Correction to: A Lagrangian dual method for two-stage robust optimization with binary uncertainties

Henri Lefebvre[1] and Anirudh Subramanyam[2]

[1]Trier University, Trier, Germany
[2]The Pennsylvania State University, University Park, PA, USA

November 8, 2024

Abstract

We provide a correction to the sufficient conditions under which closed-form expressions for the optimal Lagrange multiplier are provided in [4]. We first present a simple counterexample where the original conditions are insufficient, highlight where the original proof fails, and then provide modified conditions along with a correct proof of their validity. Finally, although the original paper discusses modifications to their method for problems that may not satisfy any sufficient conditions, we substantiate that discussion along two directions. We first show that computing an optimal Lagrange multiplier can still be done in polynomial time. We then provide complete and correct versions of the corresponding Benders and column-and-constraint generation algorithms in which the original method is used. We also discuss the implications of our findings on computational performance.

1 Background

In [4], the author considers two-stage robust optimization problems with binary-valued uncertain data and proposes a new method to construct worst-case parameter realizations in such problems. To keep our presentation succinct, we adopt all notation and assumptions from that paper, and consider the problem formulation denoted as $\mathcal{P}$, shown below.

$$\inf_{\boldsymbol{x}\in\mathcal{X}} \sup_{\boldsymbol{\xi}\in\Xi} \mathcal{Q}(\boldsymbol{x},\boldsymbol{\xi}),$$
$$\mathcal{Q}(\boldsymbol{x},\boldsymbol{\xi}) = \begin{bmatrix} \underset{\boldsymbol{y}\in\mathcal{Y}}{\text{minimize}} \;\; \boldsymbol{c}(\boldsymbol{\xi})^\top \boldsymbol{x} + \boldsymbol{d}(\boldsymbol{\xi})^\top \boldsymbol{y} \\ \text{subject to } \boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{y} \geq \boldsymbol{h}(\boldsymbol{\xi}) \end{bmatrix}. \tag{$\mathcal{P}$}$$

The central idea of the method in [4] is the development of the following Lagrangian dual with scalar-valued multiplier $\lambda \in \mathbb{R}_+$.

$$\mathcal{L}(\boldsymbol{x},\boldsymbol{\xi},\lambda) = \begin{bmatrix} \underset{\boldsymbol{y}\in\mathcal{Y},\boldsymbol{z}\in\mathbb{R}_+^{np}}{\text{minimize}} \;\; \boldsymbol{c}(\boldsymbol{\xi})^\top \boldsymbol{x} + \boldsymbol{d}(\boldsymbol{\xi})^\top \boldsymbol{y} + \lambda\phi(\boldsymbol{z},\boldsymbol{\xi}) \\ \text{subject to } \;\; \boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{y} \geq \boldsymbol{h}(\boldsymbol{z}), \;\; \boldsymbol{z} \leq \mathbf{e}. \end{bmatrix} \tag{1}$$
$$\phi(\boldsymbol{z},\boldsymbol{\xi}) = \mathbf{e}^\top \boldsymbol{z} + \mathbf{e}^\top \boldsymbol{\xi} - 2\boldsymbol{z}^\top \boldsymbol{\xi}. \tag{2}$$

It is shown in [4, Theorems 1 and 2] that the Lagrangian dual has the following attractive properties.

$$\mathcal{Q}(\boldsymbol{x}, \boldsymbol{\xi}) = \sup_{\lambda \in \mathbb{R}_+} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\xi}, \lambda) \quad \text{for all } \boldsymbol{x} \in \mathcal{X}, \boldsymbol{\xi} \in \Xi. \tag{3}$$

$$\inf_{\boldsymbol{x} \in \mathcal{X}} \sup_{\boldsymbol{\xi} \in \Xi} \mathcal{Q}(\boldsymbol{x}, \boldsymbol{\xi}) = \sup_{\lambda \in \mathbb{R}_+} \inf_{\boldsymbol{x} \in \mathcal{X}} \sup_{\boldsymbol{\xi} \in \Xi} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\xi}, \lambda) \tag{4}$$

An immediate consequence of these properties is the following equation to which we shall refer in this paper. This equation is obtained by simply taking the supremum of both sides of (3) over $\boldsymbol{\xi} \in \Xi$.

$$\sup_{\boldsymbol{\xi} \in \Xi} \mathcal{Q}(\boldsymbol{x}, \boldsymbol{\xi}) = \sup_{\lambda \in \mathbb{R}_+} \sup_{\boldsymbol{\xi} \in \Xi} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\xi}, \lambda) \quad \text{for all } \boldsymbol{x} \in \mathcal{X}. \tag{5}$$

In [4, Theorem 4], it is shown that one can compute a closed-form expression for the Lagrange multiplier that maximizes the right-hand side of (5) if "for every $\boldsymbol{x} \in \mathcal{X}$, either there exists $\boldsymbol{\xi} \in \Xi$ such that $\mathcal{Q}(\boldsymbol{x}, \boldsymbol{\xi}) = +\infty$ or $\mathcal{Q}(\boldsymbol{x}, \boldsymbol{\xi}) < +\infty$ for all $\boldsymbol{\xi} \in \{0,1\}^{n_p}$". Whenever this condition holds, the paper claims that for any feasible first-stage decision $\boldsymbol{x} \in \mathcal{X}$; that is, for which $\sup\{\mathcal{Q}(\boldsymbol{x}, \boldsymbol{\xi}) : \boldsymbol{\xi} \in \Xi\} < +\infty$,

$$u(\boldsymbol{x}) - \ell(\boldsymbol{x}) \in \arg\max_{\lambda \in \mathbb{R}_+} \left\{ \max_{\boldsymbol{\xi} \in \Xi} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\xi}, \lambda) \right\}, \tag{6}$$

where $u(\boldsymbol{x})$ is any finite upper bound on $\sup\{\mathcal{Q}(\boldsymbol{x}, \boldsymbol{\xi}) : \boldsymbol{\xi} \in \Xi\}$ and $\ell(\boldsymbol{x})$ is any finite lower bound on $\inf\{\boldsymbol{c}(\boldsymbol{\xi})^\top \boldsymbol{x} + \boldsymbol{d}(\boldsymbol{\xi})^\top \boldsymbol{y} : \boldsymbol{\xi} \in \Xi, \boldsymbol{y} \in \mathcal{Y}\}$. In other words, $u(\boldsymbol{x}) - \ell(\boldsymbol{x})$ is an optimal Lagrange multiplier.

2 Counterexample

Consider the problem with $\mathcal{X} = \{0\}$, $\Xi = \{0,1\}$, $\mathcal{Y} = \{0,1\}$, and whose second-stage value function is given by

$$\mathcal{Q}(x, \xi) = \begin{bmatrix} \underset{y \in \mathcal{Y}}{\text{minimize}} & -y \\ \text{subject to} & y \leq \dfrac{3}{2} - \xi \end{bmatrix}.$$

For simplicity, this problem has a unique feasible first-stage decision, so that the second-stage value function depends only on the uncertain parameter ξ. The counterexample can, however, be easily extended to more complex first-stage decision spaces.

It can be readily verified by enumeration that $\mathcal{Q}(x, \xi) < +\infty$ for all $x \in \mathcal{X}$ and all $\xi \in \Xi$. Hence, the conditions of [4, Theorem 4] are satisfied. We show, however, that the claimed result is wrong. To that end, observe the following obtained by simply enumerating all points in Ξ and $\mathcal{Y}$:

$$\begin{aligned} \sup_{\xi \in \Xi} \mathcal{Q}(x, \xi) &= \max \left\{ \min_{y \in \{0,1\}, y \leq \frac{3}{2}} -y, \quad \min_{y \in \{0,1\}, y \leq \frac{1}{2}} -y \right\} = \max\{-1, 0\} = 0, \\ \inf_{y \in \{0,1\}} -y &= \min\{0, -1\} = -1. \end{aligned}$$

According to [4, Theorem 4], one can choose $u(x) = 0$ and $\ell(x) = -1$ to ensure that $u(x) - \ell(x) = 1$ is an optimal Lagrange multiplier, namely that $1 \in \arg\max_{\lambda \in \mathbb{R}_+} \{\max_{\xi \in \Xi} \mathcal{L}(x, \xi, \lambda)\}$. We now proceed to show that this is false by explicitly calculating $\mathcal{L}(x, \xi, \lambda)$ by brute force enumeration. In deriving the following, we only use the fact that $\lambda \geq 0$.

$$\max_{\xi \in \Xi} \mathcal{L}(x, \xi, \lambda) = \max_{\xi \in \{0,1\}} \left\{ \min_{\substack{(y,z) \in \{0,1\} \times [0,1], \\ y + z \leq \frac{3}{2}}} -y + \lambda (z + \xi - 2z\xi) \right\}$$

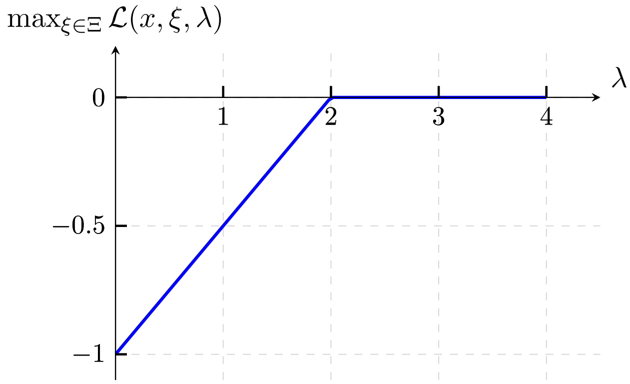

Figure 1: Plot of $\max_{\xi\in\Xi}\mathcal{L}(x,\xi,\lambda)$ versus λ for the counterexample.

$$
\begin{aligned}
&= \max\left\{ \min_{\substack{(y,z)\in\{0,1\}\times[0,1],\\ y+z\le\frac{3}{2}}} -y+\lambda z,\ \min_{\substack{(y,z)\in\{0,1\}\times[0,1],\\ y+z\le\frac{3}{2}}} -y+\lambda-\lambda z \right\} \\
&= \max\left\{ \min\left\{ \min_{z\in[0,1],z\le\frac{3}{2}} \lambda z,\ \min_{z\in[0,1],z\le\frac{1}{2}} -1+\lambda z \right\}, \right. \\
&\qquad\left. \min\left\{ \min_{z\in[0,1],z\le\frac{3}{2}} \lambda-\lambda z,\ \min_{z\in[0,1],z\le\frac{1}{2}} -1+\lambda-\lambda z \right\} \right\} \\
&= \max\left\{ \min\left\{0,-1\right\}, \min\left\{0, \frac{\lambda}{2}-1\right\} \right\} \\
&= \max\left\{ -1, \min\left\{0, \frac{\lambda}{2}-1\right\} \right\}.
\end{aligned}
$$

A plot of the function is shown in Figure 1, which indicates that the function is not maximized at $\lambda = u(x) - \ell(x) = 1$. In other words, equation (6) is false.

