{t+1}(\mathbf{d}_t) \leq \mathbb{E}_{P_{t+1|t}}[\mathbf{d}_{t+1}] \leq \overline{\boldsymbol{\mu}}_{t+1}(\mathbf{d}_t), \\
& \mathbb{E}_{P_{t+1|t}}[(\mathbf{d}_{t+1} - \boldsymbol{\mu}_{t+1}(\mathbf{d}_t))(\mathbf{d}_{t+1} - \boldsymbol{\mu}_{t+1}(\mathbf{d}_t))^\top] \preceq \boldsymbol{\Sigma}_{t+1}(\mathbf{d}_t) \Big\}.
\end{aligned}
$$

Since this set has a nonempty relative interior, by strong duality, the dual of (9) obtains the same value as $g_{t+1}(\mathbf{x}_{[t+1,T]}, \mathbf{d}_t)$ and can be expressed as

$$\begin{aligned}
\inf_{\substack{p_{t+1},\overline{\mathbf{q}}_{t+1}\\ \underline{\mathbf{q}}_{t+1},\mathbf{R}_{t+1}}} \quad & p_{t+1} + \overline{\mathbf{q}}_{t+1}^\top \overline{\boldsymbol{\mu}}_{t+1}(\mathbf{d}_t) - \underline{\mathbf{q}}_{t+1}^\top \underline{\boldsymbol{\mu}}_{t+1}(\mathbf{d}_t) + \mathbf{R}_{t+1} \cdot \boldsymbol{\Sigma}'_{t+1}(\mathbf{d}_t) \\
\text{s.t.} \quad & p_{t+1} + (\overline{\mathbf{q}}_{t+1} - \underline{\mathbf{q}}_{t+1})^\top \mathbf{d}_{t+1} + \mathbf{d}_{t+1}^\top \mathbf{R}_{t+1}\mathbf{d}_{t+1} \geq h_{t+1}(\mathbf{x}_{t+1}, \mathbf{d}_{t+1}) \\
& \qquad\qquad + g_{t+2}(\mathbf{x}_{[t+2,T]}, \mathbf{d}_{t+1}) \;\; \forall \mathbf{d}_{t+1} \in \Xi_{t+1} \\
& \overline{\mathbf{q}}_{t+1}, \underline{\mathbf{q}}_{t+1} \geq 0 \\
& \mathbf{R}_{t+1} \succeq \mathbf{0}.
\end{aligned} \tag{10}$$

The optimal solution for the above problem is a function of the uncertain component $\mathbf{d}_t$. Observe the last constraint in (8):

$$Q(p_t, \overline{\mathbf{q}}_t, \underline{\mathbf{q}}_t, \mathbf{R}_t; \mathbf{d}_t, \mathbf{d}_{t-1}) \geq h_t(\mathbf{x}_t, \mathbf{d}_t) + g_{t+1}(\mathbf{x}_{[t+1,T]}, \mathbf{d}_t) \;\; \forall \mathbf{d}_{t-1}, \mathbf{d}_t \in \Xi_{t-1} \times \Xi_t. \tag{11}$$

By (10), there exists a set of variables $p_{t+1}(\mathbf{d}_t)$, $\overline{\mathbf{q}}_{t+1}(\mathbf{d}_t) \geq \mathbf{0}$, $\underline{\mathbf{q}}_{t+1}(\mathbf{d}_t) \geq \mathbf{0}$ and $\mathbf{R}_{t+1}(\mathbf{d}_t) \succeq \mathbf{0}$ which are functions of $\mathbf{d}_t$ such that,

$$\begin{aligned}
Q(p_t, \overline{\mathbf{q}}_t, \underline{\mathbf{q}}_t, \mathbf{R}_t; \mathbf{d}_t, \mathbf{d}_{t-1}) \geq h_t(\mathbf{x}_t, \mathbf{d}_t) + p_{t+1}(\mathbf{d}_t) + \overline{\mathbf{q}}_{t+1}(\mathbf{d}_t)^\top \overline{\boldsymbol{\mu}}_{t+1}(\mathbf{d}_t) - \underline{\mathbf{q}}_{t+1}(\mathbf{d}_t)^\top \underline{\boldsymbol{\mu}}_{t+1}(\mathbf{d}_t) + \\
\mathbf{R}_{t+1}(\mathbf{d}_t) \cdot \boldsymbol{\Sigma}'_{t+1}(\mathbf{d}_t).
\end{aligned}$$

This can be compactly expressed as

$$Q(p_t, \overline{\mathbf{q}}_t, \underline{\mathbf{q}}_t, \mathbf{R}_t; \mathbf{d}_t, \mathbf{d}_{t-1}) \geq h_t(\mathbf{x}_t, \mathbf{d}_t) + S_{t+1}(p_{t+1}, \overline{\mathbf{q}}_{t+1}, \underline{\mathbf{q}}_{t+1}, R_{t+1}; \mathbf{d}_t),$$

where the variables $p_{t+1}(\mathbf{d}_t), \overline{\mathbf{q}}_{t+1}(\mathbf{d}_t), \underline{\mathbf{q}}_{t+1}(\mathbf{d}_t)$ and $\mathbf{R}_{t+1}(\mathbf{d}_t)$ satisfy the constraints of (10), which means $Q(p_{t+1}, \overline{\mathbf{q}}_{t+1}, \underline{\mathbf{q}}_{t+1}, \mathbf{R}_{t+1}; \mathbf{d}_{t+1}, \mathbf{d}_t) \geq h_{t+1}(\mathbf{x}_{t+1}, \mathbf{d}_{t+1}) + g_{t+2}(\mathbf{x}_{[t+2,T]}, \mathbf{d}_{t+1}) \;\; \forall \mathbf{d}_{t+1} \in \Xi_{t+1}$. This allows us to write constraint (4) as

$$\begin{aligned}
S_1(p_1, \overline{\mathbf{q}}_1, \underline{\mathbf{q}}_1, \mathbf{R}_1; \mathbf{d}_0) &\leq B \\
Q(p_1, \overline{\mathbf{q}}_1, \underline{\mathbf{q}}_1, \mathbf{R}_1; \mathbf{d}_1, \mathbf{d}_0) &\geq h_1(\mathbf{x}_1, \mathbf{d}_1) + S_2(p_2, \overline{\mathbf{q}}_2, \underline{\mathbf{q}}_2, \mathbf{R}_2; \mathbf{d}_1) && \forall \mathbf{d}_1 \in \Xi_1 \\
Q(p_k, \overline{\mathbf{q}}_k, \underline{\mathbf{q}}_k, \mathbf{R}_k; \mathbf{d}_k, \mathbf{d}_{k-1}) &\geq h_k(\mathbf{x}_k, \mathbf{d}_k) + S_{k+1}(p_{k+1}, \overline{\mathbf{q}}_{k+1}, \underline{\mathbf{q}}_{k+1}, \mathbf{R}_{k+1}; \mathbf{d}_k) && \forall \mathbf{d}_{k-1}, \mathbf{d}_k \in \Xi_{k-1} \\
Q(p_t, \overline{\mathbf{q}}_t, \underline{\mathbf{q}}_t, \mathbf{R}_t; \mathbf{d}_t, \mathbf{d}_{t-1}) &\geq h_t(\mathbf{x}_t, \mathbf{d}_t) + S_{t+1}(p_{t+1}, \overline{\mathbf{q}}_{t+1}, \underline{\mathbf{q}}_{t+1}, R_{t+1}; \mathbf{d}_t) \\
Q(p_{t+1}, \overline{\mathbf{q}}_{t+1}, \underline{\mathbf{q}}_{t+1}, \mathbf{R}_{t+1}; \mathbf{d}_{t+1}, \mathbf{d}_t) &\geq h_{t+1}(\mathbf{x}_{t+1}, \mathbf{d}_{t+1}) + g_{t+2}(\mathbf{x}_{[t+2,T]}, \mathbf{d}_{t+1}) && \forall \mathbf{d}_t \in \Xi_t \;\; \forall \mathbf{d}_{t+1} \in \Xi_{t+1} \\
\underline{\mathbf{q}}_k(\mathbf{d}_{k-1}), \overline{\mathbf{q}}_k(\mathbf{d}_{k-1}) &\geq 0 && \forall \mathbf{d}_{k-1} \in \Xi_{k-1} \\
\mathbf{R}_1 \succeq \mathbf{0}, \;\; \mathbf{R}_k(\mathbf{d}_{k-1}) &\succeq \mathbf{0} && \forall \mathbf{d}_{k-1} \in \Xi_{k-1} \\
\underline{\mathbf{q}}_1, \overline{\mathbf{q}}_1 &\geq 0,
\end{aligned}$$

for $k = 2, \ldots, t-1$. This completes the proof by induction. □

Note that a complete reformulation of the constraint (4) can be obtained by applying Theorem 1 to the case $t = T$ because $g_T(\mathbf{x}_T, \mathbf{d}_{T-1}) = \sup_{P_{T|T-1} \in \tilde{\mathcal{U}}_T(\mathbf{d}_{T-1})} \mathbb{E}_{P_{T|T-1}} h_T(\mathbf{x}_T, \mathbf{d}_T)$.

In the reformulation (8), the variables of the form $p_t(\mathbf{d}_{t-1}), \overline{\mathbf{q}}_t(\mathbf{d}_{t-1}), \underline{\mathbf{q}}_t(\mathbf{d}_{t-1})$ and $\mathbf{R}_t(\mathbf{d}_{t-1})$ are functions of the uncertainty realization in the sense that for each realization of $\mathbf{d}_{t-1}$, there exist separate variables $p_t(\mathbf{d}_{t-1}), \overline{\mathbf{q}}_t(\mathbf{d}_{t-1}), \underline{\mathbf{q}}_t(\mathbf{d}_{t-1})$, and $\mathbf{R}_t(\mathbf{d}_{t-1})$ in the problem. This leads to an infinite dimensional optimization problem. It also presents a path to reduce the complexity of the problem by limiting the adaptability of the variables.

2.2 Approximation

To reduce the complexity of the reformulation from Theorem 1, we replace all the uncertainty dependent variables $p_t(\mathbf{d}_{t-1}), \overline{\mathbf{q}}_t(\mathbf{d}_{t-1}), \underline{\mathbf{q}}_t(\mathbf{d}_{t-1}), \mathbf{R}_t(\mathbf{d}_{t-1})$ by decision rules, which do not adapt to the uncertainty realization. This allows for a conservative approximation to constraint (4). The following proposition investigates this further.

