An Exact Approach for Convex Adjustable Robust Optimization

Henri Lefebvre[1], Enrico Malaguti[1], and Michele Monaci[1]

Dipartimento di Ingegneria dell'Energia Elettrica e dell'Informazione "Guglielmo Marconi", Università di Bologna, Viale del Risorgimento, 2, Bologna, 40136, BO, Italy {henri.lefebvre,enrico.malaguti,michele.monaci}@unibo.it

Abstract. Adjustable Robust Optimization (ARO) is a paradigm for facing uncertainty in a decision problem, in case some recourse actions are allowed after the actual value of all input parameters is revealed. While several approaches have been introduced for the linear case, little is known regarding exact methods for the convex case. In this work, we introduce a new general method for solving ARO problems involving convex functions in the recourse problem. We first recall a semi-infinite reformulation of the problem and we show how to solve it by row generation, provided that one can tackle a non-convex separation problem. For the relevant case where the uncertainty set has an affine mapping to a 0-1 set, we show that the separation problem can be reformulated as a convex MINLP, thus allowing us to derive a computationally sound exact method. We then study the convergence of the obtained algorithm and apply it to a resource planning problem.

Keywords: Robust Optimization · Two-Stage Optimization · Convex Optimization.

1 Introduction

Adjustable Robust Optimization (ARO) is a paradigm used to face uncertainty in case some recourse actions are allowed after the actual value of all input parameters is revealed. An ARO problem can be formulated as

$$\inf_{\boldsymbol{x}\in X} \sup_{\boldsymbol{\xi}\in\Xi} \inf_{\boldsymbol{y}\in Y(\boldsymbol{x},\boldsymbol{\xi})} g_0(\boldsymbol{x},\boldsymbol{y}), \tag{1}$$

where X denotes the feasible set of decisions to be taken here-and-now (first-stage decisions), Ξ is the uncertainty set, and $Y(\boldsymbol{x},\boldsymbol{\xi})$ is the set of all feasible recourse actions (second-stage decisions) for a given $\boldsymbol{x} \in X$ and $\boldsymbol{\xi} \in \Xi$. In this paper we consider a broad class of ARO problems, characterised by convex objective function g_0 and second stage feasible set.

1.1 Literature review

Linear case ARO problems are known to be intractable even in the simple case where only linear functions appear in the constraints defining $Y(\boldsymbol{x},\boldsymbol{\xi})$ (see [2]).

For this setting, several exact approaches have been introduced in the literature. A first line of research is based on an adaptation of the Benders' decomposition algorithm, where the second-stage problem is dualized and an epigraph reformulation of the remaining maximization problem is used. Then, the obtained "Benders' cuts" are dynamically generated (see, e.g., [22], [7], [15] and [12]). An alternative approach is the column-and-constraint generation algorithm proposed by [23]. In this scheme, a restricted master problem is iteratively solved and augmented by introducing second-stage variables and constraints associated with harmful scenarios. The identification of such scenarios requires the solution of a bilevel problem. Later, [1] solve this bilevel problem by means of an MILP obtained by exploiting a description of the uncertainty set in terms of its extreme points. Finally, the Fourier-Motzkin elimination technique is used in [24] to remove the second-stage decisions from the definition of (1) and solve the problem to optimality.

Given the complexity of linear ARO problems, approximation methods have been introduced, including Finite Adaptability (see, e.g., [5], [14], [20]), where a fixed number of second-stage decisions must be identifies in the first stage so as to address any possible realization of the uncertainty set; and the Affine Decision Rule approach introduced by [2], where second-stage decisions are restricted to be affine functions of the uncertainty, see, e.g., [6] and [4].

Convex case Regarding the convex case, occurring when $Y(\boldsymbol{x},\boldsymbol{\xi})$ is defined by convex functions of the first- and second- stage variables, the techniques introduced for the linear case are not directly applicable, and the scientific literature is much more sparse.

Exact approaches focus on specific settings: in [10], the authors consider ARO problems with ellipsoidal uncertainty and conic quadratic second-stage constraints. Then, [21] considers ARO problems where the uncertainty set is expressed as the convex hull of a finite set of points and report conditions under which their problem can be reduced to a single-stage problem. Finally, [19] derive a dualized problem for a class of convex ARO by extending an approach from [8]. This allows to exploit ADR techniques to obtain a tight approximation for the objective value.

1.2 Problem formulation

Problem definition Let n_X, n_Ξ, n_Y and m be given natural numbers. We let the first-stage feasible set X be any subset of $\mathbb{R}^{n_X}$ and assume that the uncertainty set $\Xi \subseteq \mathbb{R}^{n_\Xi}$ is defined as

$$\Xi = \left\{ \boldsymbol{\xi} \in \mathbb{R}^{n_\Xi}_+ : \boldsymbol{U}\boldsymbol{\xi} \le \boldsymbol{d} \right\} \tag{2}$$

where $\boldsymbol{U} \in \mathbb{R}^{q\times n_\Xi}$ and $\boldsymbol{d} \in \mathbb{R}^{n_q}$. As is customary, we write matrices and vectors in bold case.

For each $i = 1, \dots, m$ and $j = 1, \dots, n_\Xi$, we let $f_{ij} : \mathbb{R}^{n_X} \to \mathbb{R}$ be given real-valuated convex functions; for a given $\boldsymbol{x} \in \mathbb{R}^{n_X}$, we denote by $\boldsymbol{F}(\boldsymbol{x})$ the $m \times n_\Xi$

matrix whose generic element is $f_{ij}(\boldsymbol{x})$. Similarly, for each $i = 1, \dots, m$, we let $g_i : \mathbb{R}^{n_X + n_Y} \to \mathbb{R}$ be a convex function; for a given $\boldsymbol{x} \in \mathbb{R}^{n_X}$ and $\boldsymbol{y} \in \mathbb{R}^{n_Y}$, we denote by $\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y})$ the m-dimensional vector whose generic element is $g_i(\boldsymbol{x}, \boldsymbol{y})$.

For given $\boldsymbol{x} \in X$ and $\boldsymbol{\xi} \in \Xi$, the second-stage feasible space $Y(\boldsymbol{x}, \boldsymbol{\xi})$ is defined by

$$Y(\boldsymbol{x}, \boldsymbol{\xi}) = \{\boldsymbol{y} \in \mathbb{R}^{n_Y} : \boldsymbol{F}(\boldsymbol{x})\boldsymbol{\xi} + \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}) \leq \boldsymbol{0}\}. \tag{3}$$

Finally, we assume that the objective function $g_0 : \mathbb{R}^{n_X + n_Y} \to \mathbb{R}$ is a real-valuated convex function. In other words, we consider the class of problems of type 1 where the uncertainty set has a polyhedral representation and second stage optimization is modelled as a convex problem.

Additionally, we make the following technical assumption.

Assumption 1 *For every $(x_0, \boldsymbol{x}) \in \mathbb{R} \times X$ and every $\boldsymbol{\xi} \in \Xi$, the following set*

$$Z(x_0, \boldsymbol{x}, \boldsymbol{\xi}) = \left\{ \begin{pmatrix} \beta_0 \\ \boldsymbol{\beta} \end{pmatrix} \in \mathbb{R}^{m+1} : \exists \boldsymbol{y} \in \mathbb{R}^{n_Y}, \begin{array}{ll} g_0(\boldsymbol{x}, \boldsymbol{y}) - x_0 & \leq \beta_0 \\ \boldsymbol{F}(\boldsymbol{x})\boldsymbol{\xi} + \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}) & \leq \boldsymbol{\beta} \end{array} \right\} \tag{4}$$

is closed.