3 Correct Sufficient Conditions

The proof presented in [4, Theorem 4] fails when it is claimed that "problem (1) always has an optimal solution $(\hat{\boldsymbol{y}}, \hat{\boldsymbol{z}})$ such that $\hat{\boldsymbol{z}} \in \{0,1\}^{n_p}$, for any $\boldsymbol{\xi} \in \Xi$ and $\lambda \in \mathbb{R}_+$". It is worth noting, however, that all other steps of the proof remain valid whenever this claim is true. Below, we provide complete and correct conditions under which this claim is true and hence, the conclusion of [4, Theorem 4] remains valid. Additionally, we highlight that although Theorem 5, Algorithm 3, and Algorithm 8 in the original paper invoke [4, Theorem 4], they do not require any modifications themselves. Indeed, their validity is unaffected as long as the correct sufficient conditions are satisifed.

For simplicity of our ensuing exposition, we disaggregate the second-stage decisions in $\mathcal{P}$ into their continuous and discrete components. Let $\boldsymbol{y} = (\boldsymbol{y}_{\mathrm{c}}, \boldsymbol{y}_{\mathrm{d}})$, where $\boldsymbol{y}_{\mathrm{c}}$ and $\boldsymbol{y}_{\mathrm{d}}$ are the vector of continuous and discrete second-stage variables, respectively. We extend this notation to matrices and vectors which are multiplied by $\boldsymbol{y}$ (for example, $\boldsymbol{W}\boldsymbol{y} = \boldsymbol{W}_{\mathrm{c}}\boldsymbol{y}_{\mathrm{c}} + \boldsymbol{W}_{\mathrm{d}}\boldsymbol{y}_{\mathrm{d}}$), so that $\mathcal{Q}(\boldsymbol{x}, \boldsymbol{\xi})$ can also be equivalently written as:

$$\mathcal{Q}(\boldsymbol{x},\boldsymbol{\xi}) = \begin{bmatrix} \underset{\boldsymbol{y}}{\text{minimize}} \;\; \boldsymbol{c}(\boldsymbol{\xi})^\top \boldsymbol{x} + \boldsymbol{d}_\mathrm{c}(\boldsymbol{\xi})^\top \boldsymbol{y}_\mathrm{c} + \boldsymbol{d}_\mathrm{d}(\boldsymbol{\xi})^\top \boldsymbol{y}_\mathrm{d} \\ \text{subject to } \boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}_\mathrm{c}\boldsymbol{y}_\mathrm{c} + \boldsymbol{W}_\mathrm{d}\boldsymbol{y}_\mathrm{d} \geq \boldsymbol{h}(\boldsymbol{\xi}) \\ \qquad\qquad \boldsymbol{y} = (\boldsymbol{y}_\mathrm{c}, \boldsymbol{y}_\mathrm{d}) \in \mathcal{Y} := \mathbb{R}_+^{nc_2} \times \mathcal{Y}_\mathrm{d} \end{bmatrix},$$

where $\mathcal{Y}_\mathrm{d} \subseteq \mathbb{Z}^{nd_2}$, $nc_2, nd_2 \in \mathbb{Z} \cap [0, n_2]$ and $nc_2 + nd_2 = n_2$. This representation is general enough to allow the second-stage decisions to be purely continuous ($nc_2 = n_2$), purely integer ($nd_2 = n_2$) or mixed-integer.

We also define $\mathcal{Q}(\boldsymbol{x},\boldsymbol{\xi};\boldsymbol{y}_\mathrm{d})$ and $\mathcal{L}(\boldsymbol{x},\boldsymbol{\xi},\lambda;\boldsymbol{y}_\mathrm{d})$ as restrictions of $\mathcal{Q}(\boldsymbol{x},\boldsymbol{\xi})$ and $\mathcal{L}(\boldsymbol{x},\boldsymbol{\xi},\lambda)$, respectively, where the values of the discrete second-stage decisions are fixed to $\boldsymbol{y}_\mathrm{d}$, as shown below.

$$\mathcal{Q}(\boldsymbol{x},\boldsymbol{\xi};\boldsymbol{y}_\mathrm{d}) = \begin{bmatrix} \underset{\boldsymbol{y}_\mathrm{c}\in\mathbb{R}_+^{nc_2}}{\text{minimize}} \;\; \boldsymbol{c}(\boldsymbol{\xi})^\top \boldsymbol{x} + \boldsymbol{d}_\mathrm{c}(\boldsymbol{\xi})^\top \boldsymbol{y}_\mathrm{c} + \boldsymbol{d}_\mathrm{d}(\boldsymbol{\xi})^\top \boldsymbol{y}_\mathrm{d} \\ \text{subject to } \;\; \boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}_\mathrm{c}\boldsymbol{y}_\mathrm{c} + \boldsymbol{W}_\mathrm{d}\boldsymbol{y}_\mathrm{d} \geq \boldsymbol{h}(\boldsymbol{\xi}) \end{bmatrix}, \tag{8}$$

$$\mathcal{L}(\boldsymbol{x},\boldsymbol{\xi},\lambda;\boldsymbol{y}_\mathrm{d}) = \begin{bmatrix} \underset{\boldsymbol{y}_\mathrm{c}\in\mathbb{R}_+^{nc_2},\boldsymbol{z}\in\mathbb{R}_+^{n_p}}{\text{minimize}} \;\; \boldsymbol{c}(\boldsymbol{\xi})^\top \boldsymbol{x} + \boldsymbol{d}_\mathrm{c}(\boldsymbol{\xi})^\top \boldsymbol{y}_\mathrm{c} + \boldsymbol{d}_\mathrm{d}(\boldsymbol{\xi})^\top \boldsymbol{y}_\mathrm{d} + \lambda\phi(\boldsymbol{z},\boldsymbol{\xi}) \\ \text{subject to } \;\; \boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}_\mathrm{c}\boldsymbol{y}_\mathrm{c} + \boldsymbol{W}_\mathrm{d}\boldsymbol{y}_\mathrm{d} \geq \boldsymbol{h}(\boldsymbol{z}), \;\; \boldsymbol{z} \leq \mathbf{e}. \end{bmatrix}. \tag{9}$$

The next theorem provides the correct sufficient conditions under which one can compute closed-form expressions for the optimal Lagrange multiplier.

Theorem 1. *Suppose that the following conditions are satisfied in problem $\mathcal{P}$.*

1. $\mathcal{X} \subseteq \mathbb{Z}^{n_1}$.
2. $\boldsymbol{T} \in \mathbb{Z}^{m\times n_1}$, $\boldsymbol{W} \in \mathbb{Z}^{m\times n_2}$, $\boldsymbol{h}(\boldsymbol{\xi}) = \boldsymbol{h}_0 + \boldsymbol{H}\boldsymbol{\xi}$, *where* $\boldsymbol{h}_0 \in \mathbb{Z}^m$, $\boldsymbol{H} \in \mathbb{Z}^{m\times n_p}$.
3. $[\boldsymbol{W}_\mathrm{c} \;\; -\boldsymbol{H}] \in \mathbb{Z}^{m\times(nc_2+n_p)}$ *is a totally unimodular matrix.*

Then, for any feasible first-stage decision $\boldsymbol{x} \in \mathcal{X}$; that is, for which $\sup\{\mathcal{Q}(\boldsymbol{x},\boldsymbol{\xi}) : \boldsymbol{\xi} \in \Xi\} < +\infty$, *we have that*

$$u(\boldsymbol{x}) - \ell(\boldsymbol{x}) \in \underset{\lambda\in\mathbb{R}_+}{\arg\max} \left\{ \max_{\boldsymbol{\xi}\in\Xi} \mathcal{L}(\boldsymbol{x},\boldsymbol{\xi},\lambda) \right\},$$

where $u(\boldsymbol{x})$ is any finite upper bound on $\sup\{\mathcal{Q}(\boldsymbol{x},\boldsymbol{\xi}) : \boldsymbol{\xi} \in \Xi\}$ *and $\ell(\boldsymbol{x})$ is any finite lower bound on* $\inf\{\boldsymbol{c}(\boldsymbol{\xi})^\top\boldsymbol{x} + \boldsymbol{d}(\boldsymbol{\xi})^\top\boldsymbol{y} : \boldsymbol{\xi} \in \Xi, \boldsymbol{y} \in \mathcal{Y}\}$.

Proof. Suppose that $\boldsymbol{x} \in \mathcal{X}$ is any feasible first-stage decision in $\mathcal{P}$. We shall show that under the conditions stated in the Theorem, problem (1) always has an optimal solution $(\hat{\boldsymbol{y}}, \hat{\boldsymbol{z}})$ such that $\hat{\boldsymbol{z}} \in \{0,1\}^{n_p}$, for any $\boldsymbol{\xi} \in \Xi$ and $\lambda \in \mathbb{R}_+$. The rest of the argument follows from the proof of [4, Theorem 4] and remains unchanged.

Using the definition of $\mathcal{L}(\boldsymbol{x},\boldsymbol{\xi},\lambda;\boldsymbol{y}_\mathrm{d})$ in equation (9), problem (1) can be equivalently written as the following nested optimization problem:

$$\mathcal{L}(\boldsymbol{x},\boldsymbol{\xi},\lambda) = \underset{\boldsymbol{y}_\mathrm{d}\in\mathcal{Y}_\mathrm{d}}{\text{minimize}}\, \mathcal{L}(\boldsymbol{x},\boldsymbol{\xi},\lambda;\boldsymbol{y}_\mathrm{d})$$

Under the stated conditions, it can be readily verified the constraint matrix defining the feasible region of $\mathcal{L}(\boldsymbol{x},\boldsymbol{\xi},\lambda;\boldsymbol{y}_\mathrm{d})$ is totally unimodular. Moreover, the right-hand side coefficients, $\boldsymbol{h}_0 - \boldsymbol{T}\boldsymbol{x} - \boldsymbol{W}_\mathrm{d}\boldsymbol{y}_\mathrm{d}$, are integer-valued for any $\boldsymbol{x} \in \mathcal{X} \subseteq \mathbb{Z}^{n_1}$ and $\boldsymbol{y}_\mathrm{d} \in \mathcal{Y}_\mathrm{d} \subseteq \mathbb{Z}^{nd_2}$. Therefore, the polyhedron defining the feasible region of $\mathcal{L}(\boldsymbol{x},\boldsymbol{\xi},\lambda;\boldsymbol{y}_\mathrm{d})$ has integer vertices, as does its optimal solution. Hence, any optimal solution $(\hat{\boldsymbol{y}}, \hat{\boldsymbol{z}})$ of $\mathcal{L}(\boldsymbol{x},\boldsymbol{\xi},\lambda)$ must satisfy $\hat{\boldsymbol{z}} \in \{0,1\}^{n_p}$. □

We now present a class of problems where the conditions are satisfied.

