Electronic Companion to "Optimization under Connected Uncertainty"

1. Proof of Theorem 3

The following is the proof for the robust counterpart for the polyhedral uncertainty set.

Proof: For (P), we define the joint uncertainty set $\mathcal{U}$ as

$$\begin{aligned}\mathcal{U} &= \left\{(\mathbf{d}_1^\top, \mathbf{d}_2^\top, \ldots, \mathbf{d}_T^\top)^\top \mid \mathbf{d}_1 \in \mathcal{U}_1,\ \mathbf{d}_t \in \mathcal{U}_t(\mathbf{d}_{t-1})\ \forall t = 2, \ldots, T\right\} \\ &= \left\{(\mathbf{d}_1^\top, \mathbf{d}_2^\top, \ldots, \mathbf{d}_T^\top)^\top \mid \mathbf{G}_1\mathbf{d}_1 \geq \mathbf{g}_1,\ \mathbf{G}_t\mathbf{d}_t \geq \mathbf{g}_t + \boldsymbol{\Delta}_t\mathbf{d}_{t-1}\ \forall t = 2, \ldots, T\right\},\end{aligned}$$

combining the polyhedral sets for each period. The robust counterpart of (C-RO) becomes

$$\sum_{t=1}^{T} \mathbf{d}_t^\top \mathbf{x}_t \leq B\ \forall (\mathbf{d}_1^\top, \mathbf{d}_2^\top, \ldots, \mathbf{d}_T^\top)^\top \in \mathcal{U},$$

which can be rewritten as

$$\max_{(\mathbf{d}_1^\top, \mathbf{d}_2^\top, \ldots, \mathbf{d}_T^\top)^\top \in \mathcal{U}} \sum_{t=1}^{T} \mathbf{d}_t^\top \mathbf{x}_t \leq B.$$

The LHS is computed by

$$\begin{aligned}\max_{\mathbf{d}_t}\ & \sum_{t=1}^{T} \mathbf{d}_t^\top \mathbf{x}_t \\ \text{s.t.}\ & \mathbf{G}_1\mathbf{d}_1 \geq \mathbf{g}_1 \\ & \mathbf{G}_t\mathbf{d}_t \geq \mathbf{g}_t + \boldsymbol{\Delta}_t\mathbf{d}_{t-1} \qquad \forall t = 2, \ldots, T.\end{aligned}$$

Using duality, the robust counterpart of (C-RO) is

$$\begin{aligned}\sum_{t=1}^{T} \mathbf{q}_t^\top \mathbf{b}_t &\leq B \\ \mathbf{q}_t^\top \mathbf{A}_t - \mathbf{q}_{t+1}^\top \boldsymbol{\Delta}_{t+1} &= \mathbf{x}_t^\top, \quad \forall t = 1, \ldots, T \\ \mathbf{q}_t &\leq 0 \quad \forall t = 1, \ldots, T,\end{aligned}$$

which is the desired result. $\square$

2. ARO with Ellipsoidal CU Sets for Connected Uncertainty with RO

In this section, we reformulate the connected constraints from problem (CU-ARO) for ellipsoidal uncertainty sets, where the center depends on the previous period realization as

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

where $\mathbf{L}_t\mathbf{L}_t^\top = \boldsymbol{\Sigma}_t$. The dependence of $\boldsymbol{\mu}_t$ on the previous period realization is given by

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

For this setting, the following theorem provides the reformulation for CU sets with affine decision policies.

THEOREM 1. *The two-period adjustable optimization problem* (CU-ARO) *has a tractable reformulation, when the uncertainty affects the center and the fully adaptive decisions are replaced by affine decision rules.*

Proof: We replace $\mathbf{x}_2(\mathbf{d}_1)$ with the affine decision rule $\mathbf{x}_2(\mathbf{d}_1) = \mathbf{X}_2\mathbf{d}_1$ and expand $\mathbf{b}_1(\mathbf{d}_1) = \mathbf{B}_1\mathbf{d}_1$ and $\mathbf{b}_2(\mathbf{d}_2) = \mathbf{B}_2\mathbf{d}_2$. We focus on reformulating the second constraint, which is affected by the connected uncertainty and whose robust problem is

$$\begin{aligned}
\max_{\mathbf{d}_1,\mathbf{d}_2}\ & \mathbf{B}_{2,i}^\top\mathbf{d}_2 - [\mathbf{A}_{22}\mathbf{X}_2]_i^\top\mathbf{d}_1 \\
\text{s.t.}\ & \mathbf{d}_1 = \boldsymbol{\mu}_1 + \mathbf{L}_1\mathbf{u}_1 \\
& \mathbf{d}_2 = \boldsymbol{\mu}_2(\mathbf{d}_1) + \mathbf{L}_2\mathbf{u}_2 \\
& \|\mathbf{u}_t\|_2 \le r_t \quad \forall t = 1,2.
\end{aligned}$$