Contributions As discussed in the literature review, few exact methods for convex ARO have been proposed so far, mostly relying on strong assumptions. In this work, we first recall a quite general reformulation for ARO problems, involving an exponential number of constraints. Then we start filling the literature gap by showing that separation of those constraints can be performed through the solution of a non-convex program. This result, obtained through the use of convex conjugates, can be applied to any convex ARO, including cases where the second stage is an SOCP, an SDP, or a (conic) LP. In addition, we introduce a Generalized-Benders-Decomposition-type algorithm [13] for which we prove finite ε-convergence. We show that, when the uncertainty set has some structural properties, a suitable *convex* reformulation of the separation problem can be used, thus allowing to derive a computationally sound algorithm built on top of a general-purpose solver. Finally, we give the computational evidence of the applicability of our solution method to an uncertain planning application from the literature.

Notation We conclude this section by introducing some notation. For a given function $f : \mathbb{R}^n \to \mathbb{R}$, we denote by f^* its convex conjugate defined as

$$f^*(\boldsymbol{y}) = \sup_{\boldsymbol{x} \in \mathrm{dom}(f)} \left\{ \boldsymbol{y}^T \boldsymbol{x} - f(\boldsymbol{x}) \right\} \tag{5}$$

with $\mathrm{dom}\,(f) = \{\boldsymbol{x} \in \mathbb{R}^n : f(\boldsymbol{x}) < +\infty\}$ its domain. Similarly, we let f_* denote its concave conjugate. In addition, for a given polytope S, we let $\mathrm{vert}\,(S)$ denote the set of extreme points of S. We recall that a 0-1-polytope is a polytope whose extreme points are binary vectors.

2 Theoretical development

2.1 A non-convex separation problem

Problem (1) can be reformulated as (see e.g., [21])

$$(P) \qquad \inf x_0 \tag{6}$$
$$\text{s.t. } \boldsymbol{x} \in X, x_0 \in \mathbb{R} \tag{7}$$
$$\forall \boldsymbol{\xi} \in \Xi, \exists \boldsymbol{y} \in Y(\boldsymbol{x}, \boldsymbol{\xi}), x_0 \geq g_0(\boldsymbol{x}, \boldsymbol{y}). \tag{8}$$

Since explicitly adding all constraints (8) to the formulation is not viable in practice, we follow a separation approach in which, given pair $(x_0, \boldsymbol{x})$, we check whether a violated constraint exists. Solving the *separation problem* asks to answer the following question:

"Given $(x_0, \boldsymbol{x}) \in \mathbb{R} \times X$, can we show that for all $\boldsymbol{\xi} \in \Xi$ there exists a $\hat{\boldsymbol{y}} \in Y(\boldsymbol{x}, \boldsymbol{\xi})$ such that $x_0 \geq g_0(\boldsymbol{x}, \hat{\boldsymbol{y}})$? If not, can we identify $\hat{\boldsymbol{\xi}} \in \Xi$ such that either $Y(\boldsymbol{x}, \hat{\boldsymbol{\xi}}) = \emptyset$ or $\forall \boldsymbol{y} \in Y(\boldsymbol{x}, \hat{\boldsymbol{\xi}}), x_0 < g_0(\boldsymbol{x}, \boldsymbol{y})$?".

In the following Lemma, we give a sufficient and necessary condition for answering an easier question: given $(x_0, \boldsymbol{x}) \in \mathbb{R} \times X$ and $\boldsymbol{\xi} \in \Xi$, is there a feasible second-stage decision $\hat{\boldsymbol{y}} \in Y(\boldsymbol{x}, \boldsymbol{\xi})$ such that $x_0 \geq g_0(\boldsymbol{x}, \hat{\boldsymbol{y}})$?

Lemma 1. *Given $(x_0, \boldsymbol{x}) \in \mathbb{R} \times X$ and $\boldsymbol{\xi} \in \Xi$, there exists $\hat{\boldsymbol{y}} \in Y(\boldsymbol{x}, \boldsymbol{\xi})$ such that $x_0 \geq g_0(\boldsymbol{x}, \hat{\boldsymbol{y}})$ if and only if the following condition holds*

$$\forall (\lambda_0, \boldsymbol{\lambda}) \in \mathbb{R}_+^{m_Y+1}, \quad \inf_{\boldsymbol{y} \in \mathbb{R}^{n_Y}} \left\{ \boldsymbol{\lambda}^T (\boldsymbol{F}(\boldsymbol{x})\boldsymbol{\xi} + \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y})) + \lambda_0 (g_0(\boldsymbol{x}, \boldsymbol{y}) - x_0) \right\} \leq 0. \tag{9}$$

Proof. First, it is trivial that $\hat{\boldsymbol{y}} \in Y(\boldsymbol{x}, \boldsymbol{\xi})$ implies (9).
Assume now that condition (9) holds, in which case we have

$$\sup_{(\lambda_0, \boldsymbol{\lambda}) \in \mathbb{R}_+^{m_Y+1}} \inf_{\boldsymbol{y} \in \mathbb{R}^{n_Y}} \left\{ \boldsymbol{\lambda}^T (\boldsymbol{F}(\boldsymbol{x})\boldsymbol{\xi} + \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y})) + \lambda_0 (g_0(\boldsymbol{x}, \boldsymbol{y}) - x_0) \right\} \leq 0 \tag{10}$$

Since $(\boldsymbol{0}, 0)$ is a possible choice for $(\lambda_0, \boldsymbol{\lambda})$ in (10), we have that

$$\sup_{(\lambda_0, \boldsymbol{\lambda}) \geq \boldsymbol{0}} \inf_{\boldsymbol{y} \in \mathbb{R}^{n_Y}} \left\{ \boldsymbol{\lambda}^T (\boldsymbol{F}(\boldsymbol{x})\boldsymbol{\xi} + \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y})) + \lambda_0 (g_0(\boldsymbol{x}, \boldsymbol{y}) - x_0) \right\} = 0. \tag{11}$$

As the left-hand side of (11) is the dual of the following problem

$$\inf_{\boldsymbol{y} \in \mathbb{R}^{n_Y}} \{ 0 : x_0 \geq g_0(\boldsymbol{x}, \boldsymbol{y}), \boldsymbol{y} \in Y(\boldsymbol{x}, \boldsymbol{\xi}) \} \tag{12}$$

then, the latter must have optimal value 0, i.e., it is feasible and exhibits the claimed $\hat{\boldsymbol{y}}$. Note that strong duality holds thanks to Assumption 1 (see also, [13], Theorem 5.1).

Remark 1. Condition (9) of Lemma 1 remains valid when adding the restriction $||(\lambda_0, \boldsymbol{\lambda})|| = 1$, where $||\bullet||$ is any norm of $\mathbb{R}^{m+1}$. Indeed, scaling does not impact the sign of the inner inf optimization problem in (10).

Thanks to Lemma 1, we now introduce a non-convex optimization problem which solves the separation problem.

Theorem 1. *Let $(x_0, \boldsymbol{x}) \in \mathbb{R} \times X$, the following propositions are equivalent.*

1. $\forall \boldsymbol{\xi} \in \Xi, \exists \hat{\boldsymbol{y}} \in Y(\boldsymbol{x}, \boldsymbol{\xi}), x_0 \geq g_0(\boldsymbol{x}, \hat{\boldsymbol{y}})$*;*
2. *The following non-convex optimization problem has an optimal objective value which is non-positive*

$$\sup \; -\sum_{i=0}^{m} \lambda_i g_i|_{\boldsymbol{x}}^{*} \left(\frac{\boldsymbol{u}^i}{\lambda_i} \right) + \boldsymbol{\lambda}^T \boldsymbol{F}(\boldsymbol{x}) \boldsymbol{\xi} - \lambda_0 x_0 \tag{13}$$

$$\text{s.t.} \; \sum_{i=0}^{m} \boldsymbol{u}^i = 0 \tag{14}$$

$$(\lambda_0, \boldsymbol{\lambda}) \in \Lambda \tag{15}$$

$$\boldsymbol{\xi} \in \Xi \tag{16}$$

$$\boldsymbol{u}^i \in \mathbb{R}^{n_Y} \qquad i = 0, 1, \ldots, m \tag{17}$$

where $g_i|_{\boldsymbol{x}}(\bullet) = g_i(\boldsymbol{x}, \bullet)$, and $\Lambda = \{(\lambda_0, \boldsymbol{\lambda}) \in \mathbb{R}_+^m \times \mathbb{R}_+ : ||(\lambda_0, \boldsymbol{\lambda})|| = 1\}$.