Proposition 2 *Let $\hat{p}_t$, $\hat{\overline{\mathbf{q}}}_t, \hat{\underline{\mathbf{q}}}_t$, and $\hat{\mathbf{R}}_t$ be static decision rules for $t = 1, \ldots, T$ and let the set $\widetilde{\mathcal{U}}_1$ and, for each $\mathbf{d}_{t-1}$, the sets $\widetilde{\mathcal{U}}_t(\mathbf{d}_{t-1})$ have a nonempty relative interior. Then $\mathbf{x}_1, \ldots, \mathbf{x}_T$, which satisfy*

$$\begin{aligned}
S_1(\hat{p}_1, \hat{\overline{\mathbf{q}}}_1, \hat{\underline{\mathbf{q}}}_1, \hat{\mathbf{R}}_1; \mathbf{d}_0) &\leq B \\
Q(\hat{p}_1, \hat{\overline{\mathbf{q}}}_1, \hat{\underline{\mathbf{q}}}_1, \hat{\mathbf{R}}_1; \mathbf{d}_1, \mathbf{d}_0) &\geq h_1(\mathbf{x}_1, \mathbf{d}_1) + S_2(\hat{p}_2, \hat{\overline{\mathbf{q}}}_2, \hat{\underline{\mathbf{q}}}_2, \hat{\mathbf{R}}_2; \mathbf{d}_0) && \forall \mathbf{d}_1 \in \Xi_1 \\
Q(\hat{p}_k, \hat{\overline{\mathbf{q}}}_k, \hat{\underline{\mathbf{q}}}_k, \hat{\mathbf{R}}_k; \mathbf{d}_k, \mathbf{d}_0) &\geq h_k(\mathbf{x}_k, \mathbf{d}_k) + S_{k+1}(\hat{p}_{k+1}, \hat{\overline{\mathbf{q}}}_{k+1}, \hat{\underline{\mathbf{q}}}_{k+1}, \hat{\mathbf{R}}_{k+1}; \mathbf{d}_0) && \forall \mathbf{d}_k \in \Xi_k, k = 2, .., t-1 \\
Q(\hat{p}_t, \hat{\overline{\mathbf{q}}}_t, \hat{\underline{\mathbf{q}}}_t, \hat{\mathbf{R}}_t; \mathbf{d}_t, \mathbf{d}_0) &\geq h_t(\mathbf{x}_t, \mathbf{d}_t) + g_{t+1}(\mathbf{x}_{[t+1,T]}, \mathbf{d}_t) && \forall \mathbf{d}_{t-1}, \mathbf{d}_t \in \Xi_{t-1} \times \Xi_t \\
\hat{\underline{\mathbf{q}}}_k, \hat{\overline{\mathbf{q}}}_k &\geq 0 \\
\hat{\mathbf{R}}_1 \succeq \mathbf{0},\ \hat{\mathbf{R}}_k &\succeq \mathbf{0} \\
\hat{\underline{\mathbf{q}}}_1, \hat{\overline{\mathbf{q}}}_1 &\geq 0,
\end{aligned} \tag{12}$$

for any $t = 2, \ldots, T$, also satisfy the constraint (C-DRO) *for any $P \in \widetilde{\mathcal{U}}$. Here $g_{t+1}(\mathbf{x}_{[t+1,T]}, \mathbf{d}_t)$ is defined in Theorem 1.*

Proof We proceed by induction. At each step, we prove that constraints (12) are a conservative approximation to the true constraints (8).

Base Case ($t = 1$): From the base case of Theorem 1, the reformulation of the constraint (4) for $t = 1$ is

$$\begin{aligned}
&\hat{p}_1 + \hat{\overline{\mathbf{q}}}_1^\top \overline{\boldsymbol{\mu}}_1 - \hat{\underline{\mathbf{q}}}_1^\top \underline{\boldsymbol{\mu}}_1 + \hat{\mathbf{R}}_1 \cdot \boldsymbol{\Sigma}_1' \leq B \\
&\hat{p}_k + (\hat{\overline{\mathbf{q}}}_k - \hat{\underline{\mathbf{q}}}_k)^\top \mathbf{d}_k + \mathbf{d}_k^\top \hat{\mathbf{R}}_k \mathbf{d}_k \geq h_1(\mathbf{x}_1, \mathbf{d}_1) + g_2(\mathbf{x}_{[2,T]}, \mathbf{d}_1) \quad \forall \mathbf{d}_1 \in \Xi_1 \\
&\hat{\overline{\mathbf{q}}}_1, \hat{\underline{\mathbf{q}}}_1 \geq \mathbf{0} \\
&\hat{\mathbf{R}}_1 \succeq \mathbf{0}.
\end{aligned}$$

This is the same as the set of required constraints for Proposition 2. Any $\mathbf{x}_1, \ldots, \mathbf{x}_T$, which satisfy these constraints, will also satisfy (C-DRO) . Note that $\hat{p}_1, \hat{\overline{\mathbf{q}}}_1, \hat{\underline{\mathbf{q}}}_1$ and $\hat{\mathbf{R}}_1$ are fixed variables.

Nominal case: Assume that any $\mathbf{x}_1, \ldots, \mathbf{x}_T$, which satisfy (12), also satisfy (C-DRO). Equation (10) (from the nominal case of Theorem 1) provides $g_{t+1}(\mathbf{x}_{[t+1,T]}, \mathbf{d}_t)$ as

$$\begin{aligned}
\inf_{\substack{p_{t+1}, \overline{\mathbf{q}}_{t+1} \\ \underline{\mathbf{q}}_{t+1}, \mathbf{R}_{t+1}}} \quad & p_{t+1} + \overline{\mathbf{q}}_{t+1}^\top \overline{\boldsymbol{\mu}}_{t+1}(\mathbf{d}_t) - \underline{\mathbf{q}}_{t+1}^\top \underline{\boldsymbol{\mu}}_{t+1}(\mathbf{d}_t) + \mathbf{R}_{t+1} \cdot \boldsymbol{\Sigma}_{t+1}'(\mathbf{d}_t) \\
\text{s.t.} \quad & p_{t+1} + (\overline{\mathbf{q}}_{t+1} - \underline{\mathbf{q}}_{t+1})^\top \mathbf{d}_{t+1} + \mathbf{d}_{t+1}^\top \mathbf{R}_{t+1} \mathbf{d}_{t+1} \geq h_{t+1}(\mathbf{x}_{t+1}, \mathbf{d}_{t+1}) \\
& \qquad + g_{t+2}(\mathbf{x}_{[t+2,T]}, \mathbf{d}_{t+1}) \ \ \forall \mathbf{d}_{t+1} \in \Xi_{t+1} \\
& \overline{\mathbf{q}}_{t+1}, \underline{\mathbf{q}}_{t+1} \geq 0 \\
& \mathbf{R}_{t+1} \succeq \mathbf{0}.
\end{aligned}$$

The optimal solution is a function of $\mathbf{d}_t$. If we restrict the problem to functions that are independent of $\mathbf{d}_t$, then the solution is suboptimal. Hence, for any $\hat{p}_{t+1}, \hat{\overline{\mathbf{q}}}_{t+1} \geq \mathbf{0}, \hat{\underline{\mathbf{q}}}_{t+1} \geq \mathbf{0}$, and $\hat{\mathbf{R}}_{t+1} \succeq \mathbf{0}$ which satisfy

$$\hat{p}_{t+1} + (\hat{\overline{\mathbf{q}}}_{t+1} - \hat{\underline{\mathbf{q}}}_{t+1})^\top \mathbf{d}_{t+1} + \mathbf{d}_{t+1}^\top \hat{\mathbf{R}}_{t+1} \mathbf{d}_{t+1} \geq h_{t+1}(\mathbf{x}_{t+1}, \mathbf{d}_{t+1}) + g_{t+2}(\mathbf{x}_{[t+2,T]}, \mathbf{d}_{t+1}) \ \ \forall \mathbf{d}_{t+1} \in \Xi_{t+1}, \tag{13}$$

we have the following inequality in the objective function values

$$\begin{aligned}
\hat{p}_{t+1} + \hat{\overline{\mathbf{q}}}_{t+1}^\top \overline{\boldsymbol{\mu}}_{t+1}(\mathbf{d}_t) - \hat{\underline{\mathbf{q}}}_{t+1}^\top \underline{\boldsymbol{\mu}}_{t+1}(\mathbf{d}_t) + \hat{\mathbf{R}}_{t+1} \cdot \boldsymbol{\Sigma}_{t+1}'(\mathbf{d}_t) &\geq p_{t+1} + \overline{\mathbf{q}}_{t+1}^\top \overline{\boldsymbol{\mu}}_{t+1}(\mathbf{d}_t) - \underline{\mathbf{q}}_{t+1}^\top \underline{\boldsymbol{\mu}}_{t+1}(\mathbf{d}_t) + \\
&\quad \mathbf{R}_{t+1} \cdot \boldsymbol{\Sigma}_{t+1}'(\mathbf{d}_t) \\
&= g_{t+1}(\mathbf{x}_{[t+1,T]}, \mathbf{d}_t).
\end{aligned} \tag{14}$$

The term $g_{t+1}(\mathbf{x}_{[t+1,T]}, \mathbf{d}_t)$ in the last constraint of (12) is bounded above by the LHS of the inequality (14). This means that any $\mathbf{x}_t$ which satisfy

$$\hat{p}_t + (\hat{\overline{\mathbf{q}}}_t - \hat{\underline{\mathbf{q}}}_t)^\top \mathbf{d}_t + \mathbf{d}_t^\top \hat{\mathbf{R}}_t \mathbf{d}_t \geq h_t(\mathbf{x}_t, \mathbf{d}_t) + \hat{p}_{t+1} + \hat{\overline{\mathbf{q}}}_{t+1}^\top \overline{\boldsymbol{\mu}}_{t+1}(\mathbf{d}_t) - \hat{\underline{\mathbf{q}}}_{t+1}^\top \underline{\boldsymbol{\mu}}_{t+1}(\mathbf{d}_t) + \hat{\mathbf{R}}_{t+1} \cdot \boldsymbol{\Sigma}_{t+1}'(\mathbf{d}_t), \tag{15}$$

and any $\mathbf{x}_{[t+1,T]}$, which satisfy (13), will also satisfy the last constraint in (12). This allows to replace constraint (12) for any $\mathbf{x}_{[t+1,T]}$ with constraints (13) and (15), which completes the induction. □

Proposition 2 can be interpreted as following. At any k, the RHS of the constraint is translated from h_k by S_{k+1}, which approximates the impact of the future period problems. This is not the case at T. On the hand, in the infinite dimensional problem in Theorem 1, this translation is equal to the exact value of the objective of the future periods for a fixed d_k. If a standard two-period DRO problem without connected sets was reformulated following the above steps, the translation S_{k+1} would be independent of d_k. The complete conservative reformulation of the constraint (C-DRO) is given by the following proposition.

Proposition 3 *Any* $\mathbf{x}_1, \ldots, \mathbf{x}_T$*, which satisfy*

$$\begin{aligned}
S_1(\hat{p}_1, \hat{\bar{\mathbf{q}}}_1, \hat{\underline{\mathbf{q}}}_1, \hat{\mathbf{R}}_1; \mathbf{d}_0) &\le B \\
Q(\hat{p}_k, \hat{\bar{\mathbf{q}}}_k, \hat{\underline{\mathbf{q}}}_k, \hat{\mathbf{R}}_k; \mathbf{d}_k, \mathbf{d}_0) &\ge h_k(\mathbf{x}_k, \mathbf{d}_k) + S_{k+1}(\hat{p}_{k+1}, \hat{\bar{\mathbf{q}}}_{k+1}, \hat{\underline{\mathbf{q}}}_{k+1}, \hat{\mathbf{R}}_{k+1}; \mathbf{d}_0) \ \ \forall \mathbf{d}_k \in \Xi_k, k = 2, .., T-1 \\
Q(\hat{p}_T, \hat{\bar{\mathbf{q}}}_T, \hat{\underline{\mathbf{q}}}_T, \hat{\mathbf{R}}_T; \mathbf{d}_T, \mathbf{d}_0) &\ge h_T(\mathbf{x}_T, \mathbf{d}_T) \ \ \forall \mathbf{d}_T \in \Xi_{T-1} \\
\hat{\underline{\mathbf{q}}}_k, \hat{\bar{\mathbf{q}}}_k &\ge 0 \\
\hat{\mathbf{R}}_1 \succeq \mathbf{0},\ \hat{\mathbf{R}}_k &\succeq \mathbf{0} \\
\hat{\underline{\mathbf{q}}}_1, \hat{\bar{\mathbf{q}}}_1 &\ge 0,
\end{aligned} \tag{16}$$

will also satisfy the constraint (C-DRO) *for any* $P \in \widetilde{\mathcal{U}}$*.*

Proof Apply Proposition 2 for $t = T$ and note that

$$g_T(\mathbf{x}_T, \mathbf{d}_{T-1}) = \sup_{P_{T|T-1} \in \widetilde{\mathcal{U}}_T(\mathbf{d}_{T-1})} \mathbb{E}_{P_{T|T-1}}[h_T(\mathbf{x}_T, \mathbf{d}_T)].$$

□

Proposition 3 reduces the difficulty of finding a feasible solution to the robust counterpart in (8). This comes with the caveat that while the original problem may be feasible, a static and feasible solution may not exist. This is because ARO problems do not necessarily have static feasible solutions. In some settings, more complicated decision rules have been proposed to overcome this limitation [10]. Such flexibility comes at the price of higher complexity of cut-generating algorithms that may be used to solve the optimization problem.