Substituting $\mathbf{d}_1$ and $\mathbf{d}_2$, we rewrite the problem as

$$\begin{aligned}
\max_{\mathbf{d}_1,\mathbf{d}_2}\ & \mathbf{B}_{2,i}^\top(\mathbf{A}_2\boldsymbol{\mu}_1 + \mathbf{F}_2\mathbf{d}_1 + \mathbf{c}_2) + \mathbf{B}_{2,i}^\top\mathbf{L}_2\mathbf{u}_2 - [\mathbf{A}_{22}\mathbf{X}_2]_i^\top\boldsymbol{\mu}_1 - [\mathbf{A}_{22}\mathbf{X}_2]_i^\top\mathbf{L}_1\mathbf{u}_1 \\
\text{s.t.}\ & \|\mathbf{u}_t\|_2 \le r_t \quad \forall t = 1,2.
\end{aligned}$$

Using the dual, the complete second constraint of (CU-ARO) is given by

$$\begin{aligned}
\mathbf{A}_{21,i}^\top\mathbf{x}_1 \ge\ & \mathbf{B}_{2,i}^\top(\mathbf{A}_2\boldsymbol{\mu}_1 + \mathbf{F}_2\boldsymbol{\mu}_1 + \mathbf{c}_2) - [\mathbf{A}_{22}\mathbf{X}_2]_i^\top\boldsymbol{\mu}_1 + r_2\|\mathbf{B}_{2,i}^\top\mathbf{L}_2\|_2 + r_1\|\mathbf{B}_{2,i}^\top\mathbf{F}_2\mathbf{L}_1 \\
& - [\mathbf{A}_{22}\mathbf{X}_2]_i^\top\mathbf{L}_1\|_2.
\end{aligned}$$

The remaining constraints in (CU-ARO) is then reformulated in a similar manner, leading to a tractable reformulation. $\square$

3. Extensions to Affinely adaptive Decisions for Connected Uncertainty with RO

In previous sections, we focused on the effect of connected uncertainties on *static* and here-and-now decisions. In many problems, decisions can also adapt to previously revealed uncertainties. To this end, ARO has applications in inventory management (Solyalı et al. 2015) and unit commitment (Lorca and Sun 2017), amongst others. ARO problems, however, are known to be NP-complete, which can be circumvented by leveraging affine or piece-wise static decision rules (Hanasusanto et al. 2015).

In this section, we extend the notion of modeling with CU sets to *adaptive* decisions and provide reformulations for a two-period ARO problem under affine decision rules. Such a problem can be expressed as

$$\begin{aligned}
\min_{\substack{\mathbf{x}_1\\ \mathbf{x}_2(\mathbf{d}_1)}} \; \max_{\substack{\mathbf{d}_1\in\mathcal{U}_1\\ \mathbf{d}_2\in\mathcal{U}_2(\mathbf{d}_1)}} \quad & \mathbf{c}_1^\top\mathbf{x}_1+\mathbf{c}_2^\top\mathbf{x}_2(\mathbf{d}_1) \\
\text{s.t.} \quad & \mathbf{A}_{11}\mathbf{x}_1\ge\mathbf{b}_1(\mathbf{d}_1) && \forall\mathbf{d}_1\in\mathcal{U}_1 \\
& \mathbf{A}_{21}\mathbf{x}_1+\mathbf{A}_{22}\mathbf{x}_2(\mathbf{d}_1)\ge\mathbf{b}_2(\mathbf{d}_2) && \forall\mathbf{d}_2\in\mathcal{U}_2(\mathbf{d}_1)\,\forall\mathbf{d}_1\in\mathcal{U}_1 \\
& \mathbf{x}_2(\mathbf{d}_1)\ge\mathbf{0} && \forall\mathbf{d}_1\in\mathcal{U}_1 \\
& \mathbf{x}_1\ge\mathbf{0}.
\end{aligned} \tag{CU-ARO}$$

We focus on polyhedral CU sets, where the RHS depends on the previous period realization as

$$\mathcal{U}_1=\{\mathbf{d}_1\mid\mathbf{G}_1\mathbf{d}_1\ge\mathbf{g}_1\},\qquad \mathcal{U}_2(\mathbf{d}_1)=\{\mathbf{d}_2\mid\mathbf{G}_2\mathbf{d}_2\ge\mathbf{g}_2+\mathbf{\Delta}_1\mathbf{d}_1\}.$$

For this setting, the following theorem shows the impact of CU sets with affine decision policies on the fully adaptive decisions $\mathbf{x}_2(\mathbf{d}_1)=\mathbf{x}_2^0+\mathbf{X}_2\mathbf{d}_1$. For brevity, we incorporated the constant term $\mathbf{x}_2^0$ into the matrix $\mathbf{X}_2$ by modifying the uncertainty set and setting the first component of $\mathbf{d}_1$ to be equal to 1. Then we can write $\mathbf{x}_2(\mathbf{d}_1)=\mathbf{X}_2\mathbf{d}_1$.