Proof. Let $(x_0, \boldsymbol{x}) \in \mathbb{R} \times X$. By Lemma 1, for any $\boldsymbol{\xi} \in \Xi$, there exists $\hat{\boldsymbol{y}} \in Y(\boldsymbol{x}, \boldsymbol{\xi})$ such that $x_0 \geq g_0(\boldsymbol{x}, \hat{\boldsymbol{y}})$ if and only if condition (9) is satisfied. Let $(\lambda_0, \boldsymbol{\lambda})$ be any element of Λ. We start by re-arranging the terms of (9) for $(\lambda_0, \boldsymbol{\lambda})$ as follows

$$\inf_{\boldsymbol{y} \in \mathbb{R}^{n_Y}} \left\{ \boldsymbol{\lambda}^T \boldsymbol{g}|_{\boldsymbol{x}}(\boldsymbol{y}) + \lambda_0 g_0|_{\boldsymbol{x}}(\boldsymbol{y}) \right\} + \boldsymbol{\lambda}^T \boldsymbol{F}(\boldsymbol{x}) \boldsymbol{\xi} - \lambda_0 x_0 \leq 0 \tag{18}$$

where terms which do not depend on $\boldsymbol{y}$ are moved out from the optimization problem. Now, letting $\phi(\boldsymbol{y}) = \boldsymbol{\lambda}^T \boldsymbol{g}|_{\boldsymbol{x}}(\boldsymbol{y}) + \lambda_0 g_0|_{\boldsymbol{x}}(\boldsymbol{y})$, by definition the inf problem in (18) is $(-\phi)_*(\boldsymbol{0})$. By exploiting the fact that $(-\phi)_*(\boldsymbol{z}) = -\phi^*(-\boldsymbol{z})$ for any $\boldsymbol{z}$ (see [18], p. 308), we have that (18) is equivalent to

$$-\phi^*(\boldsymbol{0}) + \boldsymbol{\lambda}^T \boldsymbol{F}(\boldsymbol{x}) \boldsymbol{\xi} - \lambda_0 x_0 \leq 0. \tag{19}$$

Using standard conjugate rules (see [18], p. 145), one obtains the following expression of ϕ^*

$$\phi^*(\boldsymbol{z}) = \inf \; \sum_{i=1}^{m} (\lambda_i g_i|_{\boldsymbol{x}})^*(\boldsymbol{u}^i) + (\lambda_0 g_0|_{\boldsymbol{x}})^*(\boldsymbol{u}^0) \tag{20}$$

$$\sum_{i=0}^{m} \boldsymbol{u}^i = \boldsymbol{z} \tag{21}$$

$$\boldsymbol{u}^i \in \mathbb{R}^{n_Y} \qquad i = 0, 1, ..., m. \tag{22}$$

Then, we have $(\lambda_i g_i|_{\boldsymbol{x}})^*(\boldsymbol{u}^i) = \lambda_i g_i|_{\boldsymbol{x}}^*(\boldsymbol{u}^i / \lambda_i)$ (see [18], p. 140). The proof is achieved by requiring that (18) be enforced for all $\boldsymbol{\xi} \in \Xi$.

Example 1 (ℓ_p-norm objective and constraints). Assume that, for each $i = 0, 1, ..., m$, it holds $g_i(\boldsymbol{x},\boldsymbol{y}) = \left|\left|\boldsymbol{K}_X^i\boldsymbol{x} + \boldsymbol{K}_Y^i\boldsymbol{y} + \boldsymbol{\chi}^i\right|\right|_{p_i} + {\boldsymbol{\delta}_X^i}^T\boldsymbol{x} + {\boldsymbol{\delta}_Y^i}^T\boldsymbol{y} + \kappa^i$ where $\boldsymbol{K}_X^i$, $\boldsymbol{K}_Y^i$, $\boldsymbol{\chi}^i$, $\boldsymbol{\delta}_X^i$, $\boldsymbol{\delta}_Y^i$ and κ^i are given. Then, after some convex conjugate algebra, the separation problem from Theorem 1 reads [1]

$$\sup \sum_{i=0}^{m} \boldsymbol{a}^i(\boldsymbol{x})^T\boldsymbol{z}^i + (\boldsymbol{b}(\boldsymbol{x}) + \boldsymbol{F}(\boldsymbol{x})\boldsymbol{\xi})^T\boldsymbol{\lambda} - \lambda_0 x_0 \tag{23}$$

$$\text{s.t.} \sum_{i=0}^{m} \boldsymbol{K}_Y^{i^T}\boldsymbol{z}^i + \boldsymbol{\Delta}\boldsymbol{\lambda} + \boldsymbol{\delta}_Y^0\lambda_0 = \boldsymbol{0} \tag{24}$$

$$\left|\left|\boldsymbol{z}^i\right|\right|_{p_i'} \leq \lambda_i \qquad i = 0, 1, ..., m \tag{25}$$

$$\boldsymbol{z}^i \in \mathbb{R}^{n_Y} \qquad i = 0, 1, ..., m \tag{26}$$

$$(\lambda_0, \boldsymbol{\lambda}) \in \Lambda \tag{27}$$

$$\boldsymbol{\xi} \in \Xi \tag{28}$$

where $\boldsymbol{a}^i(\boldsymbol{x}) = \boldsymbol{K}_X^i\boldsymbol{x} + \boldsymbol{\chi}^i$, $\boldsymbol{b}(\boldsymbol{x}) = ({\boldsymbol{\delta}_X^1}^T\boldsymbol{x} + \kappa^1, \ldots, {\boldsymbol{\delta}_X^m}^T\boldsymbol{x} + \kappa^m)^T$ and $\boldsymbol{\Delta} = (\boldsymbol{\delta}_Y^1, \ldots, \boldsymbol{\delta}_Y^m)$.

Example 2 (Linear case). Assume that $g_i(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{t}_{(i)}\boldsymbol{x} + \boldsymbol{w}_{(i)}\boldsymbol{y}$ for given matrices $\boldsymbol{w}_{(0)}$, $\boldsymbol{W}$, $\boldsymbol{t}_{(0)}$ and $\boldsymbol{T}$. Then, Theorem 1 yields the following separation problem:

$$\max\ (\boldsymbol{T}\boldsymbol{x} + \boldsymbol{F}(\boldsymbol{x})\boldsymbol{\xi})^T\boldsymbol{\lambda} + \boldsymbol{t}_{(0)}\boldsymbol{x}\lambda_0 - \lambda_0 x_0 \tag{29}$$

$$\text{s.t.}\ \boldsymbol{W}^T\boldsymbol{\lambda} + \boldsymbol{w}_{(0)}^T = \boldsymbol{0} \tag{30}$$

$$(\lambda_0, \boldsymbol{\lambda}) \in \Lambda \tag{31}$$

$$\boldsymbol{\xi} \in \Xi. \tag{32}$$

For the specific case where $\boldsymbol{F}$ is affine in $\boldsymbol{x}$, $\boldsymbol{T} = \boldsymbol{0}$ and $\boldsymbol{w}_{(0)} = \boldsymbol{0}$, we enlight that this result is equivalent to Theorem 1 in [1].

2.2 Generalized Benders Decomposition

In this section, we introduce a new Generalized-Benders-Decomposition algorithm able to solve (P) by means of successive separation of infeasible $(x_0, \boldsymbol{x})$ pairs.