Despite the reduction in problem complexity from Proposition 3, we still have to solve a feasibility problem for a set of semi-infinite constraints in (16). For this, we may use an optimization problem to identify a violated constraint. Depending on the nature of the support sets Ξ_t and the problem coefficients $\hat{\mathbf{R}}_t$ and the function h_t, this problem may not be convex in $\mathbf{d}_t$ and, hence, not straightforward to solve. In some cases, the constraint structure can be leveraged to find violated inequalities using common optimization algorithms. Let $\boldsymbol{\mu}_{t+1}(\mathbf{d}_t) = \mathbf{A}\boldsymbol{\mu}_t^0 + \mathbf{B}\mathbf{d}_t$ a linear function of $\mathbf{d}_t$ ($\boldsymbol{\mu}_t^0$ does not depend on $\mathbf{d}_t$) and let $\boldsymbol{\Sigma}'_{t+1}(\mathbf{d}_t)$ be a quadratic function of $\mathbf{d}_t$ of the form $\boldsymbol{\Sigma}'_{t+1}(\mathbf{d}_t) = \alpha \mathbf{d}_t \mathbf{d}_t^\top + \mathbf{B}\mathbf{d}_t\mathbf{d}_t^\top\mathbf{B}^\top +$ linear terms. Table 1 highlights potential algorithms for solving the cut-generation subproblems.

$h_t(\mathbf{x}_t, \mathbf{d}_t)$	$\mathbf{R}_t - \alpha\mathbf{R}_{t+1} - \mathbf{B}^\top\mathbf{R}_{t+1}\mathbf{B}$	Type of Problem	Potential Algorithm
concave in $\mathbf{d}_t$	$\succeq \mathbf{0}$	Concave maximization	Interior point
convex in $\mathbf{d}_t$	$\preceq \mathbf{0}$	Convex maximization	Leverage structure of Ξ_t : use vertex enumeration for polyhedral Ξ_t
arbitrary	undetermined	Non-convex problem	Global optimization

Table 1 Methods for cut-generation subproblem in Proposition 3

2.3 Application: Portfolio Optimization

In this section, we evaluate the performance of the CU framework on a portfolio optimization problem by using historical stock data. We consider a 2-period problem, in which the choice of the portfolio has to be made at the beginning. Here, each period corresponds to one week. At the end of the first week, the assets can be reallocated. However, this reallocation has to be specified in the beginning. For our portfolio, we choose among 5 stocks. We conduct 150 experiments for randomly selected dates. At each date, we compute the weekly returns for the past 100 weeks based on stock price data [42]. They are calculated using the difference between the closing prices of the two consecutive weeks, which results in a single number per stock for each week. To capture risk aversion, in each experiment, we maximize a concave piecewise linear utility function. It is assumed that $\overline{\boldsymbol{\mu}}_2 = \boldsymbol{\mu}_2(\mathbf{d}_1) + \boldsymbol{\delta}$ and $\underline{\boldsymbol{\mu}}_2 = \boldsymbol{\mu}_2(\mathbf{d}_1) - \boldsymbol{\delta}$, where $\boldsymbol{\mu}_2(\mathbf{d}_1) = \boldsymbol{\mu}_0 + \mathbf{A}\boldsymbol{\mu}_1 + \mathbf{B}\mathbf{d}_1$. The vector $\boldsymbol{\mu}_0$ and the matrices $\mathbf{A}$ and $\mathbf{B}$ are estimated via a time series model. Specifically, vector autoregressive moving average with lag 1 for both is used to fit the data [38]. The parameter $\boldsymbol{\Sigma}$ is set to the residual covariance matrix. The value of $\boldsymbol{\mu}_1$ is the return point estimate at the end of the first week and $\boldsymbol{\delta}$ is three times the standard error of this estimate (to cover almost all realized means under normality assumption). We compare the CU model against the standard DRO and a modified version of the latter as following:

CU: Model with connected uncertainty set,
DRO-1: DRO model with $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_1$, and
DRO-2: DRO model with $\boldsymbol{\mu}_2 = \mathbf{A}\boldsymbol{\mu}_1 + \mathbf{B}\boldsymbol{\mu}_1$.

While the parameter fit is based on the past weeks, the returns are computed over the future two weeks. The quality of solutions from the three models is then evaluated by comparing the returns. Each experiment starts with an initial \$100 wealth, which is recomputed according to the asset allocation and the realized returns.

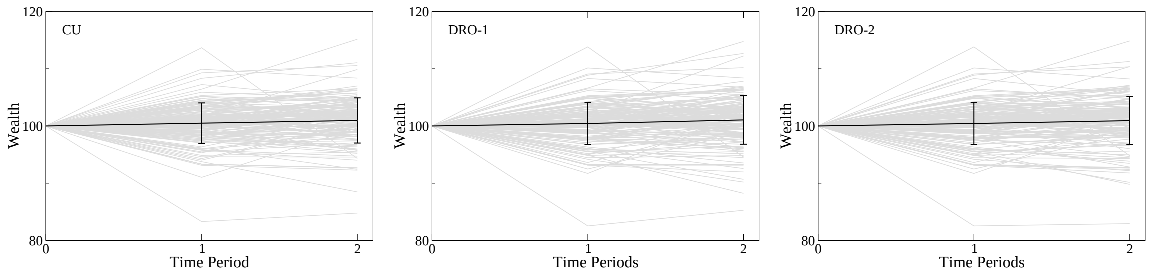

Fig. 2 Comparison of the CU and Standard DRO Models: gray lines for sample paths, bold line for the mean, and error bars for the standard deviation.

Figure 2 displays the performances of the solutions from the three different models. The light grey lines depict the sample paths of the 150 different experiments for different days and the black line depicts the mean sample path. We observe that in all three models, the mean is increasing in time, which is attributed to the mean positive return on the random samples for the five stocks. Furthermore, the

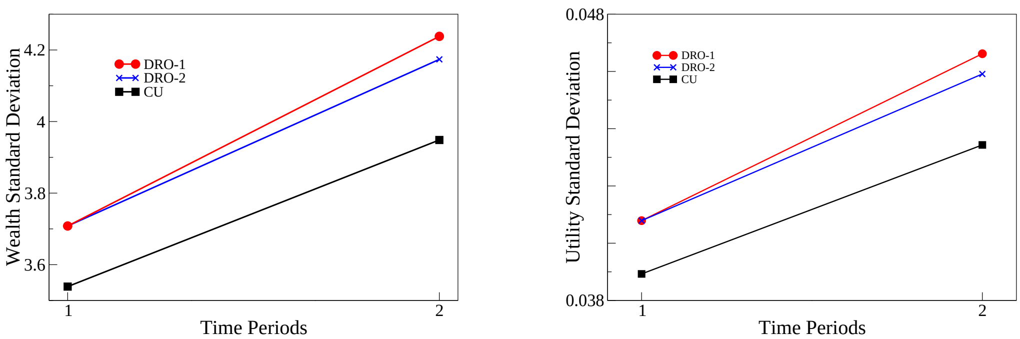

Fig. 3 Comparison of CU and standard DRO models with respect to the wealth (left) and utility (right) standard deviation.

spread of the sample paths grows as time elapses because of the inherent uncertainty in returns. When comparing the three models, we observe a lower standard deviation of the sample paths for the CU model than both of the DRO models. Figure 3(left) highlights that the CU model is able to lower the risk for (almost) the same return. It reduces the standard deviation for wealth by 4.6% for period 1 and 6.8% for period 2 compared to DRO-1. A similar reduction occurs for the utility of the realized returns, as shown in Figure 3(right). Other measures of deviation such as interquartile range depict similar decreases for both wealth and utility. This advantage arises because CU sets grow more conservatively in time than standard DRO approaches and enables the CU model to select assets that are less volatile. We also observe a marginally better increase in the means for the CU model, which we attribute to the specifics of the used sample data.

Note that solving the CU model is computationally more demanding than either of the DRO models, because of latter's convex subproblems. Therefore, depending on the application, the advantages of the CU model have to outweigh the additional computational burden. Furthermore, the DRO-2 model represents another way to leverage the connection between uncertainties of different periods. However, it only accounts for the average effect between periods and not the worst-case, which is more aligned with the robust optimization paradigm and requires solving an optimization problem.

So far we have focused on distributionally robust optimization problems. CU sets also arise in the context of standard robust optimization problems. The next section provides tractable reformulations for RO problems with CU sets.

3 RO with CU Sets

For optimization problems under uncertainty, RO is an alternative to DRO when distributions are not available and achieves tractable and exact formulations for many problems. We showed that the reformulation of DRO problems for CU sets leads to IOPs which are, in general, intractable and can be solved by approximations. In this section, we leverage the effectiveness of RO for CU sets. To fully harness the potential of RO and dual reformulations, we focus on linear constraints. First, we discuss linear dependence between sets that are polyhedral or ellipsoidal. Linear dependence occurs in various models such as autoregressive models and minimum mean square error predictors (linear or jointly normal) [20]. For polyhedral sets, our discussion focuses on RHS and LHS (left-hand side) coefficients (of the uncertainty set) depending on previous realizations. For ellipsoidal sets, we study the dependence of the set center and the covariance matrix on past observations. We then present a conservative reformulation for the case of quadratic dependence in ellipsoidal uncertainty sets.

The goal of this section is to provide a reformulation of the constraint

$$\sum_{t=1}^{T} \mathbf{d}_t^\top \mathbf{x}_t \leq B, \tag{C-RO}$$

under CU sets. Here $\mathbf{d}_t$ is the vector of uncertain coefficients, $\mathbf{x}_t$ the decision variables, and B some constant upper bound spanning all T periods. In each period t, the uncertain parameter $\mathbf{d}_t$ resides in a set $\mathcal{U}_t(\cdot)$, which may depend on previous realizations of the uncertainty. In the following, we specify the dependence for different set geometries.

3.1 Polyhedral

Polyhedral uncertainty sets, such as budgeted sets, box sets, and others, are quite common in the RO literature [4, 6, 9]. They have the advantage of preserving the complexity of the original problem, and hence maintaining tractability. This section extends these approaches to the CU setting. Specifically, it considers the case of a robust problem, in which the coefficients of each period reside in a polyhedral CU set. The parameters of this set, i.e., the LHS coefficient matrix and the RHS vector, can depend on the realization of the previous uncertainty. Here, we provide a reformulation of constraint (C-RO) for RHS coefficient dependence and then present an example for LHS dependency.

Uncertainty in the RHS vector

Consider (C-RO) and assume each term $\mathbf{d}_t$ to be uncertain and residing in a polyhedral set of the form

$$\mathcal{U}_t(\mathbf{d}_{t-1}) = \{\mathbf{d}_t \mid \mathbf{A}_t \mathbf{d}_t = \mathbf{b}_t + \boldsymbol{\Delta}_t \mathbf{d}_{t-1},\ \ \mathbf{d}_t \geq \mathbf{0}\}, \tag{P}$$

where the parameters of the first set $\mathcal{U}_1$ are known, i.e., matrix $\boldsymbol{\Delta}_1 = \mathbf{0}$. The robust counterpart of constraint (C-RO) is given by

$$\max_{\mathbf{d}_1 \in \mathcal{U}_1} \{\mathbf{d}_1^\top \mathbf{x}_1 + \max_{\mathbf{d}_2 \in \mathcal{U}_2(\mathbf{d}_1)} \{\mathbf{d}_2^\top \mathbf{x}_2 + \cdots + \max_{\mathbf{d}_T \in \mathcal{U}_T(\mathbf{d}_{T-1})} \mathbf{d}_T^\top \mathbf{x}_T\}\} \leq B.$$

The following lemma provides a partial reformulation of this constraint.