THEOREM 2. *The two-period ARO problem* (CU-ARO) *has a tractable reformulation, when the uncertainty affects the RHS linearly and the variables are replaced by affine decision rules.*

Proof: We replace $\mathbf{x}_2(\mathbf{d}_1)$ with the affine decision rule $\mathbf{x}_2(\mathbf{d}_1)=\mathbf{X}_2\mathbf{d}_1$ and expand $\mathbf{b}_1(\mathbf{d}_1)=\mathbf{B}_1\mathbf{d}_1$ and $\mathbf{b}_2(\mathbf{d}_2)=\mathbf{B}_2\mathbf{d}_2$. We focus on reformulating the second constraint which is affected by the connected uncertainty and whose robust problem is

$$\begin{aligned}
\max_{\mathbf{d}_1,\mathbf{d}_2}\; & \mathbf{B}_{2,i}^\top\mathbf{d}_2-[\mathbf{A}_{22}\mathbf{X}_2]_i^\top\mathbf{d}_1 \\
\text{s.t.}\; & \mathbf{G}_1\mathbf{d}_1\ge\mathbf{g}_1 \\
& \mathbf{G}_2\mathbf{d}_2\ge\mathbf{g}_2+\Delta_1\mathbf{d}_1.
\end{aligned}$$

Using the dual, the complete second constraint of (CU-ARO) is given by

$$\begin{aligned}
\mathbf{A}_{21}\mathbf{x}_1 &\ge \mathbf{P}^\top\mathbf{g}_1+\mathbf{Q}^\top\mathbf{g}_2 \\
\mathbf{P}^\top\mathbf{G}_1+\mathbf{Q}^\top\mathbf{\Delta}_1 &= \mathbf{B}_2 \\
\mathbf{Q}^\top\mathbf{G}_2 &= -\mathbf{A}_{22}\mathbf{X}_2 \\
\mathbf{P},\mathbf{Q} &\le \mathbf{0}.
\end{aligned}$$

The columns of $\mathbf{P}$ and $\mathbf{Q}$ correspond to dual variables of the original problem. The remaining constraints in (CU-ARO) can be reformulated similarly, leading to a tractable reformulation. ☐

Theorem 2 extends the result of modeling with standard sets for an adaptive setting, where such reformulations are possible for RHS uncertainty, to CU sets. When uncertainties affect the LHS, even problems with standard and non-connected sets are difficult to reformulate (Ben-Tal et al. 2009). Therefore, we forego their extension to the CU setting.

4. Proof of Theorem 2

Proof: The reformulation proceeds through induction. The robust counterpart of (C-RO) with respect to $\mathbf{d}_{k+1},\ldots,\mathbf{d}_T$ is

$$s_k + \Theta_{k+1} + \sum_{t=k+1}^{T} r_t\sqrt{A_{k+1,t}\mathbf{y}_t^\top \mathbf{\Sigma}_{k+1}\mathbf{y}_t} + R_{k+1} \le B \quad \forall \mathbf{n}_{[k+1]} \in \mathcal{N}_{[k+1]},$$

$$\text{with } s_k = \sum_{t=1}^{k} \mathbf{d}_t^\top \mathbf{x}_t,\ \ \Theta_{k+1} = \sum_{t=k+1}^{T} \boldsymbol{\mu}_t^\top \mathbf{x}_t \text{ and } s_0 = 0.$$

The proof proceeds parallel to that of Theorem 1.

Base Case: The constraint (C-RO) can be written as $s_{T-1} + \mathbf{d}_T^\top \mathbf{x}_T \le B$. Substituting $\mathbf{d}_T = \boldsymbol{\mu}_T + \mathbf{L}_T\mathbf{u}_T$. (C-RO) must hold for all $\mathbf{u}_T$ with $\|\mathbf{u}_T\|_2 \le r_T$, and thus it also holds for $s_{T-1} + \boldsymbol{\mu}_T^\top \mathbf{x}_T + r_T\|\mathbf{L}_T^\top \mathbf{x}_T\|_2 \le B$. Using the equality $\|\mathbf{L}_T^\top \mathbf{x}_T\|_2 = \sqrt{\mathbf{x}_T^\top \mathbf{\Sigma}_T \mathbf{x}_T}$ and $\mathbf{y}_T = \mathbf{x}_T$, we obtain

$$s_{T-1} + \Theta_T + r_T\|\mathbf{L}_T^\top \mathbf{y}_T\|_2 \le B \Leftrightarrow s_{T-1} + \Theta_T + r_T\sqrt{\mathbf{y}_T^\top \mathbf{\Sigma}_T \mathbf{y}_T} \le B.$$

Since R_{T+1} is assumed to be zero, we have achieved the desired result.

Inductive case: Assume that the result is true for $k+1$, i.e., the reformulation with respect to $\mathbf{d}_T, \mathbf{d}_{T-1}, \ldots \mathbf{d}_{k+1}$ is given by

$$s_k + \Theta_{k+1} + \sum_{t=k+1}^{T} r_t\sqrt{A_{k+1,t}\mathbf{y}_t^\top \mathbf{\Sigma}_{k+1}\mathbf{y}_t} + R_{k+1} \le B \quad \forall \mathbf{n}_{[k+1]} \in \mathcal{N}_{[k+1]}.$$

Now, we have to prove that the reformulation with respect to $\mathbf{d}_k$ is given by

$$s_{k-1} + \Theta_k + \sum_{t=k}^{T} r_t\sqrt{A_{k,t}\mathbf{y}_t^\top \mathbf{\Sigma}_k \mathbf{y}_t} + R_k \le B \quad \forall \mathbf{n}_{[k]} \in \mathcal{N}_{[k]}.$$