For notational convenience, we denote by $\boldsymbol{s}$ a generic tuple $(\boldsymbol{\xi}, \lambda_0, \boldsymbol{\lambda}, \boldsymbol{u}^0, ..., \boldsymbol{u}^m)$, and by S be the set of all such tuples satisfying constraints (14)-(17). In addition, we introduce function σ defined for each $x_0 \in \mathbb{R}$, $\boldsymbol{x} \in X$ and $\boldsymbol{s} \in S$ as the objective function (13), i.e.,

$$\sigma(x_0, \boldsymbol{x}; \boldsymbol{s}) = -\sum_{i=0}^{m} \lambda_i g_i|_{\boldsymbol{x}}^* \left(\frac{\boldsymbol{u}}{\lambda_i}\right) + \boldsymbol{\lambda}^T\boldsymbol{F}(\boldsymbol{x})\boldsymbol{\xi} - \lambda_0 x_0.$$

In the following theorem, we use the result from Theorem 1 to introduce an alternative projected formulation of (P).

[1] Full details of the derivation can be found in the Appendix.

Theorem 2. *Assume X is convex, then Problem (P) is equivalently solved by the following infinite-dimensional convex MINLP*

$$\inf x_0 \tag{33}$$
$$\text{s.t. } \boldsymbol{x} \in X, x_0 \in \mathbb{R} \tag{34}$$
$$\sigma(x_0, \boldsymbol{x}; \boldsymbol{s}) \leq 0 \qquad \forall \boldsymbol{s} \in S. \tag{35}$$

Proof. The reformulation holds by Theorem 1. To show that it is convex, we have to show that, for each $\boldsymbol{s} \in S$, function $\sigma(\bullet, \bullet; \boldsymbol{s})$ is convex with respect to x_0 and $\boldsymbol{x}$. Note that since $\boldsymbol{\lambda}, \boldsymbol{\xi} \geq \boldsymbol{0}$ and function f_{ij} is convex for each $i = 1, \dots, m$ and $j = 1, \dots, n_\Xi$, we have that $\boldsymbol{x} \mapsto \boldsymbol{\lambda}^T \boldsymbol{F}(\boldsymbol{x})\boldsymbol{\xi} - \lambda_0 x_0$ is convex since the non-negative sum of convex functions is convex and the last term is linear. We therefore focus on the remaining part, and show that $\boldsymbol{x} \mapsto g_i|_{\boldsymbol{x}}^*(\boldsymbol{\pi})$ is a concave function for any fixed $\boldsymbol{\pi} \in \mathbb{R}^{n_Y}$. Thus, let $\boldsymbol{\pi} \in \mathbb{R}^{n_Y}$ be fixed. By definition, we have

$$g_i|_{\boldsymbol{x}}^*(\boldsymbol{\pi}) = \sup_{\boldsymbol{y} \in \mathrm{dom}(g_i|_{\boldsymbol{x}})} \{ \boldsymbol{\pi}^T \boldsymbol{y} - g_i|_{\boldsymbol{x}}(\boldsymbol{y})\} = \sup_{\boldsymbol{y} \in \mathrm{dom}(g_i|_{\boldsymbol{x}})} \{ \boldsymbol{\pi}^T \boldsymbol{y} - g_i(\boldsymbol{x}, \boldsymbol{y})\}. \tag{36}$$

Let us introduce new variables $\boldsymbol{z} \in \mathbb{R}^{n_X}$ such that $\boldsymbol{z} = \boldsymbol{x}$. Then, the following holds by Lagrangian duality:

$$g_i|_{\boldsymbol{x}}^*(\boldsymbol{\pi}) = \sup_{(\boldsymbol{z},\boldsymbol{y}) \in \mathrm{dom}(g_i), \boldsymbol{z}=\boldsymbol{x}} \{ \boldsymbol{\pi}^T \boldsymbol{y} - g_i(\boldsymbol{z}, \boldsymbol{y})\} \tag{37}$$
$$= \inf_{\boldsymbol{\lambda} \in \mathbb{R}^{n_X}} \sup_{(\boldsymbol{z},\boldsymbol{y}) \in \mathrm{dom}(g_i)} \{ \boldsymbol{\lambda}^T(\boldsymbol{z} - \boldsymbol{x}) + \boldsymbol{\pi}^T \boldsymbol{y} - g_i(\boldsymbol{z}, \boldsymbol{y})\} \tag{38}$$
$$= \inf_{\boldsymbol{\lambda} \in \mathbb{R}^{n_X}} \sup_{(\boldsymbol{z},\boldsymbol{y}) \in \mathrm{dom}(g_i)} \left\{ -\boldsymbol{\lambda}^T \boldsymbol{x} + \begin{pmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\pi} \end{pmatrix}^T \begin{pmatrix} \boldsymbol{z} \\ \boldsymbol{y} \end{pmatrix} - g_i(\boldsymbol{z}, \boldsymbol{y})\right\} \tag{39}$$
$$= \inf_{\boldsymbol{\lambda} \in \mathbb{R}^{n_X}} \{-\boldsymbol{\lambda}^T \boldsymbol{x} + g_i^*(\boldsymbol{\lambda}, \boldsymbol{\pi})\}. \tag{40}$$

Thus, $g_i|_{\boldsymbol{x}}^*(\boldsymbol{\pi})$ can be expressed as the infimum of infitely many affine functions of $\boldsymbol{x}$, thus, it is concave in $\boldsymbol{x}$.

Based on Theorem 2, we can derive a cutting-plane algorithm where cuts (35) are dynamically generated, for which we discuss finite convergence. The complete procedure is reported in Algorithm 1, where problem (MP) is defined as

$$\inf\{x_0 : (x_0, \boldsymbol{x}) \in \mathbb{R} \times X \quad \sigma(x_0, \boldsymbol{x}; \boldsymbol{s}) \leq 0 \; \forall \boldsymbol{s} \in \hat{S}\} \tag{MP}$$

for some $\hat{S} \subseteq S$ which is dynamically augmented.

Before stating our convergence result, we introduce two definitions.

Definition 1 (ε-oracle). *Let $\varepsilon \geq 0$. We say that an oracle is an ε-oracle for the separation problem if, for any pair $(x_0, \boldsymbol{x}) \in \mathbb{R} \times X$, it returns $\tilde{\boldsymbol{s}} \in S$ such that* $\sup_{\boldsymbol{s} \in S} \sigma(x_0, \boldsymbol{x}; \boldsymbol{s}) - \sigma(x_0, \boldsymbol{x}, \tilde{\boldsymbol{s}}) \leq \varepsilon$.

Definition 2 (ε-relaxed feasible space). *We define the ε-relaxed feasible space as*

$$Y^\varepsilon(\boldsymbol{x}, \boldsymbol{\xi}) = \{\boldsymbol{y} \in \mathbb{R}^{n_Y} : \boldsymbol{F}(\boldsymbol{x})\boldsymbol{\xi} + \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}) \leq \varepsilon \boldsymbol{e}\}\,. \tag{41}$$

Algorithm 1 Generalized Benders decomposition

input an instance of (P), a tolerance $\varepsilon \geq 0$ and an initial set $\hat{S} \subseteq S$ such that (MP) is bounded.
$Stop \leftarrow false$
while $Stop = false$ **do**
  Solve (MP)
  if (MP) is infeasible **then**
    Declare (P) infeasible and **return**
  end if
  Let $(x_0^*, \boldsymbol{x}^*)$ be an optimal solution of (MP)
  Solve the separation problem (13)-(17) for $(x_0^*, \boldsymbol{x}^*)$
  Let $\boldsymbol{s}^* = (\boldsymbol{\xi}^*, \lambda_0^*, \boldsymbol{\lambda}^*, \boldsymbol{u}^{0*}, ..., \boldsymbol{u}^{m*})$ be an optimal solution of the separation problem
  if $\sigma(x_0^*, \boldsymbol{x}^*; \boldsymbol{s}^*) \geq 0$ **then**
    $\hat{S} \leftarrow \hat{S} \cup \{\boldsymbol{s}^*\}$
  else
    $Stop \leftarrow true$
  end if
end while

We are now ready to state our main convergence result.