Example 1 (Interdiction Constraints)**.** *Suppose that the second-stage problem is combinatorial, $\mathcal{Y} = \{0,1\}^{n_2}$, and the second-stage decisions represent resources that are being interdicted depending on some random realization of the uncertain parameters, as shown below. Such structures are common in network interdiction problems, including facility location with random facility disruptions; e.g., see [2, 3].*

$$\mathcal{Q}(\boldsymbol{x},\boldsymbol{\xi}) = \begin{bmatrix} \textit{minimize}_{\boldsymbol{y}\in\mathcal{Y}} \ \boldsymbol{c}(\boldsymbol{\xi})^\top \boldsymbol{x} + \boldsymbol{d}(\boldsymbol{\xi})^\top \boldsymbol{y} \\ \textit{subject to}\ \boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{y} \geq \boldsymbol{h}, \\ \qquad \boldsymbol{0} \leq \boldsymbol{y} \leq \boldsymbol{e} - \boldsymbol{\xi} \end{bmatrix}.$$

Suppose also that $\mathcal{X} \subseteq \mathbb{Z}^{n_1}$ and that all matrices are integer. Then, it can be verified that the assumptions of Theorem 1 are satisfied (with $nc_2 = 0$). One can then compute a closed-form expression for the optimal multiplier using expressions for $u(\boldsymbol{x})$ and $\ell(\boldsymbol{x})$ provided in [4, Theorem 5].

Our proof argument allows us to also extend Theorem 1 to problem $\mathcal{P}_{\mathcal{I}}$ (reproduced below from the original paper) without requiring significant changes.

$$\begin{aligned} &\inf_{\boldsymbol{x}\in\mathcal{X}} \sup_{\boldsymbol{\xi}\in\Xi} \mathcal{Q}_{\mathcal{I}}(\boldsymbol{x},\boldsymbol{\xi}), \\ &\mathcal{Q}_{\mathcal{I}}(\boldsymbol{x},\boldsymbol{\xi}) = \begin{bmatrix} \text{minimize}_{\boldsymbol{y}\in\mathcal{Y}} \ \boldsymbol{c}(\boldsymbol{\xi})^\top \boldsymbol{x} + \boldsymbol{d}(\boldsymbol{\xi})^\top \boldsymbol{y} \\ \text{subject to}\ \boldsymbol{g}(\boldsymbol{x},\boldsymbol{y}) \geq \boldsymbol{0} \\ \qquad \xi_j = 0 \implies g_i(\boldsymbol{x},\boldsymbol{y}) = 0,\ \ i \in \mathcal{I}_j^0,\ j \in [n_p] \\ \qquad \xi_j = 1 \implies g_i(\boldsymbol{x},\boldsymbol{y}) = 0,\ \ i \in \mathcal{I}_j^1,\ j \in [n_p] \end{bmatrix}. \end{aligned} \tag{$\mathcal{P}_{\mathcal{I}}$}$$

Here, $\boldsymbol{g}(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{y} - \boldsymbol{h}_0$ and $\mathcal{I}_j^0, \mathcal{I}_j^1 \subseteq [m]$ are some index sets. Also, the corresponding Lagrangian is given by:

$$\mathcal{L}_{\mathcal{I}}(\boldsymbol{x},\boldsymbol{\xi},\lambda) = \begin{bmatrix} \text{minimize}_{\boldsymbol{y}\in\mathcal{Y}} \ \boldsymbol{c}(\boldsymbol{\xi})^\top \boldsymbol{x} + \boldsymbol{d}(\boldsymbol{\xi})^\top \boldsymbol{y} + \lambda\phi_{\mathcal{I}}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\xi}) \\ \text{subject to}\ \ \boldsymbol{g}(\boldsymbol{x},\boldsymbol{y}) \geq \boldsymbol{0}. \end{bmatrix} \tag{10}$$

$$\phi_{\mathcal{I}}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\xi}) = \sum_{j\in[n_p]} \sum_{i\in\mathcal{I}_j^1} \xi_j g_i(\boldsymbol{x},\boldsymbol{y}) + \sum_{j\in[n_p]} \sum_{i\in\mathcal{I}_j^0} (1-\xi_j) g_i(\boldsymbol{x},\boldsymbol{y}). \tag{11}$$

We now present the following analog of Theorem 1 for problem $\mathcal{P}_{\mathcal{I}}$. Notably, we highlight that a similar result was not postulated in the original paper. The proof is similar to that of Theorem 1 and we omit it for the sake of brevity.

Theorem 2. *Suppose that the following conditions are satisfied in problem $\mathcal{P}_{\mathcal{I}}$.*

1. $\mathcal{X} \subseteq \mathbb{Z}^{n_1}$.
2. $\boldsymbol{T} \in \mathbb{Z}^{m\times n_1}$, $\boldsymbol{W} \in \mathbb{Z}^{m\times n_2}$, $\boldsymbol{h}_0 \in \mathbb{Z}^m$.
3. $\boldsymbol{W}_{\mathrm{c}} \in \mathbb{Z}^{m\times nc_2}$ *is a totally unimodular matrix.*

Then, for any feasible first-stage decision $\boldsymbol{x} \in \mathcal{X}$; that is, for which $\sup\left\{\mathcal{Q}_{\mathcal{I}}(\boldsymbol{x},\boldsymbol{\xi}) : \boldsymbol{\xi} \in \Xi\right\} < +\infty$, we have that

$$u(\boldsymbol{x}) - \ell(\boldsymbol{x}) \in \arg\max_{\lambda\in\mathbb{R}_+} \left\{ \max_{\boldsymbol{\xi}\in\Xi} \mathcal{L}_{\mathcal{I}}(\boldsymbol{x},\boldsymbol{\xi},\lambda) \right\},$$

where $u(\boldsymbol{x})$ is any finite upper bound on $\sup\left\{\mathcal{Q}_{\mathcal{I}}(\boldsymbol{x},\boldsymbol{\xi}) : \boldsymbol{\xi} \in \Xi\right\}$ and $\ell(\boldsymbol{x})$ is any finite lower bound on $\inf\left\{\boldsymbol{c}(\boldsymbol{\xi})^\top \boldsymbol{x} + \boldsymbol{d}(\boldsymbol{\xi})^\top \boldsymbol{y} : \boldsymbol{\xi} \in \Xi, \boldsymbol{y} \in \mathcal{Y}\right\}$.

4 Absence of Sufficient Conditions

We now discuss key algorithmic implications in the absence of sufficient conditions that allow for a closed-form expression. Although the original paper includes a short discussion to that end, we present a more thorough and formal treatment in this section.

4.1 Computational Complexity

We show that computing an optimal Lagrange multiplier can still be done in time that is polynomial in the size of the input data under fairly general conditions. To ease our presentation, we first paraphrase a result from [1, Lemma 4] regarding solutions of linear programs.

Lemma 1 ([1])**.** *Let $\|\boldsymbol{X}\|$ denote the maximum absolute value of any entry of a matrix or vector $\boldsymbol{X}$. Let $P = \{\boldsymbol{\omega} \in \mathbb{R}_+^{n_\omega} : \boldsymbol{A}\boldsymbol{\omega} = \boldsymbol{b}\}$ be a polyhedron with $\boldsymbol{A} \in \mathbb{Z}^{m_\omega \times n_\omega}$ and $\boldsymbol{b} \in \mathbb{Z}^{m_\omega}$. Then, any vertex $\bar{\boldsymbol{\omega}}$ of P satisfies $\|\bar{\boldsymbol{\omega}}\| \leq m_\omega! \|\boldsymbol{b}\| \|\boldsymbol{A}\|^{m_\omega - 1}$.*

We highlight that the bound in Lemma 1 can be computed in time polynomial in the input data $(\boldsymbol{A}, \boldsymbol{b})$. It allows us to prove our main complexity result, which we state next.

Theorem 3. *Suppose that the following conditions are satisfied in problem $\mathcal{P}$.*

1. $\mathcal{X} \subseteq \mathbb{Z}^{n_1} \cap [\boldsymbol{x}^\ell, \boldsymbol{x}^u]$.
2. *For all $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{\xi} \in \Xi$, we have $\{\boldsymbol{y} \in \mathcal{Y} : \boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{y} \geq \boldsymbol{h}(\boldsymbol{\xi})\} \subseteq [\boldsymbol{y}^\ell, \boldsymbol{y}^u]$.*
3. *$\boldsymbol{c}(\boldsymbol{\xi}) = \boldsymbol{C}\boldsymbol{\xi}$, $\boldsymbol{d}(\boldsymbol{\xi}) = \boldsymbol{D}\boldsymbol{\xi}$, and $\boldsymbol{h}(\boldsymbol{\xi}) = \boldsymbol{H}\boldsymbol{\xi}$ for some matrices $\boldsymbol{C}$, $\boldsymbol{D}$, and $\boldsymbol{H}$.*
4. *The matrices, $\boldsymbol{C}$, $\boldsymbol{D}$, $\boldsymbol{T}$, $\boldsymbol{W}$, $\boldsymbol{H}$, are integer-valued.*
5. *There exists $\boldsymbol{x} \in \mathcal{X}$ for which $\sup\left\{\mathcal{Q}(\boldsymbol{x}, \boldsymbol{\xi}) : \boldsymbol{\xi} \in \Xi\right\} < +\infty$.*

Then, there exists a finite $\bar{\lambda} \geq 0$ that is computable in polynomial time in the input data, $\boldsymbol{C}$, $\boldsymbol{D}$, $\boldsymbol{T}$, $\boldsymbol{W}$, $\boldsymbol{H}$, and the bounds $\boldsymbol{x}^\ell$, $\boldsymbol{x}^u$, $\boldsymbol{y}^\ell$, $\boldsymbol{y}^u$, such that

$$\inf_{\boldsymbol{x} \in \mathcal{X}} \sup_{\boldsymbol{\xi} \in \Xi} \mathcal{Q}(\boldsymbol{x}, \boldsymbol{\xi}) = \inf_{\boldsymbol{x} \in \mathcal{X}} \sup_{\boldsymbol{\xi} \in \Xi} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\xi}, \bar{\lambda}).$$

Proof. Let U denote any finite upper bound on the optimal value of problem $\mathcal{P}$,

$$\inf_{\boldsymbol{x} \in \mathcal{X}} \sup_{\boldsymbol{\xi} \in \Xi} \mathcal{Q}(\boldsymbol{x}, \boldsymbol{\xi}) \leq U.$$

Note that U exists due to the last condition in the hypothesis of the theorem. Moreover, it can be easily shown, similar to [4, Theorem 5], that under the stated hypotheses, U can be computed in polynomial time.