Substituting $s_k = s_{k-1} + \mathbf{d}_k^\top \mathbf{x}_k$ and $\mathbf{d}_k = \boldsymbol{\mu}_k + \mathbf{L}_k\mathbf{u}_k$ in the inductive assumption

$$\begin{aligned} &s_{k-1} + \Theta_{k+1} + \boldsymbol{\mu}_k^\top \mathbf{x}_k + \mathbf{u}_k^\top \mathbf{L}_k^\top \mathbf{x}_k \\ &\quad + \sum_{t=k+1}^{T} r_t\sqrt{A_{k+1,t}\mathbf{y}_t^\top \mathbf{\Sigma}_{k+1}\mathbf{y}_t} + R_{k+1} \le B\ \forall \mathbf{n}_{[k+1]} \in \mathcal{N}_{[k+1]}. \end{aligned}$$

Utilizing (5) for $t = k$ and $\Theta_k = \boldsymbol{\mu}_k^\top \mathbf{x}_k + \Theta_{k+1}$, the constraint can be reformulated as

$$\begin{aligned} &s_{k-1} + \Theta_k + \mathbf{u}_k^\top \mathbf{L}_k^\top \mathbf{x}_k \\ &+ \sum_{t=k+1}^{T} r_t\Big(A_{k+1,t}a_k\mathbf{y}_t^\top \mathbf{\Sigma}_k\mathbf{y}_t + A_{k+1,t}f_k((\mathbf{d}_k - \boldsymbol{\mu}_k)^\top \mathbf{y}_t)^2 \\ &+ A_{k+1,t}\mathbf{y}_t^\top \mathbf{C}_k\mathbf{y}_t\Big)^{\frac{1}{2}} + R_{k+1} \le B\ \forall \mathbf{n}_{[k+1]} \in \mathcal{N}_{[k+1]}. \end{aligned}$$

Using the inequality, $\sqrt{a^2+b^2} \leq |a|+|b|$, a more conservative constraint is

$$\begin{aligned}
&s_{k-1}+\Theta_k+\mathbf{u}_k^\top \mathbf{L}_k^\top \mathbf{x}_k+\sum_{t=k+1}^T r_t \sqrt{A_{k+1,t} a_k \mathbf{y}_t^\top \boldsymbol{\Sigma}_k \mathbf{y}_t} \\
&+\sum_{t=k+1}^T r_t \sqrt{A_{k+1,t} f_k}\left|(\mathbf{d}_k-\boldsymbol{\mu}_k)^\top \mathbf{y}_t\right| \\
&+\sum_{t=k+1}^T r_t \sqrt{A_{k+1,t} \mathbf{y}_t^\top \mathbf{C}_k \mathbf{y}_t}+R_{k+1} \leq B \ \forall \mathbf{n}_{[k+1]} \in \mathcal{N}_{[k+1]}.
\end{aligned}$$

The above constraint holds for $\left|(\mathbf{d}_k-\boldsymbol{\mu}_k)^\top \mathbf{y}_t\right|$, if it holds for both the positive and the negative possible values. We use $n_{k,t}$, which can take values 1 or -1 to express the above as

$$\begin{aligned}
&s_{k-1}+\Theta_k+\mathbf{u}_k^\top \mathbf{L}_k^\top \mathbf{x}_k+\sum_{t=k+1}^T r_t \sqrt{A_{k+1,t} a_k \mathbf{y}_t^\top \boldsymbol{\Sigma}_k \mathbf{y}_t} \\
&+\sum_{t=k+1}^T n_{k,t} \cdot r_t \sqrt{A_{k+1,t} f_k}(\mathbf{d}_k-\boldsymbol{\mu}_k)^\top \mathbf{y}_t \\
&+\sum_{t=k+1}^T r_t \sqrt{A_{k+1,t} \mathbf{y}_t^\top \mathbf{C}_k \mathbf{y}_t}+R_{k+1} \leq B \\
&\qquad\qquad\qquad \forall n_{k,t} \in \{1,-1\},\ \forall t=k+1,\dots,T,\ \forall \mathbf{n}_{[k+1]} \in \mathcal{N}_{[k+1]}.
\end{aligned}$$

Note that the use of $n_{k,t}$ allows to write the 2^{T-k} constraints in this concise form. Combining the set $\mathcal{N}_{[k+1]}$ with $\{1,-1\}^{T-k}$, using $\mathbf{d}_k-\boldsymbol{\mu}_k=\mathbf{L}_k \mathbf{u}_k$, and collecting the terms involving $\mathbf{u}_k^\top \mathbf{L}_k^\top$, we obtain

$$\begin{aligned}
&s_{k-1}+\Theta_k+\mathbf{u}_k^\top \mathbf{L}_k^\top (\mathbf{x}_k+\sum_{t=k+1}^T n_{k,t} \cdot r_t \sqrt{A_{k+1,t} f_k} \mathbf{y}_t) \\
&+\sum_{t=k+1}^T r_t \sqrt{A_{k+1,t} a_k \mathbf{y}_t^\top \boldsymbol{\Sigma}_k \mathbf{y}_t}+\sum_{t=k+1}^T r_t \sqrt{A_{k+1,t} \mathbf{y}_t^\top \mathbf{C}_k \mathbf{y}_t} \\
&+R_{k+1} \leq B \ \forall \mathbf{n}_{[k]} \in \mathcal{N}_{[k]}.
\end{aligned}$$