Theorem 3. *Assume that the separation problem is solved with an ε-oracle. Then, Algorithm 1 finitely terminates. Moreover, it either returns a solution $(x_0, \boldsymbol{x})$ such that $\forall \boldsymbol{\xi} \in \Xi, \exists \boldsymbol{y} \in Y^\varepsilon(\boldsymbol{x}, \boldsymbol{\xi}), x_0 \geq g_0(\boldsymbol{x}, \boldsymbol{y}) - \varepsilon$, or correctly concludes that (P) is infeasible.*

Proof. First, observe that the separation problem from Theorem 1 is an uncoupled biconvex problem. Thus, at most $|\text{vert}(\Xi)|$ iterations are needed for the algorithm to stop. Assume that, at some iteration, the oracle returns a solution $\boldsymbol{s}^*$ such that $\sigma(x_0^*, \boldsymbol{x}^*; \boldsymbol{s}^*) \leq 0$. Together with $\sup_{\boldsymbol{s} \in S} \sigma(x_0^*, \boldsymbol{x}^*; \boldsymbol{s}) - \sigma(x_0^*, \boldsymbol{x}^*, \boldsymbol{s}^*) \leq \varepsilon$, we have that $\sup_{\boldsymbol{s} \in S} \sigma(x_0^*, \boldsymbol{x}^*; \boldsymbol{s}) \leq \varepsilon$. In turn, this implies

$$\sup_{\boldsymbol{s} \in S} \sigma(x_0^*, \boldsymbol{x}^*; \boldsymbol{s}) \tag{42}$$

$$= \sup_{\boldsymbol{\xi} \in \Xi, (\lambda_0, \boldsymbol{\lambda}) \in \Lambda} \inf_{\boldsymbol{y} \in \mathbb{R}^{n_Y}} \left\{ \boldsymbol{\lambda}^T (\boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{\xi} + \boldsymbol{g}(\boldsymbol{x}^*, \boldsymbol{y})) + \lambda_0 (g_0(\boldsymbol{x}^*, \boldsymbol{y}) - x_0^*) \right\} \tag{43}$$

$$= \sup_{\boldsymbol{\xi} \in \Xi} \inf_{\boldsymbol{y} \in \mathbb{R}^{n_Y}} \sup_{(\lambda_0, \boldsymbol{\lambda}) \in \Lambda} \left\{ \boldsymbol{\lambda}^T (\boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{\xi} + \boldsymbol{g}(\boldsymbol{x}^*, \boldsymbol{y})) + \lambda_0 (g_0(\boldsymbol{x}^*, \boldsymbol{y}) - x_0^*) \right\} \tag{44}$$

$$= \sup_{\boldsymbol{\xi} \in \Xi} \inf_{\boldsymbol{y} \in \mathbb{R}^{n_Y}} \max \left\{ \max_{i=1,\dots,m} \left\{ \boldsymbol{f}_{(i)}(\boldsymbol{x}^*)\boldsymbol{\xi} + g_i(\boldsymbol{x}^*, \boldsymbol{y}) \right\}; g_0(\boldsymbol{x}^*, \boldsymbol{y}) - x_0^*; 0 \right\} \tag{45}$$

$$\leq \varepsilon. \tag{46}$$

Here, the minimax theorem from [17] was used to swap the sup and inf operators. This shows that $(x_0, \boldsymbol{x})$ is such that $\forall \boldsymbol{\xi} \in \Xi, \exists \boldsymbol{y} \in Y^\varepsilon(\boldsymbol{x}, \boldsymbol{\xi}), x_0 \geq g_0(\boldsymbol{x}, \boldsymbol{y}) - \varepsilon$.

Otherwise, assume that, at some iteration, (MP) is infeasible. This means that "$\forall (x_0, \boldsymbol{x}) \in \mathbb{R} \times X, \exists \hat{\boldsymbol{s}} \in \hat{S}, \sigma(x_0, \boldsymbol{x}; \hat{\boldsymbol{s}}) > 0$" holds, which implies

$$\forall (x_0, \boldsymbol{x}) \in \mathbb{R} \times X, \quad \sup_{\boldsymbol{s} \in S} \sigma(x_0, \boldsymbol{x}; \boldsymbol{s}) > 0. \tag{47}$$

By Theorem 1, this shows that (P) is infeasible.

2.3 0-1 polytopic uncertainty sets

Algorithm 1 gives a general scheme for solving problem (P), though its applicability depends on the possibility of solving the separation problem in practice. In this section, we present a convex MINLP formulation of the latter in the relevant case in which Ξ has an affine mapping of a 0-1 polytope. Note that, given any polytopic uncertain set, there exists an affine mapping to a 0-1 polytope. To see this, one can always express Ξ as the set of convex combinations of its extreme points. In the following, we will denote by $\Omega \subseteq \mathbb{R}^{n_\Omega}$ the 0-1 polytope associated with Ξ, and assume its dimension n_Ω be manageable. Clearly, when $\Xi \subseteq [0,1]^{n_\Xi}$, $\boldsymbol{U}$ is a totally unimodular matrix and $\boldsymbol{d}$ is integral (see (2)), the identity mapping can be used and $n_\Omega = n_\Xi$. This is notably the case for the budgeted uncertainty set (see [9]) with an integer parameter. For the case where the budget parameter is fractional, [1] shows that an affine mapping with a 0-1 polytope of size $2n_\Xi$ exists. We now state our theorem.

Theorem 4. *Let $\Omega \subseteq \mathbb{R}^{n_\Omega}$ be a given 0-1 polytope and let $\boldsymbol{\rho}^0, \boldsymbol{\rho}^1, ..., \boldsymbol{\rho}^{n_\Omega} \in \mathbb{R}^{n_\Xi}$ be some vectors. Assume that $\Xi = \tilde{\boldsymbol{\rho}}(\Omega)$ where $\tilde{\boldsymbol{\rho}} : \boldsymbol{\omega} \mapsto \boldsymbol{\rho}^0 + \sum_{k=1}^{n_\Omega} \boldsymbol{\rho}^k \omega_k$. Then, given a pair $(x_0, \boldsymbol{x}) \in \mathbb{R} \times X$, the separation model introduced in Theorem 1 can be reformulated as the following convex MINLP.*

$$\sup \ -\sum_{i=0}^{m} \lambda_i g_i|_{\boldsymbol{x}}^* \left(\frac{\boldsymbol{u}}{\lambda_i} \right) + \boldsymbol{\lambda}^T \boldsymbol{F}(\boldsymbol{x}) \boldsymbol{\rho}^0 + \sum_{k=1}^{n_\Omega} {\boldsymbol{\theta}^k}^T \boldsymbol{F}(\boldsymbol{x}) \boldsymbol{\rho}^k - \lambda_0 x_0 \tag{48}$$

$$\text{s.t.} \ \sum_{i=0}^{m} \boldsymbol{u}^i = \boldsymbol{0} \tag{49}$$

$$(\lambda_0, \boldsymbol{\lambda}) \in \Lambda \tag{50}$$

$$\boldsymbol{\theta}^k \leq \boldsymbol{\lambda} \qquad k = 1, ..., n_\Omega \tag{51}$$

$$\boldsymbol{\theta}^k \leq \omega_k \boldsymbol{e} \qquad k = 1, ..., n_\Omega \tag{52}$$

$$\boldsymbol{\theta}^k \geq \boldsymbol{\lambda} + \omega_k \boldsymbol{e} - \boldsymbol{e} \qquad k = 1, ..., n_\Omega \tag{53}$$

$$\boldsymbol{\theta}^k \in \mathbb{R}_+^m \qquad k = 1, ..., n_\Omega \tag{54}$$

$$\boldsymbol{\omega} \in \Omega \cap \{0,1\}^{n_\Omega} \tag{55}$$

$$\boldsymbol{u}^i \in \mathbb{R}^{n_Y} \qquad i = 0, 1, ..., m \tag{56}$$

Proof. By replacing each $\boldsymbol{\xi} \in \Xi$ by an $\omega \in \Omega$ such that $\boldsymbol{\xi} = \boldsymbol{\rho}^0 + \sum_{k=1}^{n_\Omega} \boldsymbol{\rho}^k \omega_k$, objective function (13) can be rewritten as

$$\sup \ -\sum_{i=0}^{m} \lambda_i g_i|_{\boldsymbol{x}}^* \left(\frac{\boldsymbol{u}^i}{\lambda_i} \right) + \boldsymbol{\lambda}^T \boldsymbol{F}(\boldsymbol{x}) \boldsymbol{\rho}^0 + \sum_{k=1}^{n_\Omega} \boldsymbol{\lambda}^T \boldsymbol{F}(\boldsymbol{x}) \boldsymbol{\rho}^k w_k - \lambda_0 x_0$$