Now, denote the feasible region of $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\xi}, \lambda; \boldsymbol{y}_\text{d})$, defined in (9), as

$$\Pi(\boldsymbol{x}, \boldsymbol{y}_\text{d}) = \left\{(\boldsymbol{y}_\text{c}, \boldsymbol{z}) \in \mathbb{R}_+^{nc_2} \times [0,1]^{n_p} : \boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}_\text{c}\boldsymbol{y}_\text{c} + \boldsymbol{W}_\text{d}\boldsymbol{y}_\text{d} \geq \boldsymbol{h}(\boldsymbol{z})\right\}.$$

The majority of the proof is concerned with showing that one can compute a finite $\bar{\lambda}$ in polynomial time, satisfying the following relationship for all $\boldsymbol{x} \in \mathcal{X}$, $\boldsymbol{\xi} \in \Xi$ and $\boldsymbol{y}_\text{d} \in \mathcal{Y}_\text{d}$:

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\xi}, \bar{\lambda}; \boldsymbol{y}_\text{d}) \begin{cases} = \mathcal{Q}(\boldsymbol{x}, \boldsymbol{\xi}; \boldsymbol{y}_\text{d}) & \text{if } \mathcal{Q}(\boldsymbol{x}, \boldsymbol{\xi}; \boldsymbol{y}_\text{d}) < +\infty \text{ or } \Pi(\boldsymbol{x}, \boldsymbol{y}_\text{d}) = \emptyset, \\ \geq U & \text{otherwise.} \end{cases} \tag{12}$$

Supposing for the moment that this can be done, observe that we immediately obtain the desired equation stated in the theorem, since (12) implies

$$\begin{aligned}\inf_{\boldsymbol{x}\in\mathcal{X}}\sup_{\boldsymbol{\xi}\in\Xi}\mathcal{L}(\boldsymbol{x},\boldsymbol{\xi},\bar{\lambda}) &= \inf_{\boldsymbol{x}\in\mathcal{X}}\sup_{\boldsymbol{\xi}\in\Xi}\inf_{\boldsymbol{y}_{\rm d}\in\mathcal{Y}_{\rm d}}\mathcal{L}(\boldsymbol{x},\boldsymbol{\xi},\bar{\lambda};\boldsymbol{y}_{\rm d})\\ &= \inf_{\boldsymbol{x}\in\mathcal{X}}\sup_{\boldsymbol{\xi}\in\Xi}\inf_{\boldsymbol{y}_{\rm d}\in\mathcal{Y}_{\rm d}}Q(\boldsymbol{x},\boldsymbol{\xi};\boldsymbol{y}_{\rm d})\\ &= \inf_{\boldsymbol{x}\in\mathcal{X}}\sup_{\boldsymbol{\xi}\in\Xi}\mathcal{Q}(\boldsymbol{x},\boldsymbol{\xi}),\end{aligned}$$

where we used the fact that U is an upper bound on the optimal value of $\mathcal{P}$.

We now proceed to establish the validity of (12). Fix $\boldsymbol{x}\in\mathcal{X}$, $\boldsymbol{\xi}\in\Xi$, and $\boldsymbol{y}_{\rm d}\in\mathcal{Y}_{\rm d}$. The key idea is that $\mathcal{Q}(\boldsymbol{x},\boldsymbol{\xi};\boldsymbol{y}_{\rm d})$ can be equivalently written as:

$$\mathcal{Q}(\boldsymbol{x},\boldsymbol{\xi};\boldsymbol{y}_{\rm d}) = \begin{bmatrix}\underset{\boldsymbol{y}_{\rm c}\in\mathbb{R}^{nc_2}_+,\boldsymbol{z}\in\mathbb{R}^{n_p}_+}{\text{minimize}} & \boldsymbol{c}(\boldsymbol{\xi})^\top\boldsymbol{x}+\boldsymbol{d}_{\rm c}(\boldsymbol{\xi})^\top\boldsymbol{y}_{\rm c}+\boldsymbol{d}_{\rm d}(\boldsymbol{\xi})^\top\boldsymbol{y}_{\rm d}\\ \text{subject to} & \boldsymbol{T}\boldsymbol{x}+\boldsymbol{W}_{\rm c}\boldsymbol{y}_{\rm c}+\boldsymbol{W}_{\rm d}\boldsymbol{y}_{\rm d}\geq\boldsymbol{H}\boldsymbol{z},\\ & \boldsymbol{z}\leq\mathbf{e},\;\;\phi(\boldsymbol{z},\boldsymbol{\xi})\leq 0.\end{bmatrix}$$

Similarly, since $\phi(\boldsymbol{z},\boldsymbol{\xi})\geq 0$ from [4, Lemma 1], we note that $\mathcal{L}(\boldsymbol{x},\boldsymbol{\xi},\lambda;\boldsymbol{y}_{\rm d})$ can be equivalently written as:

$$\mathcal{L}(\boldsymbol{x},\boldsymbol{\xi},\lambda;\boldsymbol{y}_{\rm d}) = \begin{bmatrix}\underset{\substack{\boldsymbol{y}_{\rm c}\in\mathbb{R}^{nc_2}_+,\boldsymbol{z}\in\mathbb{R}^{n_p}_+\\ w\in\mathbb{R}_+}}{\text{minimize}} & \boldsymbol{c}(\boldsymbol{\xi})^\top\boldsymbol{x}+\boldsymbol{d}_{\rm c}(\boldsymbol{\xi})^\top\boldsymbol{y}_{\rm c}+\boldsymbol{d}_{\rm d}(\boldsymbol{\xi})^\top\boldsymbol{y}_{\rm d}+\lambda w\\ \text{subject to} & \boldsymbol{T}\boldsymbol{x}+\boldsymbol{W}_{\rm c}\boldsymbol{y}_{\rm c}+\boldsymbol{W}_{\rm d}\boldsymbol{y}_{\rm d}\geq\boldsymbol{H}\boldsymbol{z},\\ & \boldsymbol{z}\leq\mathbf{e},\;\;\phi(\boldsymbol{z},\boldsymbol{\xi})\leq w.\end{bmatrix}$$

We now distinguish two cases.

1. Suppose $\mathcal{Q}(\boldsymbol{x},\boldsymbol{\xi};\boldsymbol{y}_{\rm d})<+\infty$.

   Together with [4, Assumption A1], this means that $\mathcal{Q}(\boldsymbol{x},\boldsymbol{\xi};\boldsymbol{y}_{\rm d})$ is finite. Strong linear programming duality then implies that it can be equivalently written as the (finite) optimal value of the following problem, where we have also expanded $\phi(\boldsymbol{z},\boldsymbol{\xi})$ using its definition (2).

$$\begin{aligned}\underset{\boldsymbol{\mu}\in\mathbb{R}^m_+,\boldsymbol{\beta}\in\mathbb{R}^{n_p}_+,\alpha\in\mathbb{R}_+}{\text{maximize}}\quad & \boldsymbol{c}(\boldsymbol{\xi})^\top\boldsymbol{x}+\boldsymbol{d}_{\rm d}(\boldsymbol{\xi})^\top\boldsymbol{y}_{\rm d}+(-\boldsymbol{T}\boldsymbol{x}-\boldsymbol{W}_{\rm d}\boldsymbol{y}_{\rm d})^\top\boldsymbol{\mu}+\mathbf{e}^\top(\alpha\boldsymbol{\xi}-\boldsymbol{\beta})\\ \text{subject to}\quad & \boldsymbol{W}_{\rm c}^\top\boldsymbol{\mu}\leq\boldsymbol{d}_{\rm c}(\boldsymbol{\xi}),\\ & (2\boldsymbol{\xi}-\mathbf{e})\alpha-\boldsymbol{H}^\top\boldsymbol{\mu}-\boldsymbol{\beta}\leq\mathbf{0}.\end{aligned}$$

   The dual of $\mathcal{L}(\boldsymbol{x},\boldsymbol{\xi},\lambda;\boldsymbol{y}_{\rm d})$ is identical to the above problem with the additional constraint, $\alpha\leq\lambda$. Therefore, one can ensure (12) by choosing $\bar{\lambda}$ to be any upper bound on an optimal value of α. Without loss of generality (e.g., by converting the dual problem to standard form), there exists an optimal solution to the dual problem that lies at a vertex of the polyhedron defining its feasible region. Lemma 1 ensures that all entries, and in particular α, of such a vertex can be upper bounded by

$$(nc_2+n_p)!\|\boldsymbol{d}_{\rm c}(\boldsymbol{\xi})\|\max\{\|\boldsymbol{W}_{\rm c}\|,\|\boldsymbol{H}\|,\|2\boldsymbol{\xi}-\mathbf{e}\|,1\}^{nc_2+n_p-1}.$$

   Observe now that $\|\boldsymbol{d}_{\rm c}(\boldsymbol{\xi})\|=\|\boldsymbol{D}_{\rm c}\boldsymbol{\xi}\|=\|\boldsymbol{D}_{\rm c}\boldsymbol{\xi}\|_\infty\leq\|\boldsymbol{D}_{\rm c}\|_\infty\|\boldsymbol{\xi}\|_\infty\leq\|\boldsymbol{D}_{\rm c}\|_\infty$, since $\boldsymbol{\xi}\in\{0,1\}^{n_p}$ and where $\|\boldsymbol{D}_{\rm c}\|_\infty$ denotes the vector-induced matrix norm of $\boldsymbol{D}_{\rm c}$. Also, $\|2\boldsymbol{\xi}-\mathbf{e}\|\leq 1$. One can therefore choose $\bar{\lambda}$ as follows:

$$\hat{\lambda}\geq(nc_2+n_p)!\|\boldsymbol{D}_{\rm c}\|_\infty\max\{\|\boldsymbol{W}_{\rm c}\|,\|\boldsymbol{H}\|,1\}^{nc_2+n_p-1}.$$

Note that this bound does not depend on the chosen $\boldsymbol{x}$, $\boldsymbol{\xi}$ or $\boldsymbol{y}_{\mathrm{d}}$ and is computable in polynomial time in the input data.

2. Suppose now $\mathcal{Q}(\boldsymbol{x},\boldsymbol{\xi};\boldsymbol{y}_{\mathrm{d}}) = +\infty$. We again consider two cases.

   (a) Suppose $\Pi(\boldsymbol{x},\boldsymbol{y}_{\mathrm{d}}) = \emptyset$. Then, $\mathcal{L}(\boldsymbol{x},\boldsymbol{\xi},\lambda;\boldsymbol{y}_{\mathrm{d}}) = +\infty$ for any $\lambda \geq 0$. Hence, one can safely choose any $\bar{\lambda} \geq 0$ to achieve (12).