With the definition of $\mathbf{y}_k$, we can write

$$\begin{aligned}
s_{k-1}+\Theta_k\ &+\mathbf{u}_k^\top \mathbf{L}_k^\top \mathbf{y}_k+\sum_{t=k+1}^T r_t \sqrt{A_{k+1,t} a_k \mathbf{y}_t^\top \boldsymbol{\Sigma}_k \mathbf{y}_t} \\
&+\sum_{t=k+1}^T r_t \sqrt{A_{k+1,t} \mathbf{y}_t^\top \mathbf{C}_k \mathbf{y}_t}+R_{k+1} \leq B \ \forall \mathbf{n}_{[k]} \in \mathcal{N}_{[k]}.
\end{aligned}$$

Since this constraint must hold for all $\mathbf{u}_k$ with $\|\mathbf{u}_k\|_2 \le r_k$, taking the maximum over $\mathbf{u}_k$ and substituting $\|\mathbf{L}_k^\top \mathbf{y}_k\|_2 = \sqrt{\mathbf{y}_k^\top \boldsymbol{\Sigma}_k \mathbf{y}_k}$, we obtain

$$\begin{aligned} & s_{k-1} + \Theta_k + r_k \sqrt{\mathbf{y}_k^\top \boldsymbol{\Sigma}_k \mathbf{y}_k} \\ & + \sum_{t=k+1}^{T} r_t \sqrt{A_{k+1,t} a_k \mathbf{y}_t^\top \boldsymbol{\Sigma}_k \mathbf{y}_t} \\ & + \sum_{t=k+1}^{T} r_t \sqrt{A_{k+1,t} \mathbf{y}_t^\top \mathbf{C}_k \mathbf{y}_t} + R_{k+1} \le B \ \forall \mathbf{n}_{[k]} \in \mathcal{N}_{[k]}. \end{aligned} \tag{3}$$

Using $A_{k+1,t} a_k = \prod_{j=k+1}^{t} a_j a_k = A_{k,t}$, and the fact that $A_{k,k} = 1$ we can rewrite (3) as

$$\begin{aligned} & s_{k-1} + \Theta_k + \sum_{t=k}^{T} r_t \sqrt{A_{k,t}\, \mathbf{y}_t^\top \boldsymbol{\Sigma}_k \mathbf{y}_t} \\ & \quad + \sum_{t=k+1}^{T} r_t \sqrt{A_{k+1,t} \mathbf{y}_t^\top \mathbf{C}_k \mathbf{y}_t} + R_{k+1} \le B \ \forall \mathbf{n}_{[k]} \in \mathcal{N}_{[k]}, \end{aligned}$$

by extending the first summation. Using the definition of R_k, we obtain the desired result,

$$s_{k-1} + \Theta_k + \sum_{t=k}^{T} r_t \sqrt{A_{k,t}\, \mathbf{y}_t^\top \boldsymbol{\Sigma}_k \mathbf{y}_t} + R_k \le B \ \forall \mathbf{n}_{[k]} \in \mathcal{N}_{[k]}.$$

The complete reformulation follows immediately by substituting $k = 1$. □

5. Robust Counterpart of (C-DRO)

The following proposition better illustrates the connection between constraint (C-DRO) and its robust counterpart (6). Given the uncertainty sets $\widetilde{\mathcal{U}}_1, \dots, \widetilde{\mathcal{U}}_T^{T-1}$, their joint uncertainty set is defined by

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

The joint set, $\widetilde{\mathcal{U}}$ is the set of all distributions P with the marginals lying in the specified uncertainty sets $\widetilde{\mathcal{U}}_t(\mathbf{d}_{t-1})$.

Proposition 1. *Given the sets $\widetilde{\mathcal{U}}_1, \dots, \widetilde{\mathcal{U}}_T(\mathbf{d}_{T-1})$ and their joint uncertainty set $\widetilde{\mathcal{U}}$, the robust counterpart of constraint* (C-DRO)*, given by*

$$\sup_{P \in \widetilde{\mathcal{U}}} \mathbb{E}_P \left[\sum_{t=1}^{T} h_t(\mathbf{x}_t, \mathbf{d}_t) \right] \le B \tag{4}$$

is equivalent to

$$\begin{aligned} \sup_{P_1 \in \widetilde{\mathcal{U}}_1} \mathbb{E}_{P_1} \Bigg[h_1(\mathbf{x}_1, \mathbf{d}_1) + \sup_{P_{2|1} \in \widetilde{\mathcal{U}}_2(\mathbf{d}_1)} \Bigg\{ \mathbb{E}_{P_{2|1}} \Bigg[h_2(\mathbf{x}_2, \mathbf{d}_2) + \dots \\ + \sup_{P_{T|T-1} \in \widetilde{\mathcal{U}}_T(\mathbf{d}_{T-1})} \left\{ \mathbb{E}_{P_{T|T-1}} \left[h_T(\mathbf{x}_T, \mathbf{d}_T) \right] \right\} \Bigg] \Bigg\} \Bigg] \le B. \end{aligned}$$