Lemma 1 *If* $\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_{T+1} = \mathbf{0}$*, the uncertainty is governed by* (P)*, and* $\mathbf{d}_1, \ldots, \mathbf{d}_{T-\tau}$ *for* $\tau \in \{0, \ldots, T-1\}$ *are known, then the robust counterpart of constraint* (C-RO) *with respect to* $\mathbf{d}_T, \mathbf{d}_{T-1}, \ldots, \mathbf{d}_{T-\tau+1}$ *is given by*

$$\begin{aligned}&\sum_{t=1}^{T-\tau} \mathbf{d}_t^\top \mathbf{x}_t + \sum_{t=T-\tau+1}^{T} \mathbf{q}_t^\top \mathbf{b}_t + \mathbf{q}_{T-\tau+1}^\top \boldsymbol{\Delta}_{T-\tau+1} \mathbf{d}_{T-\tau} \le B \\ &\mathbf{q}_t^\top \mathbf{A}_t \ge \mathbf{x}_t^\top + \mathbf{q}_{t+1}^\top \boldsymbol{\Delta}_{t+1} \ \ \forall i = T-\tau+1, \ldots, T.\end{aligned}$$

Proof The proof proceeds by induction, going back from $\mathbf{d}_T$ towards $\mathbf{d}_1$. First, it proves the statement for $\mathbf{d}_T$. Then it assumes the statement to be true for $\mathbf{d}_{T-\tau+1}$ and verifies it for $\mathbf{d}_{T-\tau}$.

Base case ($\tau = 1$): The constraint (C-RO) can be expanded as

$$\sum_{t=1}^{T-1} \mathbf{d}_t^\top \mathbf{x}_t + \mathbf{d}_T^\top \mathbf{x}_T \le B. \tag{17}$$

Since this constraint has to hold for all $\mathbf{d}_T \in \mathcal{U}_T$, then the term $\mathbf{d}_T^\top \mathbf{x}_T$ can be replaced by the following

$$\begin{aligned}\max_{\mathbf{d}_T}\ & \mathbf{d}_T^\top \mathbf{x}_T \\ \text{s.t. } & \mathbf{A}_T \mathbf{d}_T = \mathbf{b}_T + \boldsymbol{\Delta}_T \mathbf{d}_{T-1} \\ & \mathbf{d}_T \ge \mathbf{0}.\end{aligned}$$

Substituting the dual of the above problem into the constraint (17), the robust counterpart of the constraint (17) is

$$\begin{aligned}&\sum_{t=1}^{T-1} \mathbf{d}_t^\top \mathbf{x}_t + \mathbf{q}_T^\top \mathbf{b}_T + \mathbf{q}_T^\top \boldsymbol{\Delta}_T \mathbf{d}_{T-1} \le B \\ &\mathbf{q}_T^\top \mathbf{A}_T \ge \mathbf{x}_T^\top.\end{aligned}$$

Since $\boldsymbol{\Delta}_{T+1} = 0$, we can write

$$\begin{aligned}&\sum_{t=1}^{T-1} \mathbf{d}_t^\top \mathbf{x}_t + \mathbf{q}_T^\top \mathbf{b}_T + \mathbf{q}_T^\top \boldsymbol{\Delta}_T \mathbf{d}_{T-1} \le B \\ &\mathbf{q}_T^\top \mathbf{A}_T \ge (\mathbf{x}_T + \boldsymbol{\Delta}_{T+1}^\top \mathbf{q}_{T+1})^\top.\end{aligned}$$

Inductive case ($\tau = k$): To simplify notation, let $m = T - k$. Then, the robust counterpart of constraint (C-RO) with respect to the terms $\mathbf{d}_T, \mathbf{d}_{T-1}, \ldots, \mathbf{d}_{m+1}$ is given by

$$\begin{aligned}&\sum_{t=1}^{m} \mathbf{d}_t^\top \mathbf{x}_t + \sum_{t=m+1}^{T} \mathbf{q}_t^\top \mathbf{b}_t + \mathbf{q}_{m+1}^\top \boldsymbol{\Delta}_{m+1} \mathbf{d}_m \le B \\ &\mathbf{q}_t^\top \mathbf{A}_t \ge (\mathbf{x}_t + \boldsymbol{\Delta}_{t+1}^\top \mathbf{q}_{t+1})^\top \ \ \forall i = m+1, \ldots, T.\end{aligned} \tag{18}$$

The first constraint of (18) can be rewritten as

$$\sum_{t=1}^{m-1} \mathbf{d}_t^\top \mathbf{x}_t + \mathbf{d}_m^\top (\mathbf{x}_m + \boldsymbol{\Delta}_{m+1}^\top \mathbf{q}_{m+1}) + \sum_{t=m+1}^{T} \mathbf{q}_t^\top \mathbf{b}_t \le B. \tag{19}$$

Since $\mathbf{d}_m$ is uncertain, and the constraint (19) has to hold for $\mathbf{d}_m \in \mathcal{U}_m(\mathbf{d}_{m-1})$, it also holds if $\mathbf{d}_m^\top (\mathbf{x}_m + \boldsymbol{\Delta}_{m+1}^\top \mathbf{q}_{m+1})$ is replaced by the following problem

$$\begin{aligned}\max_{\mathbf{d}_m}\ & \mathbf{d}_m^\top (\mathbf{x}_m + \boldsymbol{\Delta}_{m+1}^\top \mathbf{q}_{m+1}) \\ \text{s.t. } & \mathbf{A}_m \mathbf{d}_m = \mathbf{b}_m + \boldsymbol{\Delta}_m \mathbf{d}_{m-1} \\ & \mathbf{d}_m \ge \mathbf{0}.\end{aligned}$$

Substituting the dual of this problem into constraint (19), the robust counterpart of the first constraint in (18) can be written as

$$\begin{aligned}&\sum_{t=1}^{m-1} \mathbf{d}_t^\top \mathbf{x}_t + \mathbf{q}_m^\top(\mathbf{b}_m + \boldsymbol{\Delta}_m \mathbf{d}_{m-1}) + \sum_{t=m+1}^{T} \mathbf{q}_t^\top \mathbf{b}_t \le B \\ &\mathbf{q}_m^\top \mathbf{A}_m \ge (\mathbf{x}_m + \boldsymbol{\Delta}_{m+1}^\top \mathbf{q}_{m+1})^\top.\end{aligned}$$

Hence, by induction, the proof is complete. □

In the above lemma, the term $\mathbf{q}_t^\top \mathbf{b}_t$ safeguards against uncertainty realizations. The dual variable $\mathbf{q}_t$ is appropriately adjusted by the constraint $\mathbf{q}_t^\top \mathbf{A}_t \ge (\mathbf{x}_t + \boldsymbol{\Delta}_{t+1}^\top \mathbf{q}_{t+1})^\top$ to account for connected uncertainties. The following proposition leads to a complete reformulation by repeated application of Lemma 1.

Proposition 4 *The robust counterpart of constraint* (C-RO) *under the uncertainty set* (P) *is given by*

$$\begin{aligned}&\sum_{t=1}^{T} \mathbf{q}_t^\top \mathbf{b}_t \le B \\ &\mathbf{q}_t^\top \mathbf{A}_t \ge \mathbf{x}_t^\top + \mathbf{q}_{t+1}^\top \boldsymbol{\Delta}_{t+1} \ \ \forall i = 1, \dots, T,\end{aligned}$$

where $\boldsymbol{\Delta}_1 = 0$ *and* $\boldsymbol{\Delta}_{T+1} = 0$.

Proof By Lemma 1, the robust counterpart of (C-RO) with respect to $\mathbf{d}_T, \mathbf{d}_{T-1}, \dots, \mathbf{d}_{T-\tau+1}$ is

$$\begin{aligned}&\sum_{t=1}^{T-\tau} \mathbf{d}_t^\top \mathbf{x}_t + \sum_{t=T-\tau+1}^{T} \mathbf{q}_t^\top \mathbf{b}_t + \mathbf{q}_{T-\tau+1}^\top \boldsymbol{\Delta}_{T-\tau+1} \mathbf{d}_{T-\tau} \le B \\ &\mathbf{q}_t^\top \mathbf{A}_t \ge \mathbf{x}_t^\top + \mathbf{q}_{t+1}^\top \boldsymbol{\Delta}_{t+1} \ \ \forall t = T-\tau+1, \dots, T.\end{aligned}$$

By repeating this step inductively until $\tau = T-1$, we obtain

$$\begin{aligned}&\mathbf{d}_1 \mathbf{x}_1 + \sum_{t=2}^{T} \mathbf{q}_t^\top \mathbf{b}_t + \mathbf{q}_2^\top \boldsymbol{\Delta}_2 \mathbf{d}_1 \le B \\ &\mathbf{q}_t^\top \mathbf{A}_t \ge \mathbf{x}_t^\top + \mathbf{q}_{t+1}^\top \boldsymbol{\Delta}_{t+1} \ \ \forall t = 2, \dots, T.\end{aligned}$$

Taking the dual with respect to $\mathbf{d}_1$ and using $\boldsymbol{\Delta}_1 = 0$, we obtain

$$\begin{aligned}&\sum_{t=1}^{T} \mathbf{q}_t^\top \mathbf{b}_t \le B \\ &\mathbf{q}_t^\top \mathbf{A}_t \ge \mathbf{x}_t^\top + \mathbf{q}_{t+1}^\top \boldsymbol{\Delta}_{t+1} \ \ \forall t = 1, \dots, T,\end{aligned}$$

which proves the proposition. □

Note that an alternative way of proving Proposition 4 is by imposing the constraints for each period at once and directly computing the dual over the intersection of the sets.

The RHS coefficients of (P) depend on the previous realization, where the exact dependence is controlled by $\boldsymbol{\Delta}_t$. This matrix can be created based on the application and the prediction mechanism for future uncertainties constructed, e.g., by time series models.

Uncertainty in the coefficient matrix

We now consider an example in which the LHS coefficients of the CU set are affected by the dependence. Consider the constraint (C-RO) with $T = 2$,

$$\mathbf{d}_1^\top \mathbf{x}_1 + \mathbf{d}_2^\top \mathbf{x}_2 \le B, \tag{20}$$

where $\mathbf{d}_1 \in \mathcal{U}_1$ and $\mathbf{d}_2 \in \mathcal{U}_2(\mathbf{d}_1)$. The uncertainty sets are described by

$$\begin{aligned}\mathcal{U}_1 &= \{\mathbf{d}_1 \mid \mathbf{A}_1 \mathbf{d}_1 = \mathbf{b}_1, \ \mathbf{d}_1 \ge \mathbf{0}\} \\ \mathcal{U}_2(\mathbf{d}_1) &= \{\mathbf{d}_2 \mid (\mathbf{A}_{2,k} + \boldsymbol{\Delta}_k \mathbf{d}_1)^\top \mathbf{d}_2 = b_{2k} \ \ \forall k \ , \ \mathbf{d}_2 \ge \mathbf{0}\}.\end{aligned}$$

Note the presence of additional constraints $\mathbf{d}_1 \geq \mathbf{0}$ and $\mathbf{d}_2 \geq \mathbf{0}$, which ensure that the dual problem does not have equality constraints. Also note that each $\mathbf{A}_{2,k,\cdot}^\top$ is a vector representing the kth row of the matrix $\mathbf{A}_2$ and the element $\boldsymbol{\Delta}_k$ is a matrix influencing the update of the kth row. In order to develop the robust counterpart, consider the problem

$$\begin{aligned}&\max \mathbf{d}_2^\top \mathbf{x}_2\\ &\text{s.t. } (\mathbf{A}_{2,k,\cdot} + \boldsymbol{\Delta}_k \mathbf{d}_1)^\top \mathbf{d}_2 = b_{2,k} \ \ \forall k\\ &\quad\ \mathbf{d}_2 \geq \mathbf{0}.\end{aligned}$$

Substituting the dual of the given problem back into (20) leads to

$$\begin{aligned}&\mathbf{d}_1^\top \mathbf{x}_1 + \mathbf{q}_2^\top \mathbf{b}_2 \leq B\\ &\sum_k q_{2,k} (\mathbf{A}_{2,k,\cdot} + \boldsymbol{\Delta}_k \mathbf{d}_1)^\top \geq \mathbf{x}_2^\top.\end{aligned}$$

The uncertain coefficient $\mathbf{d}_1$ is present in both constraints. To simplify, let $\mathbf{Q} = \sum_k q_{2,k} \boldsymbol{\Delta}_k^\top$, hence the second set of constraints in the above problem can be written as

$$\mathbf{q}_2^\top \mathbf{A}_2 + \mathbf{d}_1^\top \mathbf{Q} \geq \mathbf{x}_2^\top. \tag{21}$$