Noting that this function includes bilinear terms, we can restrict our attention to $\omega \in \text{vert}(\Omega) \subseteq \{0,1\}^{n_\Omega}$. By introducing variables $\theta_i^k = \lambda_i \omega_k$ $(i = 1, ..., m$ and

$k = 1, ..., n_\Omega$), the bilinear term can be linearized as follows

$$\sum_{k=1}^{n_\Omega} \boldsymbol{\lambda}^T \boldsymbol{F}(\boldsymbol{x})\boldsymbol{\rho}^k w_k = \sum_{k=1}^{n_\Omega}\sum_{i=1}^{m}\sum_{j=1}^{n_\Xi} f_{ij}(\boldsymbol{x})\rho_j^k \underbrace{\lambda_i \omega_k}_{=\theta_i^k} = \sum_{k=1}^{n_\Omega} {\boldsymbol{\theta}^k}^T \boldsymbol{F}(\boldsymbol{x})\boldsymbol{\rho}^k. \tag{57}$$

The result follows as (51)–(53) are linearization constraints and $0 \leq \lambda_i \leq 1$ by assumption.

3 Application: resource planning problem

We tested our method on a stochastic resource planning problem from the literature, modified so as to incorporate resource congestion.

Our algorithm was implemented in C++17 using Gurobi 9.5 to solve the underlying optimization problems. All experiments were conducted on an AMD Ryzen 5 PRO 4650GE at 3.3 GHz, with a time limit equal to 3,600 CPU seconds per run.

3.1 Problem description

We are given a set I of resources (e.g., server types) and a set J of customers. Each resource $i \in I$ is associated to a unitary cost c_i, and each customer $j \in J$ has a demand d_j. We denote by μ_{ij} the service rate of resource i for customer j, i.e., the fraction of demand of j can be served by a unit of i. The deterministic problem introduced in [16] asks to serve all customers' demands while minimizing the total cost of the used resources, and is formulated as follows

$$\min \sum_{i \in I} c_i x_i \tag{58}$$

$$\text{s.t.} \sum_{j \in J} y_{ij} \leq x_i \qquad i \in I \tag{59}$$

$$\sum_{i \in I} \mu_{ij} y_{ij} \geq d_j \qquad j \in J \tag{60}$$

$$x_i \geq 0 \qquad i \in I \tag{61}$$

$$y_{ij} \geq 0 \qquad (i,j) \in I \times J, \tag{62}$$

Here, each variable x_i ($i \in I$) represents the amount of resource i to buy, while y_{ij} ($(i,j) \in I \times J$) is the amount of resource i allocated to customer j.

We modify the model to consider congestion, which may reduce the possibility for the customers to access the resources, thus requiring an increased amount of resource to guarantee a given service rate. Accordingly, constraints (59) are replaced by

$$\Big(1 + \alpha_i \big[\sum_{j \in J} y_{ij}\big]^{\beta_i}\Big)\Big(\sum_{j \in J} y_{ij}\Big) \leq x_i \qquad i \in I, \tag{63}$$

where α_i and β_i are given parameters related to resource $i \in I$. This congestion function, which is convex for any non-negative α_i and β_i, was introduced by [11] for a facility location problem.

In a two-stage setting, customers' demands are typically not known when the resources have to be bought, whereas the assignment of customers to resources can be postponed after the actual realization of the demands reveals. We assume that the demand of each customer j has a nominal value $\bar{d}_j$ and maximum deviation $\tilde{d}_j$. The resulting demand is $d_j = \bar{d}_j + \xi_j \tilde{d}_j$, where ξ_j is a random parameter and $\boldsymbol{\xi}$ belongs to a budgeted uncertainty set $\Xi = \{\boldsymbol{\xi} \in [0,1]^{|J|} : \sum_{j \in J} \xi_j \leq \Gamma\}$.

3.2 Instance generation

We generated random instances as follows: for each resource i and customer j, the service rate μ_{ij} is defined as $\mathcal{U}(0,1)$, where $\mathcal{U}(a,b)$ denotes the uniform distribution between a and b. For each customer j, the nominal demand $\bar{d}_j$ is taken from $\mathcal{U}(1,50)$, and the maximum deviation $\tilde{d}_j$ is either $0.05\bar{d}_j$ or $0.10\bar{d}_j$. For each resource i, the cost c_i is in $\mathcal{U}(8,10) \sum_{j \in J} \mu_{ij}/|J|$, so that the higher the average efficiency the more costly the resource. In addition, congestion of the resource is defined by α_i in $\mathcal{U}(0,1)$ and $\beta_i = 1$.

We generated instances of different sizes $(|I|,|J|)$ equal to $(10,20)$, $(10,30)$, $(15,30)$, $(15,40)$, $(20,40)$ and $(20,50)$. For each size, 5 instances were generated. Finally, the uncertainty budget Γ was set to $\lfloor p|J| \rfloor$ where $p \in \{0.05, 0.10, 0.20\}$, i.e., up to p% of the clients change their demands to the maximum value.

3.3 Results

Table 1 gives the outcome of our computational experiments and reports, for each group of 5 homogeneous instances, the number of cases in which the algorithm computes a provably optimal solution, and the corresponding average computing time and average number of times the separation problem was solved.

The results show that our approach is able to solve to optimality all instances with $p = 0.05$, with an average computing time raising from few seconds to almost 15 minutes. Increasing the value of p yields more challenging instances. This is shown both by the number of optimal solution (51 out of 60) and by the average computing time for solved instances, which can be as large as twice the time needed in the previous case. This trend is more evident for $p = 0.20$, in which case the algorithm solves to optimality only 36 instances (out of 60) with a significantly larger computing time. Concerning the number of separation rounds, it grows with the size of the instance, whereas the value of p is immaterial. Finally, instances with $\bar{d}_j/\tilde{d}_j = 0.05$ appear to be more challenging for our algorithm.

			$p = 0.05$			$p = 0.10$			$p = 0.20$		
$\lvert I\rvert$	$\lvert J\rvert$	$\bar{d}_j/\tilde{d}_j$	opt	time	$\lvert\hat{S}\rvert$	opt	time	$\lvert\hat{S}\rvert$	opt	time	$\lvert\hat{S}\rvert$
10	20	0.05	5	13.92	187.00	5	15.82	201.00	5	19.63	201.40
		0.10	5	15.13	194.60	5	16.39	197.40	5	21.05	202.40
	30	0.05	5	23.41	213.80	5	35.76	216.80	5	386.51	222.00
		0.10	5	23.81	207.80	5	45.37	223.20	5	290.78	223.40
15	30	0.05	5	64.01	406.60	5	118.86	412.00	4	517.51	451.00
		0.10	5	70.41	429.20	5	175.26	426.00	5	357.54	420.60
	40	0.05	5	129.24	381.60	3	1133.50	346.00	1	215.05	311.00
		0.10	5	143.41	387.40	5	743.03	382.60	1	2507.04	313.00
20	40	0.05	5	383.06	664.40	3	1900.71	675.33	0	—	—
		0.10	5	455.72	669.00	5	847.54	695.40	3	952.38	669.67
	50	0.05	5	785.16	692.60	1	801.34	532.00	0	—	—
		0.10	5	853.19	683.60	4	1507.89	646.25	2	3117.22	692.00

Table 1. Computational results on the resource allocation problem with different uncertainty budgets.