Proof: We first show the forward direction of the equivalence. Observe that for any distribution $P \in \widetilde{\mathcal{U}}$ with the corresponding marginal distribution P_1, and for each $\mathbf{d}_{t-1}$ with the conditional distribution $P_{t|t-1}$, it holds that

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

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

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

That means the LHS of (6) is ϵ-optimal to the LHS of (4). Since P^* is in $\widetilde{\mathcal{U}}$, there exist marginal distributions $P_1^*,\dots,P_{T|T-1}^*$ such that 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 $t=2,\dots,T$. Using (5), this means

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

Now for the opposite side of the inequality, let $P_1^*,P_{2|1}^*$ up to $P_{T|T-1}^*$ be ϵ-optimal to (6). The constraints of (6) ensure that $P_1^*\in\widetilde{\mathcal{U}}_1$ and $P_{t|t-1}^*\in\widetilde{\mathcal{U}}_t(\mathbf{d}_{t-1})$. Since $\widetilde{\mathcal{U}}$ is the set of all joint distributions with these specified marginals, 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

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

This gives the opposite inequality, and the result follows by letting ϵ shrink to zero. $\square$

Note that this result does not depend on the structure of the moment based uncertainty set and can be extended to sets $\widetilde{\mathcal{U}}_1, \dots, \widetilde{\mathcal{U}}_T(\mathbf{d}_{T-1})$ of any structure, in which the parameters depend on previous realizations. An important part of this result is the additive nature of the constraint.

6. RO Application: Knapsack Problem with Negative Correlations

We repeat the experiments conducted for the Knapsack problem with a negative correlation among the weights over time i.e., $\mathbf{\Psi} = -0.2 \times \mathbf{I}$. Figure 1 presents the value of average objective for any level of constraint satisfaction (left) and the number of non-zero allocations vs. set size. Figure 2 displays how the average constraint satisfaction (left) and the average objective value (right) change with size of the uncertainty set r. These results show that when uncertainties are negatively correlated, CU solutions achieve a better objective value but a lower constraint satisfaction than NC solutions. However, for any level of constraint satisfaction both have similar performance.

- ***Effect of Uncertainty Set Size***: For both models, constraint satisfaction increases and average objective value decreases with r. Note that the objective value is only measured, if constraints are satisfied.
- ***CU vs. NC***: For any r, CU solutions have lower constraint satisfaction than NC solutions. CU solutions also have higher average objective value. This is because connected sets depend on the first period, and the negative correlation instead of magnifying the worst case causes one of the realizations to take a lower value. This reduces the protection but increases the objective value.

  Note that for any level of satisfaction, the average objective of CU is almost the same as NC (see Figure 2 left).
- ***First vs Second period solutions***: For a single estimate in (ii) shown in Figure 2 (right), the optimal solution gradually concentrates only on $\mathbf{x}_2$ for both CU and NC as r increases. For NC, this is because $\mathbf{c}_2$ tends to be higher. For CU, the negative correlation prevents the magnification of the weights for both periods as such second period allocations which have higher objective coefficients can be selected.

7. DRO Application: Portfolio Optimization

We study a practical portfolio optimization problem on historical stock data. For our portfolio, we choose among 5 stocks. The experiment is repeated 150 times for randomly selected dates. At each date, we compute the weekly returns for the previous 100 weeks based on stock price data and fit them to a time series model. The two weeks following the selected dates serve to evaluate the performance of the model.

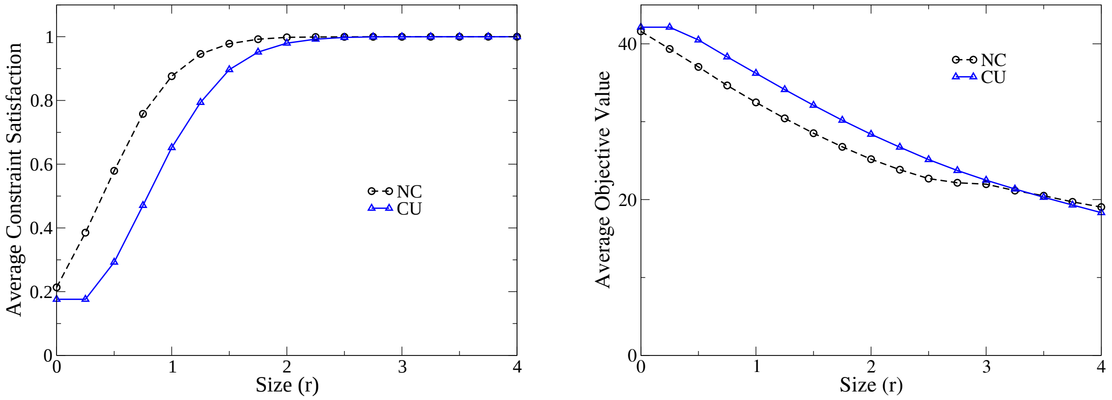

Figure 1 Comparison of connected and non-connected sets for the robust knapsack problem at different set sizes: (left) average fraction of constraint satisfaction, and (right) average objective value.