The set of constraints (21) is actually a set of constraints corresponding to different elements of $\mathbf{x}_2$. Since robust optimization is constraint wise, the above set of constraints can be expressed by

$$\begin{aligned}&\mathbf{d}_1^\top \mathbf{x}_1 + \mathbf{q}_2^\top \mathbf{b}_2 \leq B\\ &\mathbf{q}_2^\top \mathbf{A}_{2,\cdot,i} + \mathbf{d}_1^\top \mathbf{Q}_{\cdot,i} \geq x_{2,i} \ \ \forall i.\end{aligned}$$

Here $\mathbf{A}_{2,\cdot,i}$ and $\mathbf{Q}_{\cdot,i}$ represents the ith columns of $\mathbf{A}_2$ and $\mathbf{Q}$, respectively. Substituting the dual of the maximization of $\mathbf{d}_1^\top \mathbf{x}_1$ and the minimization of $\mathbf{Q}_{\cdot,i}^\top \mathbf{d}_1$ over the uncertainty set $\mathcal{U}_1$, the robust counterpart is

$$\begin{aligned}&\mathbf{q}_1^\top \mathbf{b}_1 + \mathbf{q}_2^\top \mathbf{b}_2 \leq B\\ &\mathbf{q}_2^\top \mathbf{A}_{2,\cdot,i} + \mathbf{r}_t^\top \mathbf{b}_1 \geq x_{2,i} \ \ \forall i\\ &\mathbf{q}_1^\top \mathbf{A}_1 \geq \mathbf{x}_1^\top\\ &\mathbf{r}_i^\top \mathbf{A}_1 \leq \sum_k q_{2,k} \boldsymbol{\Delta}_{k,i,\cdot}^\top \ \ \forall i.\end{aligned}$$

3.2 Ellipsoid Center Dependence

Ellipsoids are commonly used as uncertainty sets because of their natural interpretation as confidence or predictive regions [4]. For these sets, there are three main parameters that can be affected by previous uncertainty realizations: the center, the covariance matrix, and the radius of the ellipsoid. A major difference from the DRO case is that the parameters depend on realizations from all past periods and not just the previous period. This section assumes that the center of the ellipsoid depends on the previous period realization and the other two parameters are constant. For constraints of the form (C-RO), the uncertainty set for each $\mathbf{d}_t$ is

$$\mathcal{U}_t(\mathbf{d}_{t-1}) = \{\mathbf{d}_t \mid \mathbf{d}_t = \boldsymbol{\mu}_t(\mathbf{d}_{t-1}) + \mathbf{L}_t \mathbf{u}_t : \|\mathbf{u}_t\|_2 \leq r_t\}, \tag{E}$$

where $\mathbf{L}_t \mathbf{L}_t^\top = \boldsymbol{\Sigma}_t$, with $\boldsymbol{\Sigma}_t$ as the covariance matrix of the ellipsoid. The robust counterpart of constraint (C-RO) is given by

$$\max_{\mathbf{d}_1 \in \mathcal{U}_1} \left\{ \mathbf{d}_1^\top \mathbf{x}_1 + \max_{\mathbf{d}_2 \in \mathcal{U}_2(\mathbf{d}_1)} \{\mathbf{d}_2^\top \mathbf{x}_2 + \cdots + \max_{\mathbf{d}_T \in \mathcal{U}_T(\mathbf{d}_{T-1})} \mathbf{d}_T^\top \mathbf{x}_T\} \right\} \leq B.$$

Let the dependence of the center on the previous period uncertainty realization be described by

$$\boldsymbol{\mu}_t(\mathbf{d}_{t-1}) = \mathbf{A}_t \boldsymbol{\mu}_{t-1}(\mathbf{d}_{t-2}) + \mathbf{F}_t \mathbf{d}_{t-1} + \mathbf{c}_t,$$

where $\mathbf{A}_t$ and $\mathbf{F}_t$ are matrices and $\mathbf{c}_t$ is a vector. The following lemma provides the reformulation of constraint (C-RO).

Lemma 2 *For $s_{T-\tau} = \sum_{t=1}^{T-\tau} \mathbf{d}_t^\top \mathbf{x}_t$, the robust counterpart of the constraint* (C-RO) *for the ellipsoidal uncertainty set* (E) *is given by*

$$s_{T-\tau} + \boldsymbol{\mu}_{T-\tau+1}(\mathbf{d}_{T-\tau})^\top \mathbf{y}_{T-\tau+1} + C_{T-\tau+2} + R_{T-\tau+1} \leq B,$$

for any $\tau \in \{1, \dots, T\}$, where $\mathbf{y}_t$ is recursively defined as $\mathbf{y}_{T-\tau} = \mathbf{x}_{T-\tau} + (\mathbf{F}_{T-\tau+1} + \mathbf{A}_{T-\tau+1})^\top \mathbf{y}_{T-\tau+1}$ and $\mathbf{y}_T = \mathbf{x}_T$. The aggregates $C_{T-\tau}$ and $R_{T-\tau}$ are recursively defined as

$C_{T-\tau} = \mathbf{c}_{T-\tau+1}^\top \mathbf{y}_{T-\tau+1} + C_{T-\tau+1}$, *with* $C_{T+1} = 0$ *and*
$R_{T-\tau} = r_{T-\tau} \| L_{T-\tau}^\top (\mathbf{x}_{T-\tau} + \mathbf{F}_{T-\tau+1} \mathbf{y}_{T-\tau+1}) \| + R_{T-\tau+1}$ *with* $R_T = r_T \| \mathbf{L}_T^\top \mathbf{x}_T \|$.

Proof The procedure for the proof is similar to the polyhedral case. It starts by proving the statement for the term involving $\mathbf{d}_T$. Then the reformulation is assumed to be true for $\mathbf{d}_{T-k+1}$, before being proven for $\mathbf{d}_{T-k}$.

Base case ($\tau = 1$): The constraint (C-RO) can be expanded as

$$s_{T-1} + \mathbf{d}_T^\top \mathbf{x}_T + C_{T+1} \leq B.$$

Substitute $\mathbf{d}_T = \boldsymbol{\mu}_T(\mathbf{d}_{T-1}) + \mathbf{L}_T \mathbf{u}_T$. Since the constraint must hold for all $\mathbf{u}_T \in \{\mathbf{u}_T \mid \|\mathbf{u}_T\|_2 \leq r_T\}$, it also holds for the robust counterpart

$$s_{T-1} + \boldsymbol{\mu}_T(\mathbf{d}_{T-1})^\top \mathbf{x}_T + r_T \|\mathbf{L}_T^\top \mathbf{x}_T\|_2 + C_{T+1} \leq B.$$

Since we know that $\mathbf{y}_T = \mathbf{x}_T$, $C_{T+1} = 0$, and $R_T = r_T \|\mathbf{L}_T^\top \mathbf{x}_T\|$, we have the desired result

$$s_{T-1} + \boldsymbol{\mu}_T(\mathbf{d}_{T-1})^\top \mathbf{y}_T + C_{T+1} + R_T \leq B.$$

Inductive case ($\tau = k$): Assume that the result holds for $T-k+1$. Then the robust counterpart of (C-RO) with respect to $\mathbf{d}_T, \mathbf{d}_{T-1}, \dots, \mathbf{d}_{T-k+1}$ is given by

$$s_{T-k} + \boldsymbol{\mu}_{T-k+1}(\mathbf{d}_{T-k})^\top \mathbf{y}_{T-k+1} + C_{T-k+2} + R_{T-k+1} \leq B.$$

Substituting the mean dependence, this can be expressed as

$$s_{T-k} + \boldsymbol{\mu}_{T-k}(\mathbf{d}_{T-k-1})^\top \mathbf{A}_{T-k+1}^\top \mathbf{y}_{T-k+1} + \mathbf{d}_{T-k}^\top \mathbf{F}_{T-k+1}^\top \mathbf{y}_{T-k+1} + \mathbf{c}_{T-k+1}^\top \mathbf{y}_{T-k+1} + C_{T-k+2} + R_{T-k+1} \leq B.$$

After rearranging the terms, we obtain

$$\begin{aligned} s_{T-k-1} + \mathbf{d}_{T-k}^\top (\mathbf{x}_{T-k} + \mathbf{F}_{T-k+1}^\top \mathbf{y}_{T-k+1}) + \boldsymbol{\mu}_{T-k}(\mathbf{d}_{T-k-1})^\top \mathbf{A}_{T-k+1}^\top \mathbf{y}_{T-k+1} + \\ \mathbf{c}_{T-k+1}^\top \mathbf{y}_{T-k+1} + C_{T-k+2} + R_{T-k+1} \leq B. \end{aligned}$$

Using the uncertainty set (E), this can be rewritten as

$$\begin{aligned} s_{T-k-1} + \boldsymbol{\mu}_{T-k}(\mathbf{d}_{T-k-1})^\top (\mathbf{x}_{T-k} + \mathbf{F}_{T-k+1}^\top \mathbf{y}_{T-k+1}) + \mathbf{u}_{T-k}^\top \mathbf{L}_{T-k}^\top (\mathbf{x}_{T-k} + \mathbf{F}_{T-k+1}^\top \mathbf{y}_{T-k+1}) \\ + \boldsymbol{\mu}_{T-k}(\mathbf{d}_{T-k-1})^\top \mathbf{A}_{T-k+1}^\top \mathbf{y}_{T-k+1} + \mathbf{c}_{T-k+1}^\top \mathbf{y}_{T-k+1} + C_{T-k+2} + R_{T-k+1} \leq B. \end{aligned}$$

Taking the robust counterpart with respect to $\mathbf{u}_{T-k}$ we can write

$$\begin{aligned} & s_{T-k-1} + \boldsymbol{\mu}_{T-k}(\mathbf{d}_{T-k-1})^\top (\mathbf{x}_{T-k} + \mathbf{F}_{T-k+1}^\top \mathbf{y}_{T-k+1} + \mathbf{A}_{T-k+1}^\top \mathbf{y}_{T-k+1}) + \\ & r_{T-k} \|\mathbf{L}_{T-k}^\top (\mathbf{x}_{T-k} + \mathbf{F}_{T-k+1}^\top \mathbf{y}_{T-k+1})\| + \mathbf{c}_{T-k+1}^\top \mathbf{y}_{T-k+1} + C_{T-k+2} + R_{T-k+1} \leq B. \end{aligned}$$

Using $\mathbf{y}_{T-k} = \mathbf{x}_{T-k} + \mathbf{F}_{T-k+1}^\top \mathbf{y}_{T-k+1} + \mathbf{A}_{T-k+1}^\top \mathbf{y}_{T-k+1}$ and the definitions of R_{T-k}, and C_{T-k} we obtain the desired result

$$s_{T-k-1} + \boldsymbol{\mu}_{T-k}(\mathbf{d}_{T-k-1})^\top \mathbf{y}_{T-k} + C_{T-k+1} + R_{T-k} \leq B.$$

This completes the induction. □

The complete reformulation of constraint (C-RO) for the uncertainty set (E) follows straightforwardly by using conic duality and Lemma 2 as in the following proposition.

Proposition 5 *The robust counterpart of constraint* (C-RO) *for set* (E) *is given by*

$$\boldsymbol{\mu}_1^\top \mathbf{y}_1 + C_2 + R_1 \leq B,$$

where $\mathbf{y}_{T-k} = \mathbf{x}_{T-k} + (\mathbf{F}_{T-k+1} + \mathbf{A}_{T-k+1})^\top \mathbf{y}_{T-k+1}$ and $\mathbf{y}_T = \mathbf{x}_T$. The aggregate C_{T-k} is recursively defined as $C_{T-k} = \mathbf{c}_{T-k+1}^\top \mathbf{y}_{T-k+1} + C_{T-k+1}$ with $C_{T+1} = 0$. Corresponding, the aggregate R_{T-k} is defined as $R_{T-k} = r_{T-k} \|L_{T-k}^\top (\mathbf{x}_{T-k} + \mathbf{F}_{T-k+1} \mathbf{y}_{T-k+1})\| + R_{T-k+1}$ with $R_T = r_T \|\mathbf{L}_T^\top \mathbf{x}_T\|$.