4 Conclusion

In this paper, we studied general adjustable robust optimization problems where the second-stage feasible space is defined by means of convex constraints. Using Fenchel duality, we were able to derive a nonconvex separation problem for a suited reformulation of the problem. We also showed that, in case the uncertainty set has a special structure, separation reduces to solve a convex MINLP, thus allowing the design of an effective Benders-like exact algorithm. We computationally tested an implementation of this method on an uncertain variant of a planning problem from the literature, showing the practical applicability of the proposed approach.

References

1. Ayoub, J., Poss, M.: Decomposition for adjustable robust linear optimization subject to uncertainty polytope. Computational Management Science **13**(2), 219–239 (2016)
2. Ben-Tal, A., Goryashko, A., Guslitzer, E., Nemirovski, A.: Adjustable robust solutions of uncertain linear programs. Mathematical Programming **99**(2), 351–376 (2004)
3. Ben-Tal, A., den Hertog, D., Vial, J.P.: Deriving robust counterparts of nonlinear uncertain inequalities. Mathematical Programming **149**(1-2), 265–299 (2014)
4. Bertsimas, D., Bidkhori, H.: On the performance of affine policies for two-stage adaptive optimization: a geometric perspective. Mathematical Programming **153**(2), 577–594 (2014)
5. Bertsimas, D., Caramanis, C.: Finite adaptability in multistage linear optimization. IEEE Transactions on Automatic Control **55**(12), 2751–2766 (2010)

6. Bertsimas, D., Goyal, V.: On the power and limitations of affine policies in two-stage adaptive optimization. Mathematical Programming **134**(2), 491–531 (2011)
7. Bertsimas, D., Litvinov, E., Sun, X.A., Zhao, J., Zheng, T.: Adaptive robust optimization for the security constrained unit commitment problem. IEEE Transactions on Power Systems **28**(1), 52–63 (2013)
8. Bertsimas, D., de Ruiter, F.J.C.T.: Duality in two-stage adaptive linear optimization: Faster computation and stronger bounds. INFORMS Journal on Computing **28**(3), 500–511 (2016)
9. Bertsimas, D., Sim, M.: The price of robustness. Operations Research **52**(1), 35–53 (2004)
10. Boni, O., Ben-Tal, A.: Adjustable robust counterpart of conic quadratic problems. Mathematical Methods of Operations Research **68**(2), 211–233 (2008)
11. Desrochers, M., Marcotte, P., Stan, M.: The congested facility location problem. Location Science **3**(1), 9–23 (1995)
12. Gabrel, V., Lacroix, M., Murat, C., Remli, N.: Robust location transportation problems under uncertain demands. Discrete Applied Mathematics **164**, 100–111 (2014)
13. Geoffrion, A.M.: Generalized benders decomposition. Journal of Optimization Theory and Applications **10**(4), 237–260 (1972)
14. Hanasusanto, G.A., Kuhn, D., Wiesemann, W.: K-adaptability in two-stage robust binary programming. Operations Research **63**(4), 877–891 (2015)
15. Jiang, R., Zhang, M., Li, G., Guan, Y.: Two-stage network constrained robust unit commitment problem. European Journal of Operational Research **234**(3), 751–762 (2014)
16. Luedtke, J.: A branch-and-cut decomposition algorithm for solving chance-constrained mathematical programs with finite support. Mathematical Programming **146**(1), 219–244 (2014)
17. Perchet, V., Vigeral, G.: A minmax theorem for concave-convex mappings with no regularity assumptions. Journal of Convex Analysis **22** (2015)
18. Rockafellar, R.T.: Convex analysis. Princeton Landmarks in Mathematics and Physics, Princeton University Press, Princeton, NJ (1996)
19. de Ruiter, F.J.C.T., Zhen, J., den Hertog, D.: Dual approach for two-stage robust nonlinear optimization. Operations Research (2022)
20. Subramanyam, A., Gounaris, C.E., Wiesemann, W.: K-adaptability in two-stage mixed-integer robust optimization. Mathematical Programming Computation **12**(2), 193–224 (2019)
21. Takeda, A., Taguchi, S., Tütüncü, R.H.: Adjustable robust optimization models for a nonlinear two-period system. Journal of Optimization Theory and Applications **136**(2), 275–295 (2007)
22. Terry, T., Epelman, M., Thiele, A.: Robust linear optimization with recourse. technical report, Lehigh University (2009)
23. Zeng, B., Zhao, L.: Solving two-stage robust optimization problems using a column-and-constraint generation method. Operations Research Letters **41**(5), 457–461 (2013)
24. Zhen, J., den Hertog, D., Sim, M.: Adjustable robust optimization via fourier–motzkin elimination. Operations Research **66**(4), 1086–1100 (2018)

A Proof of Example 1

Assume that g_i $(i = 0, 1, ..., m)$ is generically defined by means of ℓ_p-norms, i.e., assume that $g_i(\boldsymbol{x}, \boldsymbol{y}) = \left|\left|\boldsymbol{K}_X^i \boldsymbol{x} + \boldsymbol{K}_Y^i \boldsymbol{y} + \boldsymbol{\chi}^i\right|\right|_{p_i} + {\boldsymbol{\delta}_X^i}^T \boldsymbol{x} + {\boldsymbol{\delta}_Y^i}^T \boldsymbol{y} + \kappa^i$ where $\boldsymbol{K}_X^i, \boldsymbol{K}_Y^i, \boldsymbol{\chi}^i, \boldsymbol{\delta}_X^i, \boldsymbol{\delta}_Y^i$ and κ^i are given.

A.1 Computing $g_i|_{\boldsymbol{x}}^*$

First, observe that

$$g_i|_{\boldsymbol{x}}(\boldsymbol{y}) = h_1(\boldsymbol{y}) + {\boldsymbol{\delta}_X^i}^T \boldsymbol{x} + {\boldsymbol{\delta}_Y^i}^T \boldsymbol{y} + \kappa^i \tag{64}$$

where $h_1(\boldsymbol{y}) = \left|\left|\boldsymbol{K}_X^i \boldsymbol{x} + \boldsymbol{K}_Y^i \boldsymbol{y} + \boldsymbol{\chi}^i\right|\right|_{p_i}$. By addition to an affine function, we have

$$g_i|_{\boldsymbol{x}}^*(\boldsymbol{\pi}) = h_1^*(\boldsymbol{\pi} - \boldsymbol{\delta}_Y^i) - {\boldsymbol{\delta}_X^i}^T \boldsymbol{x} - \kappa^i. \tag{65}$$

Now, we may write h_1 as

$$h_1(\boldsymbol{y}) = h_2(\boldsymbol{K}_Y^i \boldsymbol{y}) \tag{66}$$

where $h_2(\boldsymbol{y}) = \left|\left|\boldsymbol{K}_X^i \boldsymbol{x} + \boldsymbol{y} + \boldsymbol{\chi}^i\right|\right|_{p_i}$. By composition with a linear mapping (see [3], Lemma 6.7) and since $\mathrm{dom}\,(h_2) = \mathbb{R}^{n_Y}$, we have

$$h_1^*(\boldsymbol{\pi}) = \inf_{\boldsymbol{\omega}} \{h_2^*(\boldsymbol{\omega}) : {\boldsymbol{K}_Y^i}^T \boldsymbol{\omega} = \boldsymbol{\pi}\}. \tag{67}$$