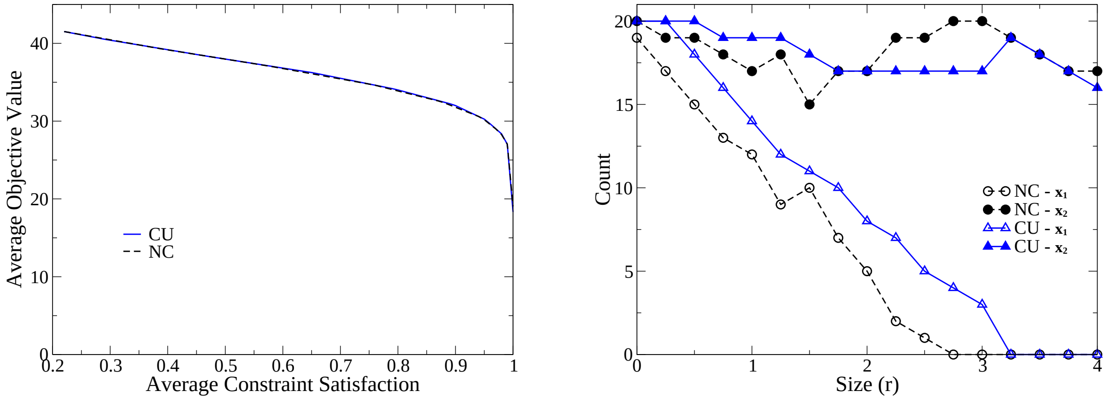

Figure 2 Comparison of connected and non-connected sets: (left) average objective value vs. constraint satisfaction, and (right) for a single iteration, the number of non-zero variables of a period.

To capture risk aversion, in each experiment, we maximize a concave piecewise linear utility function $u = \min(1.5r, 0.015 + r, 0.06 + 0.1r)$, where r denotes the return on the portfolio. It is assumed that $\overline{\boldsymbol{\mu}}_2 = \boldsymbol{\mu}_2(\mathbf{d}_1) + \boldsymbol{\delta}$ and $\underline{\boldsymbol{\mu}}_2 = \boldsymbol{\mu}_2(\mathbf{d}_1) - \boldsymbol{\delta}$ with $\boldsymbol{\mu}_2(\mathbf{d}_1) = \boldsymbol{\mu}_0 + \mathbf{A}\boldsymbol{\mu}_1 + \mathbf{B}\mathbf{d}_1$. The vector $\boldsymbol{\mu}_0$ and the matrices $\mathbf{A}$ and $\mathbf{B}$ are estimated from data via a vector autoregressive moving average with lag 1. The parameters $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are set to the residual covariance matrices. 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 (in order to cover almost all realized means under normality assumption).

To probe the performance of the proposed CU model, we compare it against the standard DRO and a modified DRO, described as following:

CU: Model with connected uncertainty set,

DRO-1: Model with $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_1$, and

DRO-2: Model with $\boldsymbol{\mu}_2 = \mathbf{A}\boldsymbol{\mu}_1 + \mathbf{B}\boldsymbol{\mu}_1$.

DRO-1 represents the standard model as used widely in the literature. It is computationally attractive because of a simpler reformulation. However, it does not take into account potential connections to previous periods. We propose a modified version (DRO-2), which captures some of the potential connections to previous periods using the unconditional mean for the second period.

In these experiments, the parameters $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ are defined based on past weeks, and the returns are computed over the future two weeks. We evaluate the solution quality from these three models by comparing the returns. Each experiment starts with an initial wealth of $W_0 = \$100$, which is recomputed using the total return on the portfolio

$$W_{t+1} = W_t \cdot (\widehat{\mathbf{d}}_t^\top \mathbf{x}_t),\ t = 1, 2,$$

The realized demand $\widehat{\mathbf{d}}_t$ is taken from the actual stock data. This wealth (W) is then averaged over the different experiments and reported in Table 1, along with the standard deviation (Std) as a measure for robustness of the solution.

Model	Period 1		Period 2		Period 3	
	W	Std	W	Std	W	Std
CU	100	0	100.1	1.97	101.2	5.38
DRO1	100	0	99.9	2.36	100.9	5.92
DRO2	100	0	99.9	2.36	100.9	6.06

Table 1 Average wealth (W) and its standard deviation (Std) over time for various models

Table 1 displays the performances of the three different models. For all three, we observe a positive average return at the end of two weeks, which is attributed to the mean positive return on the random samples for the five stocks. Furthermore, the standard deviation (spread) of the sample paths grows as time elapses, because the inherent uncertainty in returns is compounding over time.

When comparing the three models, we observe marginal differences in the average wealth, reflecting the relatively short time horizon. However, we observe the wealth standard deviation for the CU model to be lower than both of the DRO models. This is attributed to the fact that the CU model captures the compounding worst-case effect of connected periods and yields a more conservative solution. the CU model reduces the wealth standard deviation for wealth by 0.417 (17.6% of the original wealth standard deviation) for period 1 and 0.563 (9.5%) for period 2 compared to DRO-1. Other measures of deviation, such as interquartile range, reveal a similar decrease in wealth. Therefore, the CU model is able to select assets that are less volatile in order to provide more robust allocations.