3.3 Matrix Dependence

We examine the case, in which the previous uncertainty realizations affect the covariance matrix of the current ellipsoid. The mean $\boldsymbol{\mu}_t$ is assumed to be known and not affected by the realization. The general form of the matrix update is described by

$$\boldsymbol{\Sigma}_{t+1}(\mathbf{d}_t) = a_t \boldsymbol{\Sigma}_t(\mathbf{d}_{t-1}) + f_t(\mathbf{d}_t - \boldsymbol{\mu}_t)(\mathbf{d}_t - \boldsymbol{\mu}_t)^\top + \mathbf{C}_t. \tag{22}$$

The above dependence is inspired by updates of the covariance matrix estimates using a frequentist or Bayesian paradigm, where the former assumes no prior distribution, and the latter assumes one. For the updates in (22), the uncertainty set is the same ellipsoidal set (E) as used in Lemma 2. For brevity, we do not explicitly denote the dependence of $\boldsymbol{\Sigma}_{t+1}(\mathbf{d}_t)$ on $\mathbf{d}_t$. The following lemma provides a partial reformulation of constraint (C-RO).

Lemma 3 *For some $\tau \in \{1, \ldots, T\}$, let $\mathbf{d}_{T-\tau}$ reside in an ellipsoid $\mathcal{E}$ with center $\boldsymbol{\mu}_{T-\tau}$, covariance matrix $\boldsymbol{\Sigma}_{T-\tau}$, and radius $r_{T-\tau}$ and let $\boldsymbol{\Sigma}_{T-\tau}$ be updated as*

$$\boldsymbol{\Sigma}_{T-\tau+1} = a_\tau \boldsymbol{\Sigma}_{T-\tau} + f_\tau(\mathbf{d}_{T-\tau} - \boldsymbol{\mu}_{T-\tau})(\mathbf{d}_{T-\tau} - \boldsymbol{\mu}_{T-\tau})^\top + \mathbf{C}_\tau,$$

where $a_\tau \geq 0, f_\tau \geq 0, \mathbf{C}_\tau \succeq 0$ and $\boldsymbol{\Sigma}_1 \succeq 0$ are constants, and $\mathbf{L}_t\mathbf{L}_t^\top = \boldsymbol{\Sigma}_t$.
Let $\mathbf{y}_{T-\tau} = \mathbf{x}_{T-\tau} + \sum_{j=1}^{\tau} \mathbf{y}_{T-j+1} r_{T-j+1}\sqrt{f_\tau A_{\tau,j}}$ with $\mathbf{y}_{T+1} = 0$ and $\mathbf{y}_T = \mathbf{x}_T$. A conservative robust reformulation of constraint (C-RO) *with respect to uncertain coefficients $\mathbf{d}_{T-\tau+1}, \ldots, \mathbf{d}_T$ is given by*

$$s_{T-\tau} + \Theta_{T-\tau+1} + \sum_{i=1}^{\tau} r_{T-i+1}\sqrt{A_{\tau,i}\mathbf{y}_{T-i+1}^\top \boldsymbol{\Sigma}_{T-\tau+1}\mathbf{y}_{T-i+1}} + R_{T-\tau+2} \leq B,$$

with

$$s_{T-\tau} = \sum_{t=1}^{T-\tau} \mathbf{d}_t^\top \mathbf{x}_t$$

$$\Theta_{T-\tau+1} = \sum_{t=T-\tau+1}^{T} \boldsymbol{\mu}_t^\top \mathbf{x}_t$$

$$R_{T-\tau+1} = R_{T-\tau+2} + \sum_{i=1}^{\tau} r_{T-i+1}\sqrt{A_{\tau,i}\mathbf{y}_{T-i+1}^\top \mathbf{C}_\tau \mathbf{y}_{T-i+1}}$$

$$R_{T+1} = 0, A_{\tau,i} = \prod_{j=i}^{\tau-1} a_j,\ i = 1, \ldots, \tau-1, \textit{and } A_{\tau,i} = 1\ \ \forall i = \tau.$$

Proof We prove the above result by induction in a manner similar to the previous cases.

Base Case ($\tau = 1$): We must prove that the reformulation of the base case leads to

$$s_{T-1} + \Theta_T + r_T\sqrt{A_{1,1}\mathbf{y}_T^\top \boldsymbol{\Sigma}_T \mathbf{y}_T} + R_{T+1} \leq B.$$

The original constraint (C-RO) can be written as

$$s_{T-1} + \mathbf{d}_T^\top \mathbf{x}_T \leq B. \tag{23}$$

When substituting $\mathbf{d}_T = \boldsymbol{\mu}_T + \mathbf{L}_T\mathbf{u}_T$, the constraint must hold for all $\mathbf{u}_T$ such that $\|\mathbf{u}_T\|_2 \leq r_T$, and hence it also must hold for the following problem

$$s_{T-1} + \boldsymbol{\mu}_T^\top \mathbf{x}_T + r_T\|\mathbf{L}_T^\top \mathbf{x}_T\|_2 \leq B.$$

However, $\|\mathbf{L}_T^\top \mathbf{x}_T\|_2 = \sqrt{\mathbf{x}_T^\top \mathbf{L}_T \mathbf{L}_T^\top \mathbf{x}_T} = \sqrt{\mathbf{x}_T^\top \boldsymbol{\Sigma}_T \mathbf{x}_T}$. Using the fact that $\mathbf{y}_T = \mathbf{x}_T$, we obtain

$$s_{T-1} + \Theta_T + \|\mathbf{L}_T^\top \mathbf{y}_T\| \leq B \Leftrightarrow s_{T-1} + \Theta_T + \sqrt{\mathbf{y}_T^\top \boldsymbol{\Sigma}_T \mathbf{y}_T} \leq B,$$

and since $R_{T+1} = 0$, by assumption, we have the desired result.

Inductive case ($\tau = k$): By induction, the given result is true for $T-k+1$, i.e., the robust reformulation with respect to $\mathbf{d}_T, \mathbf{d}_{T-1}, \ldots \mathbf{d}_{T-k+1}$ is given by

$$s_{T-k} + \Theta_{T-k+1} + \sum_{i=1}^{k} r_{T-i+1}\sqrt{A_{k,i}\mathbf{y}_{T-i+1}^\top \boldsymbol{\Sigma}_{T-k+1}\mathbf{y}_{T-i+1}} + R_{T-k+2} \leq B.$$

We must prove that the reformulation with respect to $\mathbf{d}_{T-k}$, i.e., $\tau = k+1$ is given by

$$s_{T-k-1} + \Theta_{T-k} + \sum_{i=1}^{k+1} r_{T-i+1}\sqrt{A_{k+1,i}\mathbf{y}_{T-i+1}^\top \boldsymbol{\Sigma}_{T-k}\mathbf{y}_{T-i+1}} + R_{T-k+1} \leq B.$$

Substituting $s_{T-k} = s_{T-k-1} + \mathbf{d}_{T-k}^\top \mathbf{x}_{T-k}$ and $\mathbf{d}_{T-k} = \boldsymbol{\mu}_{T-k} + \mathbf{L}_{T-k}\mathbf{u}_{T-k}$ in the inductive assumption, we achieve

$$\begin{aligned}
&s_{T-k-1} + \Theta_{T-k+1} + \boldsymbol{\mu}_{T-k}^\top \mathbf{x}_{T-k} + \mathbf{u}_{T-k}^\top \mathbf{L}_{T-k}^\top \mathbf{x}_{T-k} \\
&+ \sum_{i=1}^{k} r_{T-i+1}\sqrt{A_{k,i}\mathbf{y}_{T-i+1}^\top \boldsymbol{\Sigma}_{T-k+1}\mathbf{y}_{T-i+1}} + R_{T-k+2} \leq B.
\end{aligned}$$

Since $\Theta_{T-k} = \boldsymbol{\mu}_{T-k}^\top \mathbf{x}_{T-k} + \Theta_{T-k+1}$, the above can be expressed as

$$s_{T-k-1} + \Theta_{T-k} + \mathbf{u}_{T-k}^\top \mathbf{L}_{T-k}^\top \mathbf{x}_{T-k} + \sum_{i=1}^{k} r_{T-i+1}\sqrt{A_{k,i}\mathbf{y}_{T-i+1}^\top \boldsymbol{\Sigma}_{T-k+1}\mathbf{y}_{T-i+1}} + R_{T-k+2} \leq B.$$

Utilizing the uncertainty dependence $\boldsymbol{\Sigma}_{T-k+1} = a_k \boldsymbol{\Sigma}_{T-k} + f_k(\mathbf{d}_{T-k} - \boldsymbol{\mu}_{T-k})(\mathbf{d}_{T-k} - \boldsymbol{\mu}_{T-k})^\top + \mathbf{C}_k$, the constraint can be reformulated to obtain

$$\begin{aligned}
&s_{T-k-1} + \Theta_{T-k} + \mathbf{u}_{T-k}^\top \mathbf{L}_{T-k}^\top \mathbf{x}_{T-k} \\
&+ \sum_{i=1}^{k} r_{T-i+1}\sqrt{A_{k,i}a_k\mathbf{y}_{T-i+1}^\top \boldsymbol{\Sigma}_{T-k}\mathbf{y}_{T-i+1} + A_{k,i}f_k((\mathbf{d}_{T-k} - \boldsymbol{\mu}_{T-k})^\top \mathbf{y}_{T-i+1})^2 + A_{k,i}\mathbf{y}_{T-i+1}^\top \mathbf{C}_k \mathbf{y}_{T-i+1}} \\
&+ R_{T-k+2} \leq B.
\end{aligned}$$

Using the Cauchy-Schwarz inequality $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$, a more conservative constraint can be written as

$$\begin{aligned}
&s_{T-k-1} + \Theta_{T-k} + \mathbf{u}_{T-k}^\top \mathbf{L}_{T-k}^\top \mathbf{x}_{T-k} + \sum_{i=1}^{k} r_{T-i+1}\sqrt{A_{k,i}a_k\mathbf{y}_{T-i+1}^\top \boldsymbol{\Sigma}_{T-k}\mathbf{y}_{T-i+1}} \\
&+ \sum_{i=1}^{k} r_{T-i+1}\sqrt{A_{k,i}f_k}(\mathbf{d}_{T-k} - \boldsymbol{\mu}_{T-k})^\top \mathbf{y}_{T-i+1} + \sum_{i=1}^{k} r_{T-i+1}\sqrt{A_{k,i}\mathbf{y}_{T-i+1}^\top \mathbf{C}_k \mathbf{y}_{T-i+1}} + R_{T-k+2} \leq B.
\end{aligned}$$

Using the fact that $\mathbf{d}_{T-k} - \boldsymbol{\mu}_{T-k} = \mathbf{L}_{T-k}\mathbf{u}_{T-k}$, and collecting the terms involving $\mathbf{u}_{T-k}^\top \mathbf{L}_{T-k}^\top$

$$\begin{aligned}
&s_{T-k-1} + \Theta_{T-k} + \mathbf{u}_{T-k}^\top \mathbf{L}_{T-k}^\top (\mathbf{x}_{T-k} + \sum_{i=1}^{k} r_{T-i+1}\sqrt{A_{k,i}f_k}\mathbf{y}_{T-i+1}) \\
&+ \sum_{i=1}^{k} r_{T-i+1}\sqrt{A_{k,i}a_k\mathbf{y}_{T-i+1}^\top \boldsymbol{\Sigma}_{T-k}\mathbf{y}_{T-i+1}} + \sum_{i=1}^{k} r_{T-i+1}\sqrt{A_{k,i}\mathbf{y}_{T-i+1}^\top \mathbf{C}_k \mathbf{y}_{T-i+1}} + R_{T-k+2} \leq B.
\end{aligned}$$