Together with (65), we have

$$g_i|_{\boldsymbol{x}}^*(\boldsymbol{\pi}) = \inf_{\boldsymbol{\omega}} \{h_2^*(\boldsymbol{\omega}) : {\boldsymbol{K}_Y^i}^T \boldsymbol{\omega} = \boldsymbol{\pi} - \boldsymbol{\delta}_Y^i\} - {\boldsymbol{\delta}_X^i}^T \boldsymbol{x} - \kappa^i. \tag{68}$$

Then,

$$h_2(\boldsymbol{y}) = h_3(\boldsymbol{y} + \boldsymbol{K}_X^i \boldsymbol{x} + \boldsymbol{\chi}^i) \tag{69}$$

with $h_3(\boldsymbol{y}) = ||\boldsymbol{y}||_{p_i}$. Thus, by translation of argument, we have

$$h_2^*(\boldsymbol{\pi}) = h_3^*(\boldsymbol{\pi}) - (\boldsymbol{K}_X^i \boldsymbol{x} + \boldsymbol{\chi}^i)^T \boldsymbol{\pi}. \tag{70}$$

Now, h_3 being a norm, its convex conjugate is the indicator of the unit ball for the dual norm, thus,

$$h_3^*(\boldsymbol{\pi}) = \delta(\boldsymbol{\pi} | B_{p_i'}(\boldsymbol{0}, 1)) \tag{71}$$

with $1/p_i + 1/p_i' = 1$. Together with (68) and (70), we have

$$g_i|_{\boldsymbol{x}}^*(\boldsymbol{\pi}) = \inf_{\boldsymbol{\omega}} \{\delta(\boldsymbol{\omega} | B_{p_i'}(\boldsymbol{0}, 1)) - (\boldsymbol{K}_X^i \boldsymbol{x} + \boldsymbol{\chi}^i)^T \boldsymbol{\omega} : {\boldsymbol{K}_Y^i}^T \boldsymbol{\omega} = \boldsymbol{\pi} - \boldsymbol{\delta}_Y^i\} \\ - {\boldsymbol{\delta}_X^i}^T \boldsymbol{x} - \kappa^i. \tag{72}$$

By optimality, we get

$$g_i|_{\boldsymbol{x}}^*(\boldsymbol{\pi}) = \inf - (\boldsymbol{K}_X^i \boldsymbol{x} + \boldsymbol{\chi}^i)^T \boldsymbol{\omega} - {\boldsymbol{\delta}_X^i}^T \boldsymbol{x} - \kappa^i \tag{73}$$

$$\text{s.t. } {\boldsymbol{K}_Y^i}^T \boldsymbol{\omega} = \boldsymbol{\pi} - \boldsymbol{\delta}_Y^i \tag{74}$$

$$||\boldsymbol{\omega}||_{p_i'} \leq 1 \tag{75}$$

$$\boldsymbol{\omega} \in \mathbb{R}^{n_Y} \tag{76}$$

A.2 Applying Theorem 1

By substitution, we get

$$\lambda_i g_i|_{\boldsymbol{x}}^*(\boldsymbol{u}^i/\lambda_i) = \inf \lambda_i \left(-(\boldsymbol{K}_X^i \boldsymbol{x} + \boldsymbol{\chi}^i)^T \boldsymbol{\omega}^i - {\boldsymbol{\delta}_X^i}^T \boldsymbol{x} - \kappa^i\right) \tag{77}$$

$$\text{s.t. } {\boldsymbol{K}_Y^i}^T \boldsymbol{\omega}^i = \boldsymbol{u}^i/\lambda_i - \boldsymbol{\delta}_Y^i \tag{78}$$

$$||\boldsymbol{\omega}||_{p_i'} \leq 1 \tag{79}$$

$$\boldsymbol{\omega}^i \in \mathbb{R}^{n_Y}. \tag{80}$$

Introducing $\boldsymbol{z}^i = \lambda_i \boldsymbol{\omega}^i$, we have

$$\lambda_i g_i|_{\boldsymbol{x}}^*(\boldsymbol{u}^i/\lambda_i) = \inf -(\boldsymbol{K}_X^i \boldsymbol{x} + \boldsymbol{\chi}^i)^T \boldsymbol{z}^i - \lambda_i({\boldsymbol{\delta}_X^i}^T \boldsymbol{x} + \kappa^i) \tag{81}$$

$$\text{s.t. } {\boldsymbol{K}_Y^i}^T \boldsymbol{z}^i = \boldsymbol{u}^i - \lambda_i \boldsymbol{\delta}_Y^i \tag{82}$$

$$\left|\left|\boldsymbol{z}^i\right|\right|_{p_i'} \leq \lambda_i \tag{83}$$

$$\boldsymbol{z}^i \in \mathbb{R}^{n_Y}. \tag{84}$$

We therefore obtain

$$\sup \sum_{i=0}^{m} \left((\boldsymbol{K}_X^i \boldsymbol{x} + \boldsymbol{\chi}^i)^T \boldsymbol{z}^i + \lambda_i({\boldsymbol{\delta}_X^i}^T \boldsymbol{x} + \kappa^i)\right) + \boldsymbol{\lambda}^T \boldsymbol{F}(\boldsymbol{x})\boldsymbol{\xi} - \lambda_0 x_0 \tag{85}$$

$$\text{s.t. } \sum_{i=0}^{m} \left({\boldsymbol{K}_Y^i}^T \boldsymbol{z}^i + \lambda_i \boldsymbol{\delta}_Y^i\right) = \boldsymbol{0} \tag{86}$$

$$\left|\left|\boldsymbol{z}^i\right|\right|_{p_i'} \leq \lambda_i \qquad i = 0, 1, ..., m \tag{87}$$

$$\boldsymbol{z}^i \in \mathbb{R}^{n_Y} \qquad i = 0, 1, ..., m \tag{88}$$

$$(\lambda_0, \boldsymbol{\lambda}) \in \Lambda \tag{89}$$

$$\boldsymbol{\xi} \in \Xi. \tag{90}$$

By letting $\boldsymbol{a}^i(\boldsymbol{x}) = \boldsymbol{K}_X^i \boldsymbol{x} + \boldsymbol{\chi}^i$, $\boldsymbol{b}(\boldsymbol{x}) = ({\boldsymbol{\delta}_X^1}^T \boldsymbol{x} + \kappa^1, \ldots, {\boldsymbol{\delta}_X^m}^T \boldsymbol{x} + \kappa^m)^T$ and $\boldsymbol{\Delta} = (\boldsymbol{\delta}_Y^1, \ldots, \boldsymbol{\delta}_Y^m)$, we may rewrite it as follows.

$$\sup \sum_{i=0}^{m} \boldsymbol{a}^i(\boldsymbol{x})^T \boldsymbol{z}^i + (\boldsymbol{b}(\boldsymbol{x}) + \boldsymbol{F}(\boldsymbol{x})\boldsymbol{\xi})^T \boldsymbol{\lambda} - \lambda_0 x_0 \tag{91}$$

$$\text{s.t. } \sum_{i=0}^{m} \boldsymbol{K}_Y^{i T} \boldsymbol{z}^i + \boldsymbol{\Delta}\boldsymbol{\lambda} + \boldsymbol{\delta}_Y^0 \lambda_0 = \boldsymbol{0} \tag{92}$$

$$\left|\left|\boldsymbol{z}^i\right|\right|_{p_i'} \leq \lambda_i \qquad i = 0, 1, ..., m \tag{93}$$

$$\boldsymbol{z}^i \in \mathbb{R}^{n_Y} \qquad i = 0, 1, ..., m \tag{94}$$

$$(\lambda_0, \boldsymbol{\lambda}) \in \Lambda \tag{95}$$

$$\boldsymbol{\xi} \in \Xi \tag{96}$$