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

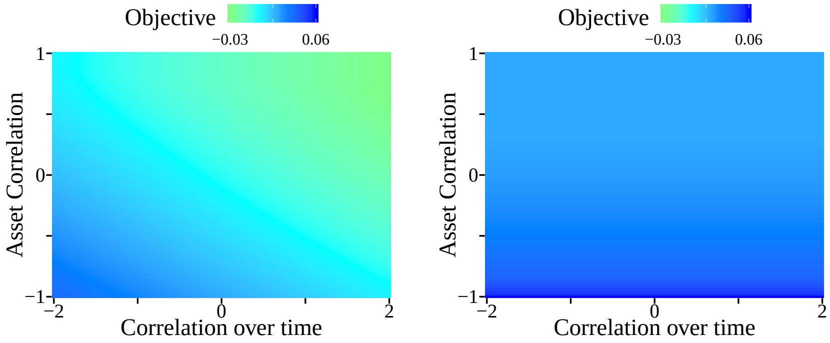

Figure 3 **CU and DRO problem objective**

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.

8. DRO Application: Portfolio Optimization II

1. *Objective function:* to measure performance of the model in the nominal setting (Figure 3).
2. *Worst case realized wealth:* to study the effect of adversarial settings (Figure 5).
3. *Standard deviation of wealth:* to evaluate the robustness of solutions (Figure 4).

Objective

Asset Correlation Figure 3 indicates that positive correlations among the assets reduce the objective performance for the CU model. This is due to the fact that positive correlation worsens the worst-case returns everywhere simultaneously, whereas a negative correlation prevents this from occurring as a result the objective (which is worst-case by definition) reduces.

Correlation over time Figure 3 also presents similar behavior for correlation over time leading to worse objective performance for the same reasons as asset correlation. The DRO model does not capture the correlation over time but it also performs better when the assets are negatively correlated.

Standard Deviation and Worst Case Wealth

Figure 4 shows the standard deviation for the realized wealth at the end of the second period.

Asset Correlation The standard deviation is higher, when the assets are positively correlated among themselves and over time.

Figure 5 shows the worst case realized wealth at the end of the second period. The worst case wealth is higher, when assets are negatively correlated among themselves and over time. This is because the negative correlation makes it unlikely for both returns to be very low. The pattern

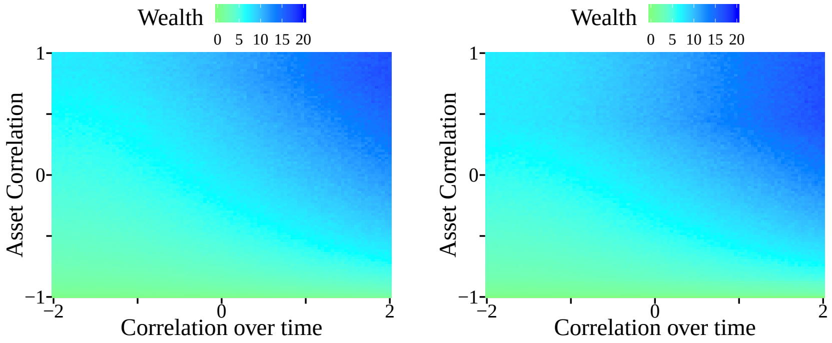

Figure 4 **Realized wealth standard deviation for CU and DRO at end of period 2**

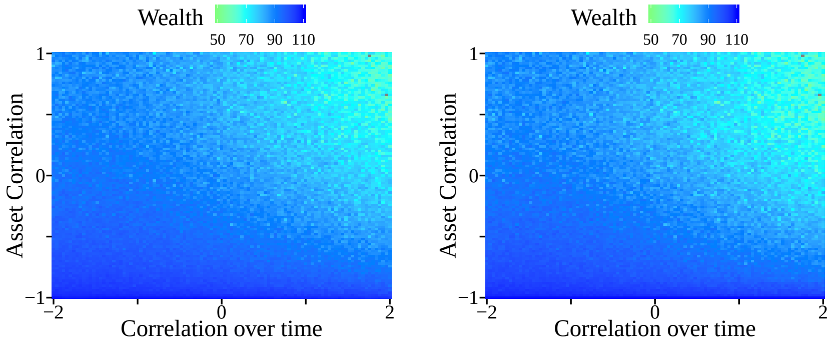

Figure 5 **Worst case realized wealth for period 2**

observed here is similar to that in Figure 3, with the worst case wealth being worse when assets are positively correlated.

References

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

Hanasusanto GA, Kuhn D, Wiesemann W (2015) K-adaptability in two-stage robust binary programming. *Oper. Res.* 63(4):877–891.

Lorca A, Sun XA (2017) Multistage robust unit commitment with dynamic uncertainty sets and energy storage. *IEEE Trans. Pow. Sys.* 32(3):1678–1688.

Solyalı O, Cordeau JF, Laporte G (2015) The impact of modeling on robust inventory management under demand uncertainty. *Management Sci.* 62(4):1188–1201.