Since $\mathbf{y}_{T-k} = \mathbf{x}_{T-k} + \sum_{j=1}^{k} \mathbf{y}_{T-j+1} r_{T-j+1}\sqrt{f_k A_{k,j}}$, we can write

$$\begin{aligned}
s_{T-k-1} + \Theta_{T-k} &+ \mathbf{u}_{T-k}^\top \mathbf{L}_{T-k}^\top \mathbf{y}_{T-k} + \sum_{i=1}^{k} r_{T-i+1}\sqrt{A_{k,i}a_k\mathbf{y}_{T-i+1}^\top \boldsymbol{\Sigma}_{T-k}\mathbf{y}_{T-i+1}} \\
&+ \sum_{i=1}^{k} r_{T-i+1}\sqrt{A_{k,i}\mathbf{y}_{T-i+1}^\top \mathbf{C}_k \mathbf{y}_{T-i+1}} + R_{T-k+2} \leq B.
\end{aligned}$$

Since the constraint must hold for all $\mathbf{u}_{T-k}$ such that $\|\mathbf{u}_{T-k}\|_2 \leq r_{T-k}$, taking the maximum over $\mathbf{u}_{T-k}$, we obtain

$$s_{T-k-1} + \Theta_{T-k} + r_{T-k}\|\mathbf{L}_{T-k}^\top \mathbf{y}_{T-k}\|_2 + \sum_{i=1}^{k} r_{T-i+1}\sqrt{A_{k,i}a_k \mathbf{y}_{T-i+1}^\top \boldsymbol{\Sigma}_{T-k}\mathbf{y}_{T-i+1}}$$
$$+ \sum_{i=1}^{k} r_{T-i+1}\sqrt{A_{k,i}\mathbf{y}_{T-i+1}^\top \mathbf{C}_k \mathbf{y}_{T-i+1}} + R_{T-k+2} \leq B.$$

We know that $\|\mathbf{L}_{T-k}^\top \mathbf{y}_{T-k}\|_2 = \sqrt{\mathbf{y}_{T-k}^\top \boldsymbol{\Sigma}_{T-k}\mathbf{y}_{T-k}}$, which leads to

$$s_{T-k-1} + \Theta_{T-k} + r_{T-k}\sqrt{\mathbf{y}_{T-k}^\top \boldsymbol{\Sigma}_{T-k}\mathbf{y}_{T-k}} + \sum_{i=1}^{k} r_{T-i+1}\sqrt{A_{k,i}a_k \mathbf{y}_{T-i+1}^\top \boldsymbol{\Sigma}_{T-k}\mathbf{y}_{T-i+1}}$$
$$+ \sum_{i=1}^{k} r_{T-i+1}\sqrt{A_{k,i}\mathbf{y}_{T-i+1}^\top \mathbf{C}_k \mathbf{y}_{T-i+1}} + R_{T-k+2} \leq B. \tag{24}$$

Using $A_{k,i}a_k = \prod_{j=i}^{k-1} a_j a_k = \prod_{j=i}^{k} a_j = A_{k+1,i}$ and $A_{k+1,k+1} = 1$, we can rewrite (24) as

$$s_{T-k-1} + \Theta_{T-k} + r_{T-k}\sqrt{A_{k+1,k+1}\mathbf{y}_{T-k}^\top \boldsymbol{\Sigma}_{T-k}\mathbf{y}_{T-k}} + \sum_{i=1}^{k} r_{T-i+1}\sqrt{A_{k+1,i}\ \mathbf{y}_{T-i+1}^\top \boldsymbol{\Sigma}_{T-k}\mathbf{y}_{T-i+1}}$$
$$+ \sum_{i=1}^{k} r_{T-i+1}\sqrt{A_{k,i}\mathbf{y}_{T-i+1}^\top \mathbf{C}_k \mathbf{y}_{T-i+1}} + R_{T-k+2} \leq B.$$

Since $r_{T-k}\sqrt{A_{k+1,k+1}\mathbf{y}_{T-k}^\top \boldsymbol{\Sigma}_{T-k}\mathbf{y}_{T-k}} = r_{T-(k+1)+1}\sqrt{A_{k+1,k+1}\mathbf{y}_{T-(k+1)+1}^\top \boldsymbol{\Sigma}_{T-(k+1)+1}\mathbf{y}_{T-(k+1)+1}}$, we can combine this term into the summation as

$$s_{T-k-1} + \Theta_{T-k} + \sum_{i=1}^{k+1} r_{T-i+1}\sqrt{A_{k+1,i}\ \mathbf{y}_{T-i+1}^\top \boldsymbol{\Sigma}_{T-k}\mathbf{y}_{T-i+1}}$$
$$+ \sum_{i=1}^{k} r_{T-i+1}\sqrt{A_{k,i}\mathbf{y}_{T-i+1}^\top \mathbf{C}_k \mathbf{y}_{T-i+1}} + R_{T-k+2} \leq B.$$

Using the fact that $R_{T-k+1} = R_{T-k+2} + \sum_{i=1}^{k} r_{T-i+1}\sqrt{A_{k,i}\mathbf{y}_{T-i+1}^\top \mathbf{C}_k \mathbf{y}_{T-i+1}}$, we obtain the desired result.

$$s_{T-k-1} + \Theta_{T-k} + \sum_{i=1}^{k+1} r_{T-i+1}\sqrt{A_{k+1,i}\ \mathbf{y}_{T-i+1}^\top \boldsymbol{\Sigma}_{T-k}\mathbf{y}_{T-i+1}} + R_{T-k+1} \leq B.$$

□

The complete reformulation follows immediately by substituting by $\tau = T$ in Lemma 3, as in the following proposition.

Proposition 6 *A conservative robust reformulation of the constraint* (C-RO) *under the assumptions of Lemma 3 is given by*

$$\sum_{t=1}^{T} \boldsymbol{\mu}_t^\top \mathbf{x}_t + \sum_{i=1}^{T} r_{T-i+1}\sqrt{A_{T,i}}\|\mathbf{L}_1^\top \mathbf{y}_{T-i+1}\|_2 + R_2 \leq B,$$

where $\mathbf{y}_{T-\tau} = \mathbf{x}_{T-\tau} + \sum_{j=1}^{\tau} \mathbf{y}_{T-j+1} r_{T-j+1}\sqrt{f_\tau A_{\tau,j}}$*, and* $R_{T+1} = 0$*,* $A_{\tau,i} = \prod_{j=i}^{\tau-1} a_j$*,* $i = 1, \ldots, \tau - 1$*,* $A_{\tau,i} = 1\ \forall i = \tau$*, and* $R_{T-\tau+1} = R_{T-\tau+2} + \sum_{i=1}^{\tau} r_{T-i+1}\sqrt{A_{\tau,i}\mathbf{y}_{T-i+1}^\top \mathbf{C}_\tau \mathbf{y}_{T-i+1}}$.

3.4 Application: Knapsack Problem

Knapsack problems form an insightful test case since they arise in many common applications as a sub-problem. We consider a sequential knapsack problem with known objective but uncertain constraint coefficients. The problem spans two periods and consists of two uncertain weight realizations, which occur successively. In our setting, the second weight value depends on the first realized weight. These uncertain parameters reside in CU sets. Such a robust knapsack problem can be described by

$$\begin{aligned}\max_{\mathbf{x}_1,\mathbf{x}_2} \ & \mathbf{c}_1^\top \mathbf{x}_1 + \mathbf{c}_2^\top \mathbf{x}_2 \\ \text{s.t.} \ & \mathbf{d}_1^\top \mathbf{x}_1 + \mathbf{d}_2^\top \mathbf{x}_2 \le B \qquad \forall \mathbf{d}_2 \in \mathcal{U}_2(\mathbf{d}_1),\ \forall \mathbf{d}_1 \in \mathcal{U}_1 \\ & \mathbf{x}_1 \in \{0,1\}^{m_1},\ \mathbf{x}_2 \in \{0,1\}^{m_2}.\end{aligned}$$

The two sets of binary decisions $\mathbf{x}_1$ and $\mathbf{x}_2$ correspond to different periods with value and uncertain weight coefficients $\mathbf{c}_1, \mathbf{d}_1$ and $\mathbf{c}_2, \mathbf{d}_2$, respectively. However, both decisions are taken before either of the weights are realized, i.e., both variables $\mathbf{x}_1$ and $\mathbf{x}_2$ are here-and-now decisions. The uncertain parameters $\mathbf{d}_t, t = 1, 2$ are modeled as residing in ellipsoidal sets with given centers and covariance matrices. The uncertainty dependence arises by allowing the center of the second period ellipsoid to depend on the realization of the first stage weights.

It is assumed that weights are characterized by a normal distribution and that there is a normal prior for the mean. The covariance matrix of both the prior and the distribution are known in advance. For stage one, the covariance matrices are denoted by $\boldsymbol{\Sigma}_1$ for the prior and $\boldsymbol{\Sigma}$ for the main distribution with $\mathbf{L}_1{\mathbf{L}_1}^\top = \boldsymbol{\Sigma}_0 + \boldsymbol{\Sigma}$. Then the corresponding first stage uncertainty set is given by

$$\mathcal{U}_1 = \{\mathbf{d}_1 \mid \mathbf{d}_1 = \boldsymbol{\mu} + \mathbf{L}_1 \mathbf{u} \ : \ \|\mathbf{u}\|_2 \le r_1\}.$$

The first stage posterior serves as the second stage prior, and is obtained by updating the prior of the first stage (Bayesian update). Consequently, the center of the second stage is given by $\boldsymbol{\Sigma}_2 \left(\boldsymbol{\Sigma}^{-1}\mathbf{d}_1 + {\boldsymbol{\Sigma}_1}^{-1}\boldsymbol{\mu}\right)$ and the covariance matrix of the second stage prior $\boldsymbol{\Sigma}_2$ is given by $\boldsymbol{\Sigma}_2 = \left({\boldsymbol{\Sigma}_1}^{-1} + \boldsymbol{\Sigma}^{-1}\right)^{-1}$. Using the Cholesky decomposition $\mathbf{L}_2{\mathbf{L}_2}^\top = \boldsymbol{\Sigma}_2 + \boldsymbol{\Sigma}$, the second stage CU set is given by

$$\mathcal{U}_2(\mathbf{d}_1) = \left\{\mathbf{d}_2 \mid \mathbf{d}_2 = \boldsymbol{\Sigma}_2 \left(\boldsymbol{\Sigma}^{-1}\mathbf{d}_1 + \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}\right) + \mathbf{L}_2\mathbf{w},\ \ \|\mathbf{w}\|_2 \le r_2\right\}.$$

When uncertainty dependence does not occur, the non-adaptive second stage set is

$$\mathcal{U}_{2,\mathrm{NA}} = \{\mathbf{d}_2 \mid \mathbf{d}_2 = \boldsymbol{\mu} + \mathbf{L}_1\mathbf{w},\ \ \|\mathbf{w}\|_2 \le r_2\}.$$

Note that the difference between $\mathcal{U}_2(\mathbf{d}_1)$ and $\mathcal{U}_{2,\mathrm{NA}}$ is in their centers, where the latter is not updated according to the realization of $\mathbf{d}_1$. We now describe the experimental setting.

Numerical Experiments

In order to test the sensitivity of the models to the distribution of the underlying uncertainty, we conduct two experiments using different underlying distributions: normal and lognormal for the knapsack weights. The following set of modules is repeated for each distribution.

(i) Generate k estimates of the mean and covariance matrix of the true distribution, using l samples;
(ii) For each estimate, solve the knapsack problem for s different sizes of the uncertainty set;
(iii) For each of these solutions, generate n samples of uncertain $\mathbf{d}_1$ and $\mathbf{d}_2$ from the true distribution to measure the average performance of the solution.

We consider a problem with 20 items in each stage, i.e., $m_1 = m_2 = 20$. The experiment with a normal distribution has a mean of $10\mathbf{e}$ ($\mathbf{e}$ is a vector of all ones) and a randomly generated covariance matrix. These mean and covariance matrix also serve as parameters for the lognormal experiment.

In module (i), let $l = 25$ and $k = 25$. Sample averages are used to estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. The covariance matrix of the prior is $\boldsymbol{\Sigma}_1 = \frac{1}{25}\boldsymbol{\Sigma}$. The objective function coefficients $\mathbf{c}_1$ and $\mathbf{c}_2$ are uniformly sampled between on $[0, 10]$ and the budget (size of the knapsack) is 200.