   (b) Suppose $\Pi(\boldsymbol{x},\boldsymbol{y}_{\mathrm{d}}) \neq \emptyset$. Using strong Lagrangian duality (3), it follows that for any finite V, there must exist a sufficiently large yet finite λ satisfying $\mathcal{L}(\boldsymbol{x},\boldsymbol{\xi},\lambda;\boldsymbol{y}_{\mathrm{d}}) \geq V$. Since $\mathcal{L}(\boldsymbol{x},\boldsymbol{\xi},\lambda;\boldsymbol{y}_{\mathrm{d}})$ is also finite (because $\Pi(\boldsymbol{x},\boldsymbol{y}_{\mathrm{d}}) \neq \emptyset$), finding λ satisfying $\mathcal{L}(\boldsymbol{x},\boldsymbol{\xi},\lambda;\boldsymbol{y}_{\mathrm{d}}) \geq V$ is equivalent to replacing $\mathcal{L}(\boldsymbol{x},\boldsymbol{\xi},\lambda;\boldsymbol{y}_{\mathrm{d}})$ by its (finite) dual and finding a feasible vector $(\boldsymbol{\mu},\boldsymbol{\beta},\alpha,\lambda) \in \mathbb{R}^m_+ \times \mathbb{R}^{n_p}_+ \times \mathbb{R}_+ \times \mathbb{R}_+$ satisfying the linear constraints:

$$
\begin{aligned}
&(-\boldsymbol{T}\boldsymbol{x} - \boldsymbol{W}_{\mathrm{d}}\boldsymbol{y}_{\mathrm{d}})^\top \boldsymbol{\mu} + \mathbf{e}^\top(\alpha\boldsymbol{\xi} - \boldsymbol{\beta}) \geq V - \boldsymbol{c}(\boldsymbol{\xi})^\top \boldsymbol{x} - \boldsymbol{d}_{\mathrm{d}}(\boldsymbol{\xi})^\top \boldsymbol{y}_{\mathrm{d}},\\
&\boldsymbol{W}_{\mathrm{c}}^\top \boldsymbol{\mu} \leq \boldsymbol{d}_{\mathrm{c}}(\boldsymbol{\xi}),\\
&(2\boldsymbol{\xi} - \boldsymbol{e})\alpha - \boldsymbol{H}^\top \boldsymbol{\mu} - \boldsymbol{\beta} \leq \boldsymbol{0},\\
&\alpha - \lambda \leq 0.
\end{aligned}
$$

   We can then achieve (12) by choosing $V = U$ and choosing $\bar{\lambda}$ to be an upper bound on the entry λ of a feasible solution of the above inequality system. To that end, it suffices to bound the vertices of the (equivalent standard form) polyhedron defined by the above inequalities in variables $(\boldsymbol{\mu},\boldsymbol{\beta},\alpha,\lambda)$. Lemma 1 can be used to compute such a bound in polynomial time. In particular, it can be shown that one can choose

$$
\bar{\lambda} \geq (nc_2 + n_p + 2)! \cdot \theta_1^{nc_2+n_p+2} \cdot \theta_2 \cdot \theta_3^{nc_2+n_p+1},
$$

   with $\theta_1 = \max\{\|\boldsymbol{x}^\ell\|, \|\boldsymbol{x}^u\|, \|\boldsymbol{y}^\ell_{\mathrm{d}}\|, \|\boldsymbol{y}^u_{\mathrm{d}}\|, 1\}$, $\theta_2 = \max\{U + \|\boldsymbol{C}\|_\infty + \|\boldsymbol{D}_{\mathrm{d}}\|_\infty, \|\boldsymbol{D}_{\mathrm{c}}\|_\infty\}$ and $\theta_3 = \max\{\|\boldsymbol{W}_{\mathrm{d}}\|_\infty + \|\boldsymbol{T}\|_\infty, \|\boldsymbol{W}_{\mathrm{c}}\|, \|\boldsymbol{H}\|, 1\}$. As before, note that this bound does not depend on the chosen $\boldsymbol{x}$, $\boldsymbol{\xi}$ or $\boldsymbol{y}_{\mathrm{d}}$.

The validity of (12) now simply follows by defining $\bar{\lambda}$ to be the maximum of the three bounds obtained in the three disjunctions. □

4.2 Algorithmic Modifications

The positive complexity result from the previous section is mostly of theoretical value. Although one can compute an optimal multiplier in polynomial time, it does not preclude the possibility that verifying optimality of a given multiplier remains computationally intractable. To that end, we propose practical yet simple modifications in cases where a candidate multiplier may be suboptimal, in the context of the Benders decomposition and column-and-constraint generation algorithms presented in [4].

Classical versions of these algorithms obtain upper bounds on the two-stage problem by solving $\sup\{\mathcal{Q}(\boldsymbol{x},\boldsymbol{\xi}) : \boldsymbol{\xi} \in \Xi\}$ for some fixed $\boldsymbol{x} \in \mathcal{X}$. Instead, [4] propose to solve $\sup\{\mathcal{L}(\boldsymbol{x},\boldsymbol{\xi},\lambda) : \boldsymbol{\xi} \in \Xi\}$, where the second-stage value function is replaced by the Lagrangian function. In the absence of sufficient conditions that ensure optimality of λ, it may be possible that the calculated upper bounds are no longer rigorous.

This issue can be addressed by a simple modification. The key idea is to use [4, Theorem 3] which provides *necessary conditions* for the optimality of a Lagrange multiplier. This theorem is exploited in Algorithms 4 and 5 of the original paper, proposed for problems $\mathcal{P}$ and $\mathcal{P}_{\mathcal{I}}$, respectively, which output either an uncertain parameter realization that makes the second-stage problem infeasible or a Lagrange multiplier which satisfies the necessary conditions. These algorithms are then embedded within the corresponding Benders and column-and-constraint generation algorithms. To ensure that the latter do not terminate incorrectly, the proposed modification indirectly verifies optimality of the calculated Lagrange multipler *ex post*.

To simplify exposition and maintain consistency with the original paper, we first illustrate this modification in the context of the Benders decomposition and column-and-constraint generation algorithms for solving formulation $\mathcal{P}_{\mathcal{I}}$ with continuous second-stage decisions ($\mathcal{Y} = \mathbb{R}_+^{n_2}$). In particular, this problem structure arises in the first two numerical experiments of the original paper. We then present the modifications for the more general formulation $\mathcal{P}$ with mixed-integer second-stage decisions, which arises in the third experiment of the original paper as well as the counterexample in Section 2. We omit presenting modifications for problem $\mathcal{P}_{\mathcal{I}}$ with mixed-integer second-stage decisions since it is very similar to the latter. We finally close the paper with a discussion about the computational efficiency of the proposed modifications.

4.2.1 Modifications for Problem $\mathcal{P}_{\mathcal{I}}$

The updated versions of the Benders and column-and-constraint generation algorithms are shown in Algorithm 1. The algorithm indirectly checks if the estimated upper bound is less than the optimal value of the original problem, by solving

$$\begin{array}{ll}\underset{\substack{\boldsymbol{\xi}\in\Xi,\boldsymbol{\mu}\in\mathbb{R}_+^m \\ \boldsymbol{\rho}\in\mathbb{R}_+^{n_p},\boldsymbol{\nu}\in\mathbb{R}_+^{n_p}}}{\text{maximize}} & \boldsymbol{c}(\boldsymbol{\xi})^\top \boldsymbol{x} + (\boldsymbol{h}_0 - \boldsymbol{T}\boldsymbol{x})^\top \boldsymbol{\psi}(\boldsymbol{\mu},\boldsymbol{\rho},\boldsymbol{\nu}) \\ \text{subject to} & \boldsymbol{W}^\top \boldsymbol{\psi}(\boldsymbol{\mu},\boldsymbol{\rho},\boldsymbol{\nu}) \le \boldsymbol{d}(\boldsymbol{\xi}), \\ & \xi_j = 0 \implies \rho_j = 0,\ j \in [n_p], \\ & \xi_j = 1 \implies \nu_j = 0,\ j \in [n_p], \end{array} \tag{13}$$

where $\boldsymbol{\psi}$ is defined as follows:

$$\boldsymbol{\psi}(\boldsymbol{\mu},\boldsymbol{\rho},\boldsymbol{\nu}) = \boldsymbol{\mu} - \sum_{j\in[n_p]} \sum_{i\in\mathcal{I}_j^1} \rho_j \mathbf{e}_i - \sum_{j\in[n_p]} \sum_{i\in\mathcal{I}_j^0} \nu_j \mathbf{e}_i.$$

Algorithm 1 Updated Benders decomposition and column-and-constraint generation algorithms to solve $\mathcal{P}_{\mathcal{I}}$ when $\mathcal{Y} = \mathbb{R}_+^{n_2}$

Benders: Run all lines of [4, Algorithm 6]
Column-and-constraint generation: Run all lines of [4, Algorithm 7]
if $\hat{\boldsymbol{x}} \neq \emptyset$ **then**
  Set Z and $(\hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\nu}})$ as the optimal value and (projected) solution of (13) (at $\boldsymbol{x} = \hat{\boldsymbol{x}}$)
  if $UB < Z$ **then**
    Update $UB \leftarrow Z$ and $\lambda \leftarrow \max\{\|\hat{\boldsymbol{\rho}}\|_\infty, \|\hat{\boldsymbol{\nu}}\|_\infty\}$
    Go to line 2 of the original algorithm
  end if
end if

The following theorem rigorously justifies the proposed modification. It shows that an optimal value of the Lagrange multiplier can be easily computed given an optimal solution of problem (13).

Theorem 4. *Suppose $\mathcal{Y} = \mathbb{R}^{n_2}_+$ and $\boldsymbol{x} \in \mathcal{X}$ is any feasible first-stage decision in problem $\mathcal{P}_{\mathcal{I}}$. Let $(\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\nu}})$ denote an optimal solution of problem* (13)*. Then, $\bar{\lambda} = \max\{\|\hat{\boldsymbol{\rho}}\|_\infty, \|\hat{\boldsymbol{\nu}}\|_\infty\}$ is an optimal multiplier satisfying*

$$\sup_{\boldsymbol{\xi}\in\Xi} \mathcal{Q}_{\mathcal{I}}(\boldsymbol{x}, \boldsymbol{\xi}) = \sup_{\boldsymbol{\xi}\in\Xi} \mathcal{L}_{\mathcal{I}}(\boldsymbol{x}, \boldsymbol{\xi}, \bar{\lambda}).$$

Proof. Let Z denote the optimal value of problem (13). Now, consider problem (13), where $\boldsymbol{\xi}$, $\boldsymbol{\rho}$ and $\boldsymbol{\nu}$ are fixed to $\hat{\boldsymbol{\xi}}$, $\hat{\boldsymbol{\rho}}$ and $\hat{\boldsymbol{\nu}}$, respectively. The resulting problem can be equivalently written as a linear maximization problem over $\boldsymbol{\mu} \in \mathbb{R}^m_+$. Taking the linear programming dual of this problem yields

$$Z = \inf_{\boldsymbol{y}\in\mathcal{Y}(\boldsymbol{x})} \ \boldsymbol{c}(\hat{\boldsymbol{\xi}})^\top \boldsymbol{x} + \boldsymbol{d}(\hat{\boldsymbol{\xi}})^\top \boldsymbol{y} + \sum_{j\in[n_p]} \sum_{i\in\mathcal{I}^1_j} \hat{\rho}_j g_i(\boldsymbol{x}, \boldsymbol{y}) + \sum_{j\in[n_p]} \sum_{i\in\mathcal{I}^0_j} \hat{\nu}_j g_i(\boldsymbol{x}, \boldsymbol{y})$$

where we used the definition of $\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{y} - \boldsymbol{h}_0$ and define $\mathcal{Y}(\boldsymbol{x}) = \{\boldsymbol{y} \in \mathcal{Y} : \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}) \geq \boldsymbol{0}\}$. The definition of $\bar{\lambda}$ along with the indicator constraints in problem (13) imply that the inequalities, $\hat{\rho}_j \leq \bar{\lambda}\hat{\xi}_j$ and $\hat{\nu}_j \leq \bar{\lambda}(1 - \hat{\xi}_j)$, hold for all $j \in [n_p]$. Substituting these inequalities above gives:

$$\begin{aligned}
Z &\leq \inf_{\boldsymbol{y}\in\mathcal{Y}(\boldsymbol{x})} \ \boldsymbol{c}(\hat{\boldsymbol{\xi}})^\top \boldsymbol{x} + \boldsymbol{d}(\hat{\boldsymbol{\xi}})^\top \boldsymbol{y} + \sum_{j\in[n_p]} \sum_{i\in\mathcal{I}^1_j} \bar{\lambda}\hat{\xi}_j g_i(\boldsymbol{x}, \boldsymbol{y}) + \sum_{j\in[n_p]} \sum_{i\in\mathcal{I}^0_j} \bar{\lambda}(1 - \hat{\xi}_j) g_i(\boldsymbol{x}, \boldsymbol{y}) \\
&= \inf_{\boldsymbol{y}\in\mathcal{Y}(\boldsymbol{x})} \ \boldsymbol{c}(\hat{\boldsymbol{\xi}})^\top \boldsymbol{x} + \boldsymbol{d}(\hat{\boldsymbol{\xi}})^\top \boldsymbol{y} + \bar{\lambda}\phi_{\mathcal{I}}(\boldsymbol{x}, \boldsymbol{y}, \hat{\boldsymbol{\xi}}) \\
&\leq \sup_{\boldsymbol{\xi}\in\Xi} \inf_{\boldsymbol{y}\in\mathcal{Y}(\boldsymbol{x})} \ \boldsymbol{c}(\boldsymbol{\xi})^\top \boldsymbol{x} + \boldsymbol{d}(\boldsymbol{\xi})^\top \boldsymbol{y} + \bar{\lambda}\phi_{\mathcal{I}}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\xi}) \\
&= \sup_{\boldsymbol{\xi}\in\Xi} \mathcal{L}_{\mathcal{I}}(\boldsymbol{x}, \boldsymbol{\xi}, \bar{\lambda}) \\
&\leq \sup_{\boldsymbol{\xi}\in\Xi} \mathcal{Q}_{\mathcal{I}}(\boldsymbol{x}, \boldsymbol{\xi}),
\end{aligned}$$

where the first equality holds by definition of $\phi_{\mathcal{I}}$, the next inequality follows by taking the supremum of the previous expression over $\boldsymbol{\xi} \in \Xi$, the second equality follows by definition of $\mathcal{L}_{\mathcal{I}}(\boldsymbol{x}, \boldsymbol{\xi}, \bar{\lambda})$, and the last inequality follows by weak duality [4, Lemma 2].

Now, strong duality [4, Theorem 1] implies that

$$\begin{aligned}
\sup_{\boldsymbol{\xi}\in\Xi} \mathcal{Q}_{\mathcal{I}}(\boldsymbol{x}, \boldsymbol{\xi}) &= \sup_{\lambda\geq 0} \sup_{\boldsymbol{\xi}\in\Xi} \mathcal{L}_{\mathcal{I}}(\boldsymbol{x}, \boldsymbol{\xi}, \lambda) \\
&= \sup_{\boldsymbol{\xi}\in\Xi} \inf_{\boldsymbol{y}\in\mathcal{Y}(\boldsymbol{x})} \ \boldsymbol{c}(\boldsymbol{\xi})^\top \boldsymbol{x} + \boldsymbol{d}(\boldsymbol{\xi})^\top \boldsymbol{y} + \lambda\phi_{\mathcal{I}}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\xi}) \\
&= \begin{bmatrix} \underset{\lambda\geq 0, \boldsymbol{\xi}\in\Xi, \boldsymbol{\mu}\in\mathbb{R}^m_+}{\text{maximize}} & \boldsymbol{c}(\boldsymbol{\xi})^\top \boldsymbol{x} + (\boldsymbol{h}_0 - \boldsymbol{T}\boldsymbol{x})^\top \boldsymbol{\psi}\left(\boldsymbol{\mu}, \lambda\boldsymbol{\xi}, \lambda(\mathbf{e} - \boldsymbol{\xi})\right) \\ \text{subject to} & \boldsymbol{W}^\top \boldsymbol{\psi}\left(\boldsymbol{\mu}, \lambda\boldsymbol{\xi}, \lambda(\mathbf{e} - \boldsymbol{\xi})\right) \leq \boldsymbol{d}(\boldsymbol{\xi}) \end{bmatrix},
\end{aligned}$$

where the last equality follows by linear programming duality. Let $(\tilde{\lambda}, \tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\mu}})$ be an optimal solution of the above (bilinear) optimization problem. Define $\tilde{\boldsymbol{\rho}} = \tilde{\lambda}\tilde{\boldsymbol{\xi}}$ and $\tilde{\boldsymbol{\nu}} = \tilde{\lambda}(\mathbf{e} - \tilde{\boldsymbol{\xi}})$. Then, it can be readily verified that $(\tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\rho}}, \tilde{\boldsymbol{\nu}})$ is a feasible solution in problem (13) that achieves an objective value equal to the optimal value of the above bilinear problem. This implies

$$\sup_{\boldsymbol{\xi}\in\Xi} \mathcal{Q}_{\mathcal{I}}(\boldsymbol{x}, \boldsymbol{\xi}) \leq Z,$$

which taken together with our previously established inequality,

$$Z \leq \sup_{\boldsymbol{\xi}\in\Xi} \mathcal{L}_{\mathcal{I}}(\boldsymbol{x}, \boldsymbol{\xi}, \bar{\lambda}) \leq \sup_{\boldsymbol{\xi}\in\Xi} \mathcal{Q}_{\mathcal{I}}(\boldsymbol{x}, \boldsymbol{\xi}),$$

proves the claimed result. ∎

Algorithm 1 checks if the optimal value of problem (13) is larger than the final estimate UB, and then uses the result of Theorem 4 to initialize another run of the original procedure with the updated λ and corrected UB. In doing so, it retains all data structures without re-initializing them to be empty sets. In particular, for the Benders algorithm, the feasibility and optimality sets, $\mathcal{F}$ and $\mathcal{O}$, are retained, and all previously generated Benders cuts are simply lifted with the updated value of λ. Similarly, for the column-and-constraint generation algorithm, the set of enumerated uncertain parameters $\mathcal{R}$ is retained. It is crucial to higlight that in both algorithms, all previously generated constraints continue to remain valid, and therefore, they always provide rigorous lower bounds on the optimal value of the original two-stage problem. In particular, this is also true for Benders cuts generated using suboptimal λ values. Formally, this is because of weak duality [4, Lemma 2], which implies:

$$\inf_{\boldsymbol{x}\in\mathcal{X}} \sup_{\boldsymbol{\xi}\in\Xi} \mathcal{Q}_{\mathcal{I}}(\boldsymbol{x},\boldsymbol{\xi}) \geq \inf_{\boldsymbol{x}\in\mathcal{X}} \sup_{\boldsymbol{\xi}\in\Xi} \mathcal{L}_{\mathcal{I}}(\boldsymbol{x},\boldsymbol{\xi},\lambda) \quad \text{for all } \lambda \geq 0.$$

We note that if Algorithm 5 of the original paper (which is invoked within the original Algorithms 6 and 7) outputs an optimal multiplier, then problem (13) is solved at most once. This is important for reasons of computational efficiency, which we discuss at the end of the paper.

4.2.2 Modifications for Problem $\mathcal{P}$

We now consider the general version of $\mathcal{P}$ with mixed-integer second-stage decisions. Keeping in line with the original paper, we focus only on the column-and-constraint generation algorithm. We first provide an updated version of the original method from [4] in Algorithm 2. Whereas the original method assumes that sufficient conditions for optimality of λ are already satisfied, Algorithm 2 does not make any such assumption.

As described in the original paper, the key idea of the method is that for fixed $\boldsymbol{x}\in\mathcal{X}$, one can obtain a relaxation of the worst-case Lagrangian function, $\sup\{\mathcal{L}(\boldsymbol{x},\boldsymbol{\xi},\lambda) : \boldsymbol{\xi}\in\Xi\}$, by enumerating the set of discrete second-stage decisions $\mathcal{Y}_{\mathrm{d}}$. In particular, if $\mathcal{D}\subseteq\mathcal{Y}_{\mathrm{d}}$, then it follows that

$$\sup_{\boldsymbol{\xi}\in\Xi} \mathcal{L}(\boldsymbol{x},\boldsymbol{\xi},\lambda) = \sup_{\boldsymbol{\xi}\in\Xi} \inf_{\boldsymbol{y}_{\mathrm{d}}\in\mathcal{Y}_{\mathrm{d}}} \mathcal{L}(\boldsymbol{x},\boldsymbol{\xi},\lambda;\boldsymbol{y}_{\mathrm{d}}) \leq \sup_{\boldsymbol{\xi}\in\Xi} \inf_{\boldsymbol{y}_{\mathrm{d}}\in\mathcal{D}} \mathcal{L}(\boldsymbol{x},\boldsymbol{\xi},\lambda;\boldsymbol{y}_{\mathrm{d}}).$$

Now, if $\boldsymbol{y}_{\mathrm{d}}^{(k)}\in\mathcal{D}$ denotes the k^{th} element of $\mathcal{D}$, then it can be shown [5, 4] that the the following (with $\tau=1$) is a reformulation of the problem on the right-hand side of the above inequality.

$$\begin{array}{ll}
\underset{\eta,\boldsymbol{\xi},\boldsymbol{\mu},\boldsymbol{\beta}}{\text{maximize}} & \eta \\
\text{subject to} & \eta\in\mathbb{R},\ \ \boldsymbol{\xi}\in\Xi, \\
& \left.\begin{array}{l}
\boldsymbol{\mu}^{(k)}\in\mathbb{R}^{m}_{+},\ \ \boldsymbol{\beta}^{(k)}\in\mathbb{R}^{n_p}_{+}, \\
\eta \leq \tau\boldsymbol{c}(\boldsymbol{\xi})^\top\boldsymbol{x} + \tau\boldsymbol{d}_{\mathrm{d}}(\boldsymbol{\xi})^\top\boldsymbol{y}_{\mathrm{d}}^{(k)} + \mathbf{e}^\top(\lambda\boldsymbol{\xi}-\boldsymbol{\beta}^{(k)}) \\
\qquad + (\boldsymbol{h}_0 - \boldsymbol{T}\boldsymbol{x} - \boldsymbol{W}_{\mathrm{d}}\boldsymbol{y}_{\mathrm{d}}^{(k)})^\top\boldsymbol{\mu}^{(k)}, \\
\boldsymbol{W}_{\mathrm{c}}^\top\boldsymbol{\mu}^{(k)} \leq \tau\boldsymbol{d}_{\mathrm{c}}(\boldsymbol{\xi}),\ \ (1-\tau)\boldsymbol{\mu}^{(k)}\leq\mathbf{e}, \\
2\lambda\boldsymbol{\xi} - \boldsymbol{H}^\top\boldsymbol{\mu}^{(k)} - \boldsymbol{\beta}^{(k)} \leq \lambda\mathbf{e},
\end{array}\right\}\ k\in[|\mathcal{D}|].
\end{array} \tag{14}$$

The parameter $\tau\in\{0,1\}$ serves purely to simplify notation. Indeed, when $\tau=0$, it can be shown that problem (14) also provides an upper bound on the worst-case constraint violation function. The latter is equal to 0 if and only if the first-stage decision $\boldsymbol{x}\in\mathcal{X}$ is feasible in problem $\mathcal{P}$, see [4] for further details.

Algorithm 2 Updated version of [4, Algorithm 8]

Input: $\boldsymbol{x} \in \mathcal{X}$, $\lambda^0 > 0$
Output: Either $\hat{\boldsymbol{\xi}} \in \Xi : \mathcal{Q}(\boldsymbol{x}, \hat{\boldsymbol{\xi}}) = +\infty$, $\hat{\lambda} = \lambda^0$ or $\hat{\boldsymbol{\xi}}, \hat{\lambda}$ satisfying conditions of [4, Theorem 3]

Initialize $\hat{\boldsymbol{\xi}} \in \Xi$ (arbitrary), $LB = -\infty$, $UB = +\infty$, $\mathcal{D} = \emptyset$, $\hat{\lambda} = \lambda^0$.
repeat
  Set UB and $\hat{\boldsymbol{\xi}}$ as optimal value and (projected) solution of (14) with $(\tau, \lambda) = (0, 1)$
  Set $(\hat{\boldsymbol{y}}, \hat{\boldsymbol{z}}, \hat{\boldsymbol{\sigma}}) \in \underset{(\boldsymbol{y},\boldsymbol{z},\boldsymbol{\sigma}) \in \mathcal{Y} \times [0,1]^{n_p} \times \mathbb{R}^m_+}{\arg\min} \left\{ \mathbf{e}^\top \boldsymbol{\sigma} + \phi(\boldsymbol{z}, \hat{\boldsymbol{\xi}}) : \boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{y} + \boldsymbol{\sigma} \geq \boldsymbol{h}(\boldsymbol{z}) \right\}$
  Update $\mathcal{D} \leftarrow \mathcal{D} \cup \{\hat{\boldsymbol{y}}_\mathrm{d}\}$
  Update $LB = \mathbf{e}^\top \hat{\boldsymbol{\sigma}} + \phi(\hat{\boldsymbol{z}}, \hat{\boldsymbol{\xi}})$
until $LB > 0$ or $UB = 0$
if $UB = 0$ **then**
  Set $LB = -\infty$
  repeat
    Update $\hat{\lambda} \leftarrow \hat{\lambda}/2$
    repeat
      Update $\hat{\lambda} \leftarrow 2\hat{\lambda}$
      Set UB, $\tilde{\boldsymbol{\xi}}$ as optimal value and (projected) solution of (14) with $(\tau, \lambda) = (1, \hat{\lambda})$
      Set $(\hat{\boldsymbol{y}}, \hat{\boldsymbol{z}}) \in \underset{(\boldsymbol{y},\boldsymbol{z}) \in \mathcal{Y} \times [0,1]^{n_p}}{\arg\min} \left\{ \boldsymbol{c}(\tilde{\boldsymbol{\xi}})^\top \boldsymbol{x} + \boldsymbol{d}(\tilde{\boldsymbol{\xi}})^\top \boldsymbol{y} + \hat{\lambda}\phi(\boldsymbol{z}, \tilde{\boldsymbol{\xi}}) : \boldsymbol{T}\boldsymbol{x} + \boldsymbol{W}\boldsymbol{y} \geq \boldsymbol{h}(\boldsymbol{z}) \right\}$
      Update $\mathcal{D} \leftarrow \mathcal{D} \cup \{\hat{\boldsymbol{y}}_\mathrm{d}\}$
    until $\phi(\hat{\boldsymbol{z}}, \tilde{\boldsymbol{\xi}}) = 0$
    if $LB < \boldsymbol{c}(\tilde{\boldsymbol{\xi}})^\top \boldsymbol{x} + \boldsymbol{d}(\tilde{\boldsymbol{\xi}})^\top \hat{\boldsymbol{y}}$ **then** update $LB \leftarrow \boldsymbol{c}(\tilde{\boldsymbol{\xi}})^\top \boldsymbol{x} + \boldsymbol{d}(\tilde{\boldsymbol{\xi}})^\top \hat{\boldsymbol{y}}$ and $\hat{\boldsymbol{\xi}} \leftarrow \tilde{\boldsymbol{\xi}}$ **end if**
  until $UB - LB \leq \epsilon$
end if

Algorithm 2 is almost identical to the original [4, Algorithm 8], except for the inner loop based on the condition appearing in line 17. The latter enforces necessary conditions for optimality of λ based on [4, Theorem 3]. Although these conditions are not sufficient, it is still possible that the final λ is optimal, as illustrated in the following example.

Example 2 (Counterexample revisited)**.** *We illustrate Algorithm 2 on the counterexample from Section 2, with inputs $x = 0$ and $\lambda^0 = u(x) - \ell(x) = 1$. Let $y^{(k)} \in \{0, 1\}$ for $k \in \{1, 2\}$. Then, problem (14) reduces to:*

$$\begin{aligned}
&\underset{\xi,\eta,\mu,\beta}{\text{maximize}} \;\; \eta \\
&\text{subject to} \;\; \xi \in \{0,1\}, \quad \eta \in \mathbb{R}, \\
&\qquad \left.\begin{aligned}
&\mu^{(k)} \in \mathbb{R}_+, \quad \beta^{(k)} \in \mathbb{R}_+, \\
&\eta \leq -\tau y^{(k)} + \left(y^{(k)} - \frac{3}{2}\right)\mu^{(k)} - \beta^{(k)} + \lambda\xi, \\
&2\lambda\xi - \mu^{(k)} - \beta^{(k)} \leq \lambda,
\end{aligned}\right\} \quad k \in \{1,2\}.
\end{aligned}$$

The various optimization problems solved in the algorithm often exhibit multiple optimal solutions. Straightforward computations (omitted for the sake of brevity) reveal that the algorithm will exit the first 'repeat' loop with one of the following outcomes: $\mathcal{D} = \{0, 1\}$ after three iterations; $\mathcal{D} = \{0, 1\}$ after two iterations; $\mathcal{D} = \{0\}$ after two iterations. In all outcomes, $LB = UB = 0$, indicating that the original two-stage problem is feasible.

1. *Suppose $D = \{0, 1\}$. Line 14 yields $UB = -0.5$ and $\tilde{\xi} = 1$ as the unique solution. The next line yields $(\hat{y}, \hat{z}) = (1, 0.5)$ that is also unique. Since $\phi(\hat{z}, \hat{\xi}) = 0.5$, the algorithm updates $\hat{\lambda} = 2$. Line 14 then yields $UB = 0$ and $\tilde{\xi} = 1$ (unique). The next line can yield one of two optimal solutions, $(\hat{y}, \hat{z}) = (0, 1)$ or $(\hat{y}, \hat{z}) = (1, 0.5)$.*

    - *If $(\hat{y}, \hat{z}) = (0, 1)$, then Line 18 updates $LB = 0$. The algorithm stops with $\hat{\lambda} = 2$ and $LB = UB = 0$.*
    - *If $(\hat{y}, \hat{z}) = (1, 0.5)$, then since $\phi(\hat{z}, \hat{\xi}) = 0.5$, the algorithm updates $\hat{\lambda} = 4$. Line 14 yields $UB = 0$ and $\tilde{\xi} = 1$ (unique). The next line yields $(\hat{y}, \hat{z}) = (0, 1)$ as the unique solution. Line 18 updates $LB = 0$. The algorithm stops with $\hat{\lambda} = 4$ and $LB = UB = 0$.*

2. *Suppose $D = \{0\}$. Line 14 yields $UB = 0$. Also, the optimal $\tilde{\xi} \in \{0, 1\}$.*

    - *If $\tilde{\xi} = 0$, then the next line yields $(\hat{y}, \hat{z}) = (1, 0)$ (unique). The algorithm updates $\mathcal{D} = \{0, 1\}$. This is the same state as the beginning of case 1.*
    - *If $\tilde{\xi} = 1$, then the next line yields $(\hat{y}, \hat{z}) = (1, 0.5)$ (unique). Since $\phi(\hat{z}, \hat{\xi}) = 0.5$, the algorithm updates $\mathcal{D} = \{0, 1\}$ and $\hat{\lambda} = 2$. This state of the algorithm is also reached in case 1.*

In all outcomes, the final $\hat{\lambda} \in \arg\max_{\lambda \in \mathbb{R}_+} \{\max_{\xi \in \Xi} \mathcal{L}(x, \xi, \lambda)\}$ (see Figure 1) and the algorithm terminates correctly with $0 = LB = UB = \max_{\xi \in \Xi} \mathcal{Q}(x, \xi)$.

Despite the positive result in the above example, it is nevertheless possible that the final value of λ is suboptimal. A suboptimal choice of λ may lead to an invalid upper bound and a premature termination of the algorithm. As before, we can circumvent this *ex post* by indirectly checking if a better choice of λ can lead to higher upper bound. In particular, after running the entire column-and-constraint generation algorithm to obtain the (candidate) optimal first-stage decisions $\boldsymbol{x}$, one can solve problem (15), shown below.

$$
\begin{array}{ll}
\underset{\eta,\boldsymbol{\xi},\boldsymbol{\mu},\boldsymbol{\rho}}{\text{maximize}} & \eta \\
\text{subject to} & \eta \in \mathbb{R},\ \ \boldsymbol{\xi} \in \Xi, \\
& \left.\begin{array}{l}
\boldsymbol{\mu}^{(k)} \in \mathbb{R}^m_+,\ \ \boldsymbol{\rho}^{(k)} \in \mathbb{R}^{n_p}_+, \\
\eta \leq \boldsymbol{c}(\boldsymbol{\xi})^\top \boldsymbol{x} + \boldsymbol{d}_\mathrm{d}(\boldsymbol{\xi})^\top \boldsymbol{y}_\mathrm{d}^{(k)} + \mathbf{e}^\top \boldsymbol{\rho}^{(k)}, \\
\quad + (\boldsymbol{h}_0 - \boldsymbol{T}\boldsymbol{x} - \boldsymbol{W}_\mathrm{d}\boldsymbol{y}_\mathrm{d}^{(k)})^\top \boldsymbol{\mu}^{(k)}, \\
\boldsymbol{W}_\mathrm{c}^\top \boldsymbol{\mu}^{(k)} \leq \boldsymbol{d}_\mathrm{c}(\boldsymbol{\xi}), \\
\xi_j = 1 \implies \rho_j^{(k)} \leq \mathbf{e}_j^\top \boldsymbol{H}^\top \boldsymbol{\mu}^{(k)},\ \ j \in [n_p], \\
\xi_j = 0 \implies \rho_j^{(k)} \leq 0,\ \ j \in [n_p],
\end{array}\right\}\ k \in [|\mathcal{D}|].
\end{array} \tag{15}
$$

Similar to Algorithm 1, if the optimal value of the above problem happens to be strictly larger than the final estimate UB, then we can simply uses it optimal solution to update λ (see Theorem 5 below) and restart the column-and-constraint generation algorithm. In doing so, we can retain the enumerated set $\mathcal{D}$ without re-initializing it to be empty. As in the case of problem $\mathcal{P}_\mathcal{I}$, we higlight that all previously generated constraints will continue to remain valid. Similarly, if the loop terminating in line 17 outputs an optimal multiplier, then problem (15) will be solved only at most once during the entire algorithm.