In (ii), the experiment is conducted for $s = 20$ different uncertainty set sizes $r_1 = r_2 \in [0, 4]$.

In (iii), the average performance is measured using $n = 100$ different samples.

To probe each setting, the nominal problem is solved for both CU and non-adaptive (NA) sets. In order to estimate the average performance of these solutions, the uncertain weight vectors are sampled from the true distribution. The fraction of cases, in which the constraints are satisfied, is also recorded.

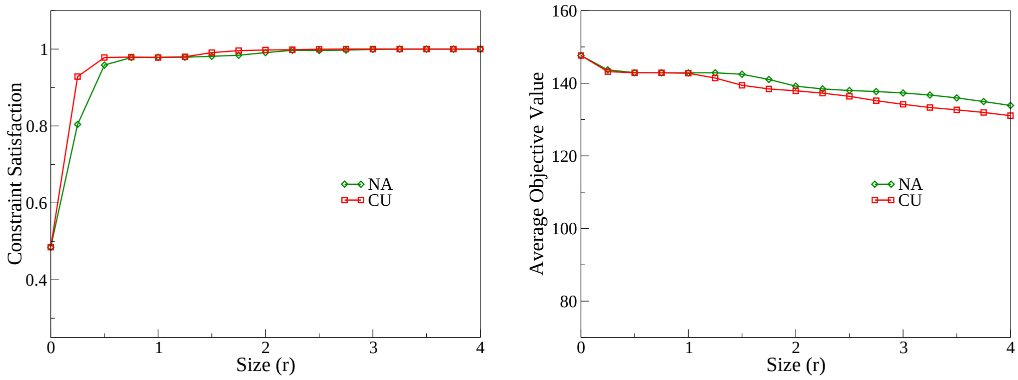

Fig. 4 For normal distribution, comparison of CU and non-adaptive (NA) sets for the robust knapsack problem at different set sizes: (left) the fraction of constraint satisfaction, and (right) the objective value.

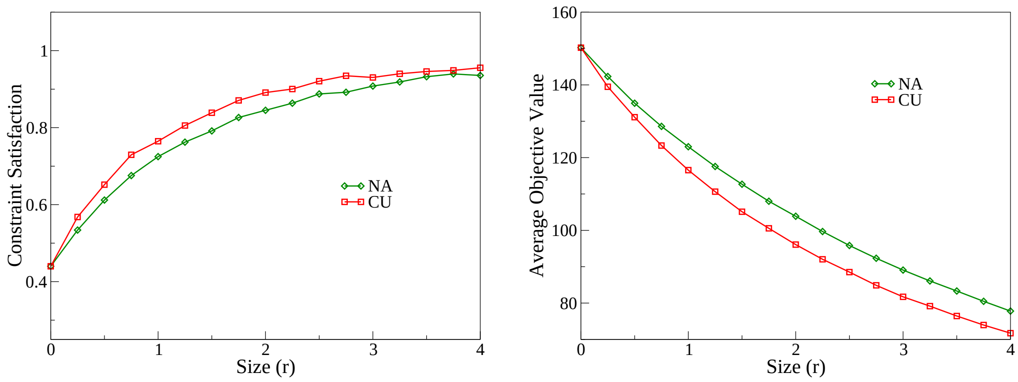

Fig. 5 For lognormal distribution, comparison of CU and NA for the robust knapsack problem at different uncertainty set sizes: (left) the fraction of constraint satisfaction, and (right) the objective value.

For Figures 4 and 5, the data points at $r_1 = r_2 = 0$ correspond to constraint satisfaction (left) and average objective value (right) of the nominal problem. In this setting, a low constraint satisfaction and high objective function value can be observed for all solutions. For larger sizes $r = r_1 = r_2 > 0$, we consolidate our observations as follows.

Effect of Size: For both solutions, the constraint satisfaction increases and the average objective value decreases with r.

Normal vs. Lognormal: The lognormal distribution is more variable then the normal. This leads to lower constraint satisfaction for the lognormal case. Since the normal distribution has ellipsoidal contours, the ellipsoidal uncertainty set is a better model for the normal than the lognormal distribution. We observe a rapid increase in constraint satisfaction for the normal distribution followed by a plateau for larger sizes. This increase is more gradual in the lognormal case.

CU vs. NA: CU solutions take into account the dependency on the first stage, which provides additional protection beyond the standard uncertainty set. This increases the constraint satisfaction for CU over NA, for both the normal and lognormal distribution. However, this also leads to lower objective function values as size increases.

4 Conclusion

Connected uncertainty sets provide a way to adapt the set to past observations. This paper extends distributionally robust optimization and standard robust optimization to the connected uncertainty paradigm by providing reformulations for commonly used constraints and uncertainty set dependencies. We provide reformulations of the constraint $\sum_{t=1}^{T} h_t(\mathbf{x}_t, \mathbf{d}_t) \leq B$ for moment based distributional uncertainty sets and reformulations of the constraint $\sum_{t=1}^{T} \mathbf{d}_t^\top \mathbf{x}_t \leq B$ for polyhedral and ellipsoidal uncertainty sets with linear and quadratic dependence. In multistage problems, uncertainties are naturally connected across periods and the proposed connected uncertainty approach offers a general modeling framework that can be applied to a broad range of applications.

References

1. Alwan, L.C., Xu, M., Yao, D.Q., Yue, X.: The dynamic newsvendor model with correlated demand. Decision Sci. **47**(1), 11–30 (2016)
2. Analui, B., Pflug, G.C.: On distributionally robust multiperiod stochastic optimization. Comput. Manage. Sci. **11**(3), 197–220 (2014)
3. Bandi, C., Bertsimas, D.: Tractable stochastic analysis in high dimensions via robust optimization. Math. Prog. **134**(1), 23–70 (2012)
4. Ben-Tal, A., El Ghaoui, L., Nemirovski, A.: Robust optimization. Princeton University Press (2009)
5. Ben-Tal, A., Goryashko, A., Guslitzer, E., Nemirovski, A.: Adjustable robust solutions of uncertain linear programs. Math. Prog. **99**(2), 351–376 (2004)
6. Bertsimas, D., Brown, D.B., Caramanis, C.: Theory and applications of robust optimization. SIAM Rev. **53**(3), 464–501 (2011)
7. Bertsimas, D., Goyal, V., Sun, X.A.: A geometric characterization of the power of finite adaptability in multistage stochastic and adaptive optimization. Math. Oper. Res. **36**(1), 24–54 (2011)
8. Bertsimas, D., Gupta, V., Kallus, N.: Robust SAA. arXiv preprint arXiv:1408.4445 (2014)
9. Bertsimas, D., Sim, M.: Robust discrete optimization and network flows. Math. Prog. **98**(1-3), 49–71 (2003)
10. Bertsimas, D., Sim, M., Zhang, M.: A practicable framework for distributionally robust linear optimization. Preprint (2014)
11. Bertsimas, D., Vayanos, P.: Data-driven learning in dynamic pricing using adaptive optimization. Preprint (2015)
12. Buishand, T.A., Brandsma, T.: Multisite simulation of daily precipitation and temperature in the rhine basin by nearest-neighbor resampling. Water Resour. Res. **37**(11), 2761–2776 (2001)
13. Calafiore, G.C., El Ghaoui, L.: On distributionally robust chance-constrained linear programs. J. Optim. Theory Appl. **130**(1), 1–22 (2006)
14. Chu, M., Zinchenko, Y., Henderson, S.G., Sharpe, M.B.: Robust optimization for intensity modulated radiation therapy treatment planning under uncertainty. Phys. Med. Biol. **50**(23), 5463 (2005)
15. De Queiroz, A.R., Morton, D.P.: Sharing cuts under aggregated forecasts when decomposing multi-stage stochastic programs. Oper. Res. Lett. **41**(3), 311–316 (2013)
16. Delage, E., Iancu, D.: Robust multi-stage decision making. INFORMS Tutor. Oper. Res. pp. 20–46 (2015)
17. Delage, E., Ye, Y.: Distributionally robust optimization under moment uncertainty with application to data-driven problems. Oper. Res. **58**(3), 595–612 (2010)
18. Goh, J., Sim, M.: Distributionally robust optimization and its tractable approximations. Oper. Res. **58**(4-part-1), 902–917 (2010)
19. Hanasusanto, G.A., Kuhn, D.: Conic programming reformulations of two-stage distributionally robust linear programs over wasserstein balls. arXiv preprint arXiv:1609.07505 (2016)
20. Hogg, R.V., McKean, J.: Introduction to mathematical Statistics. Prentice Hall (2005)
21. Infanger, G., Morton, D.P.: Cut sharing for multistage stochastic linear programs with interstage dependency. Math. Prog. **75**(2), 241–256 (1996)
22. Iyengar, G.N.: Robust dynamic programming. Math. Oper. Res. **30**(2), 257–280 (2005)
23. Jiang, R., Wang, J., Guan, Y.: Robust unit commitment with wind power and pumped storage hydro. IEEE Trans. Power Syst. **27**(2), 800–810 (2012)
24. Kuhn, D., Wiesemann, W., Georghiou, A.: Primal and dual linear decision rules in stochastic and robust optimization. Math. Prog. **130**(1), 177–209 (2011)
25. Lam, H.: Robust sensitivity analysis for stochastic systems. Math. Oper. Res. **41**(4), 1248–1275 (2016)
26. Lewellen, J.: Momentum and autocorrelation in stock returns. Rev. Financ. Stud. **15**(2), 533–564 (2002)
27. Lorca, A., Sun, X.A.: Adaptive robust optimization with dynamic uncertainty sets for multi-period economic dispatch under significant wind. IEEE Trans. Power Syst. **30**(4), 1702–1713 (2015)
28. Lorca, A., Sun, X.A., Litvinov, E., Zheng, T.: Multistage adaptive robust optimization for the unit commitment problem. Oper. Res. **64**(1), 32–51 (2016)

29. Mehrotra, S., Papp, D.: A cutting surface algorithm for semi-infinite convex programming with an application to moment robust optimization. SIAM J. Optim. **24**(4), 1670–1697 (2014)
30. Mehrotra, S., Zhang, H.: Models and algorithms for distributionally robust least squares problems. Math. Prog. **146**(1-2), 123–141 (2014)
31. Nilim, A., El Ghaoui, L.: Robust control of markov decision processes with uncertain transition matrices. Oper. Res. **53**(5), 780–798 (2005)
32. Nohadani, O., Roy, A.: Robust optimization with time-dependent uncertainty in radiation therapy. submitted (2016)
33. Scarf, H., Arrow, K., Karlin, S.: A min-max solution of an inventory problem. Studies Math. theory Inv. Prod. **10**, 201–209 (1958)
34. Shapiro, A. (2001) On duality theory of conic linear problems. In: Semi-infinite programming, Springer, pp. 135–165
35. Shapiro, A.: Minimax and risk averse multistage stochastic programming. Eur. J. Oper. Res. **219**(3), 719–726 (2012)
36. Shapiro, A.: Rectangular sets of probability measures. Oper. Res. **64**(2), 528–541 (2016)
37. Shapiro, A., Dentcheva, D., Ruszczyński, A.: Lectures on stochastic programming: modeling and theory, vol. 16. SIAM (2014)
38. Tsay, R.S.: Multivariate Time Series Analysis: with R and financial applications. John Wiley & Sons (2013)
39. Wiesemann, W., Kuhn, D., Sim, M.: Distributionally robust convex optimization. Oper. Res. **62**(6), 1358–1376 (2014)
40. Xin, L., Goldberg, D.A.: Distributionally robust inventory control when demand is a martingale. arXiv preprint arXiv:1511.09437 (2015)
41. Xu, H., Mannor, S. (2010) Distributionally robust markov decision processes. In: Adv. Neural Inf. Process. Syst., pp. 2505–2513
42. Yahoo Finance!: (2017). URL https://finance.yahoo.com
43. Zhao, L., Zeng, B. (2012) Robust unit commitment problem with demand response and wind energy. In: 2012 IEEE Power Energy Soc. Gen. Meet., IEEE, pp. 1–8