The following theorem is the counterpart of Theorem 4 for problem $\mathcal{P}$.

Theorem 5. *Suppose $\mathcal{D} \subseteq \mathcal{Y}_\mathrm{d}$ and $\boldsymbol{x} \in \mathcal{X}$ is any feasible first-stage decision in problem $\mathcal{P}$. Let $(\hat{\eta}, \hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\rho}})$ denote an optimal solution of problem* (13)*. Then, $\bar{\lambda} = \max_{k \in [|\mathcal{D}|]} \left\{ \|\hat{\boldsymbol{\rho}}^{(k)}\|_\infty, \|\boldsymbol{H}^\top \hat{\boldsymbol{\mu}}^{(k)}\|_\infty \right\}$ is an optimal multiplier satisfying*

$$\sup_{\boldsymbol{\xi}\in\Xi} \inf_{\boldsymbol{y}_\mathrm{d}\in\mathcal{D}} \mathcal{Q}(\boldsymbol{x}, \boldsymbol{\xi}; \boldsymbol{y}_\mathrm{d}) = \sup_{\boldsymbol{\xi}\in\Xi} \inf_{\boldsymbol{y}_\mathrm{d}\in\mathcal{D}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\xi}, \bar{\lambda}; \boldsymbol{y}_\mathrm{d}).$$

Proof. Define $\hat{\boldsymbol{\beta}}^{(k)} = \bar{\lambda}\hat{\boldsymbol{\xi}} - \hat{\boldsymbol{\rho}}^{(k)}$ for $k \in [|\mathcal{D}|]$. We claim that $(\bar{\lambda}, \hat{\eta}, \hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\beta}})$ is feasible in problem (14) with $(\tau, \lambda) = (1, \bar{\lambda})$. To see why, first note that if $\hat{\xi}_j = 1$, then $-\hat{\rho}_j^{(k)} \geq 0$ by definition of $\hat{\boldsymbol{\rho}}^{(k)}$ and if $\hat{\xi}_j = 1$, then $\bar{\lambda} - \hat{\rho}_j^{(k)} \geq 0$ by definition of $\bar{\lambda}$. The two cases together imply $\hat{\boldsymbol{\beta}}^{(k)} \geq \mathbf{0}$. Next, note that the expression, $2\bar{\lambda}\hat{\boldsymbol{\xi}} - \boldsymbol{H}^\top \hat{\boldsymbol{\mu}}^{(k)} - \hat{\boldsymbol{\beta}}^{(k)} = \hat{\boldsymbol{\rho}}^{(k)} - \boldsymbol{H}^\top \hat{\boldsymbol{\mu}}^{(k)}$, satisfies $\hat{\rho}_j^{(k)} - \mathbf{e}_j^\top \boldsymbol{H}^\top \hat{\boldsymbol{\mu}}^{(k)} \leq -\mathbf{e}_j^\top \boldsymbol{H}^\top \hat{\boldsymbol{\mu}}^{(k)}$ whenever $\hat{\xi}_j = 0$, and it satisfies $\hat{\rho}_j^{(k)} - \mathbf{e}_j^\top \boldsymbol{H}^\top \hat{\boldsymbol{\mu}}^{(k)} \leq 0$ whenever $\hat{\xi}_j = 1$. These two cases together with the definition of $\bar{\lambda}$ imply $2\bar{\lambda}\hat{\boldsymbol{\xi}} - \boldsymbol{H}^\top \hat{\boldsymbol{\mu}}^{(k)} - \hat{\boldsymbol{\beta}}^{(k)} \leq \bar{\lambda}\mathbf{e}$. In summary, $(\bar{\lambda}, \hat{\eta}, \hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\beta}})$ is feasible in problem problem (14) with $(\tau, \lambda) = (1, \bar{\lambda})$. The optimal value of the latter problem is precisely equal (by construction) to the right-hand side problem of the equation stated in this theorem. If Z denotes the optimal value of problem (15), then this implies

$$Z \leq \sup_{\boldsymbol{\xi}\in\Xi} \inf_{\boldsymbol{y}_\mathrm{d}\in\mathcal{D}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\xi}, \bar{\lambda}; \boldsymbol{y}_\mathrm{d}).$$

Define now $\mathcal{T}$ to be the (bilinear) optimization problem, which is identical to problem (14) with $\tau = 1$ and the addition of $\lambda \geq 0$ as a decision variable. By construction, its optimal value $Z_\mathcal{T}$ satisfies the relation,

$$
\begin{aligned}
Z \leq \sup_{\boldsymbol{\xi}\in\Xi} \inf_{\boldsymbol{y}_\mathrm{d}\in\mathcal{D}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\xi}, \bar{\lambda}; \boldsymbol{y}_\mathrm{d}) \leq Z_\mathcal{T} &= \sup_{\lambda\geq 0} \sup_{\boldsymbol{\xi}\in\Xi} \inf_{\boldsymbol{y}_\mathrm{d}\in\mathcal{D}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\xi}, \lambda; \boldsymbol{y}_\mathrm{d}) \\
&= \sup_{\boldsymbol{\xi}\in\Xi} \inf_{\boldsymbol{y}_\mathrm{d}\in\mathcal{D}} \mathcal{Q}(\boldsymbol{x}, \boldsymbol{\xi}; \boldsymbol{y}_\mathrm{d}),
\end{aligned}
$$

where the first inequality was established in the previous paragraph and the second equality follows by strong duality [4, Theorem 1]. Observe now that if we have $Z \geq Z_{\mathcal{T}}$, then the equation in the theorem follows. To see why $Z \geq Z_{\mathcal{T}}$, let $(\tilde{\lambda}, \tilde{\eta}, \tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\beta}})$ be a feasible solution in problem $\mathcal{T}$ with objective value $\tilde{\eta}$. Then, an identical line of argument as before establishes that $(\tilde{\eta}, \tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\rho}})$ is feasible in problem (15) with the same objective value, where we define $\tilde{\boldsymbol{\rho}}^{(k)} = \bar{\lambda}\tilde{\boldsymbol{\xi}} - \tilde{\boldsymbol{\beta}}^{(k)}$ for all $k \in [|\mathcal{D}|]$. □

4.2.3 Impact on Computational Performance and Practical Considerations

Although the proposed modifications ensure correctness of the overall method and its finite termination to an optimal solution of the two-stage problem, they are computationally inefficient. This is because of the presence of the indicator constraints in problems (13) and (15). Indeed, the main motivation of the original paper [4] in developing their Lagrangian method was to circumvent the use of indicator constraints (and arbitrary upper bounds on dual variables), which were common in previous solution methods. Indicator constraints are supported by very few solvers and their presence often significantly slows down the overall search process. The lack of indicator constraints and arbitrary upper bounds directly contribute to the large computational speedups of the Lagrangian method over other solution methods. Problems (13) and (15) should therefore be solved as few times as possible.

With a goal toward offering practical guidelines, we perform experiments across the entire set of 378 benchmark problems from [4]. This includes Benders decomposition for solving instances of $\mathcal{P}_{\mathcal{I}}$ and column-and-constraint generation for solving instances of $\mathcal{P}$ and $\mathcal{P}_{\mathcal{I}}$, with both continuous and mixed-integer decisions, across three problem classes. The first two problem classes (network design and facility location) are instances of $\mathcal{P}_{\mathcal{I}}$ with continuous second-stage decisions, whereas the third (staff rostering) is an instance of $\mathcal{P}$ with mixed-integer second-stage decisions.

None of the three problem classes satisfy sufficient conditions that allow closed-form expressions of the optimal Lagrange multiplier. The original paper explicitly acknowledges this for the first two problem classes but incorrectly assumes that the third class satisfies the sufficient conditions. We therefore re-run the latter class of instances using the updated Algorithm 2. In doing so, we specify an initial multiplier value of $\lambda^0 = u(\boldsymbol{x}) - \ell(\boldsymbol{x})$. Interestingly, we find that this initial multiplier value already satisfies the necessary conditions in line 17 of Algorithm 2 across all runs of all instances. Without the *ex post* verification of the optimality of λ, the number of iterations and computational times across the entire set of benchmark instances therefore remains unchanged compared to what was originally reported in [4].

We now examine the potential suboptimality of the final Lagrange multiplier across all problem instances. The *ex post* solution of problems (13) and (15) reveals that their optimal values are equal to the final upper bounds computed using the original methods in 377 out of 378 benchmark problems. In other words, the final Lagrange multipliers computed in [4] are provably optimal across all of these instances. Moreover, the objective values of the corresponding final solutions rigorously upper bound (or equal, if terminated within the time limit) the optimal value of the original two-stage problem. Moreover, in the only case where this was not true[1], the estimated upper bound was less than its true optimal value by less than 0.3%.

The solution times of the indicator constrained optimization problems (13) and (15) are slower than their counterparts with fixed λ. Compared to the latter, their solution times are slower by a (geometric average) factor of 11.9 (network design), 1.3 (facility location), and 2.8 (staff rostering). However, because the indicator constrained problems end up being solved only once during the

[1]Network design instance ‘di-yuan’ with budget parameter $k = 3$.

entire algorithm, this slower solution time has a negligible effect on the overall computational times, which remain similar to those originally reported in [4].

We offer the following conclusions based on our observations. First, the necessary conditions that are enforced within the algorithms of the original paper appear to be almost sufficient in experiments, hinting at opportunities to generalize the conditions in Theorems 1 and 2. Moreover, the computational speedups offered by the original algorithms over traditional methods are only possible because of their use of Lagrangian functions (with fixed λ) as proxies for the second-stage value function. We showed how to check if these proxies are rigorous by solving alternate but computationally difficult optimization problems. We therefore suggest to solve these only at the end to assess the potential suboptimality of the final first-stage decisions. If the optimality gap is not satisfactorily small, then one can obtain better solutions by warm-starting the original algorithms in the manner we showed in this paper.

Acknowledgements

The first author acknowledges support by the German Bundesministerium für Bildung und Forschung within the project "RODES" (Förderkennzeichen 05M22UTB).

References

[1] Christoph Buchheim. Bilevel linear optimization belongs to NP and admits polynomial-size KKT-based reformulations. *Operations Research Letters*, 51(6):618–622, 2023.

[2] Henri Lefebvre, Enrico Malaguti, and Michele Monaci. Adjustable Robust Optimization with Discrete Uncertainty. *INFORMS Journal on Computing*, 36(1):78–96, 2024.

[3] Henri Lefebvre, Martin Schmidt, and Johannes Thürauf. Using Column Generation in Column-and-Constraint Generation for Adjustable Robust Optimization. Available at https://optimization-online.org/?p=24462, 2024.

[4] Anirudh Subramanyam. A Lagrangian dual method for two-stage robust optimization with binary uncertainties. *Optimization and Engineering*, 23(4):1831–1871, 2022.

[5] Bo Zeng and Long Zhao. Solving two-stage robust optimization problems using a column-and-constraint generation method. *Operations Research Letters*, 41(5):457–461, 2013.