Inverse linear optimization for the recovery of constraint parameters in robust and non-robust problems

by

Neal Kaw

A thesis submitted in conformity with the requirements
for the degree of Master of Applied Science
Graduate Department of Mechanical and Industrial Engineering
University of Toronto

Abstract

Inverse linear optimization for the recovery of constraint parameters in robust and non-robust problems

Neal Kaw
Master of Applied Science
Graduate Department of Mechanical and Industrial Engineering
University of Toronto
2017

Most inverse optimization models impute unspecified parameters of an objective function to make an observed solution optimal for a given optimization problem. In this thesis, we propose two approaches to impute unspecified left-hand-side constraint coefficients in addition to a cost vector for a given linear optimization problem. The first approach minimally perturbs prior estimates of the unspecified parameters to satisfy strong duality, while the second identifies parameters minimizing the duality gap, if it is not possible to satisfy the optimality conditions exactly. We apply these two approaches to the general linear optimization problem. We also use them to impute unspecified parameters of the uncertainty set for robust linear optimization problems under interval and cardinality constrained uncertainty. Each inverse optimization model we propose is nonconvex, but we show that a globally optimal solution can be obtained by solving a finite number of tractable optimization problems.

Acknowledgements

I would like to thank my supervisor, Timothy Chan, for being an outstanding teacher and role model throughout the years that I have known him. I especially thank him for his patience and encouragement, for giving me numerous opportunities, and for helping me reach my goals.

I thank my committee members, Merve Bodur and Roy Kwon, for the helpful comments they provided on my work. I also want to thank Justin Boutilier, Michael Carter, Timothy Chan, Leslie Sinclair, Deborah Tihanyi, and Nicholas Yeung for giving me the opportunity to be a TA during my graduate studies.

I want to thank Ali Vahit Esensoy for his mentorship and continued support since our work together at Cancer Care Ontario. Finally, I thank my friends and colleagues in the Applied Optimization Lab for making sure that the day-to-day work of graduate school was always fun: Aaron Babier, Justin Boutilier, Iman Dayarian, Derya Demirtas, Minha Lee, Rafid Mahmood, Philip Mar, Ben Potter, Chris Sun, Islay Wright, and Ian Zhu. It was a pleasure to work with you all.

Contents

Chapter 1

Introduction

Inverse optimization (IO) aims to determine the unspecified parameters of an optimization problem (the *forward problem*) that make a given solution optimal. To date, most of the literature has focused on formulating and solving IO problems that determine parameters of the objective function, under the assumption that parameters specifying the feasible region are fixed. These methods are appropriate regardless of whether the given solution is or is not a candidate to be exactly optimal (e.g., a boundary vs. interior point in a linear optimization problem). In the former case, the optimality conditions can be satisfied exactly (Ahuja and Orlin, 2001, Iyengar and Kang, 2005, Schaefer, 2009), and to choose among multiple satisfactory imputations, it is typical to include an objective function that minimally perturbs "prior" estimates such that the observed solution is exactly optimal. In the latter case, the imputed parameters minimize some measure of suboptimality (Chan et al., 2014, 2017, Keshavarz et al., 2011).

While most IO literature has focused on imputing an objective function, some recent work has considered imputing constraint coefficients in a linear optimization problem. The problem of imputing a cost vector alone can be solved by a linear IO model, but the models for imputing constraint coefficients are all nonconvex. In particular, only a small number of these papers consider imputing "left-hand-side" constraint coefficients: Birge et al. (2017) use observed electricity prices to impute parameters of electricity market structure in the economic dispatch problem, and Brucker and Shakhlevich (2009) use an observed job schedule to impute job processing times in the minimax lateness scheduling problem. In both cases, the authors exploit characteristics of their particular forward problem to derive a tractable inverse problem.

The goal of this thesis is to solve the IO problem of imputing unspecified constraint parameters for the general linear programming problem, such that an observed solution is optimal with respect to some nonzero cost vector. The motivation for this problem is that we may have estimates of constraint parameters, but due to inaccuracy in these estimates the observed solution is not a candidate to be optimal for the forward problem; or we may altogether lack estimates of some constraint coefficients. Accordingly, the first part of this thesis considers the problem of recovering a left-hand-side constraint matrix for a general linear programming problem, given an observed solution and a known "right-hand-side" constraint vector.

In the second part of this thesis, we adapt this IO approach to the situation that the forward problem is a robust optimization problem: this can be viewed as a special case of the general problem from the first part of the thesis, but we will first motivate this problem independently. Suppose that we have estimates of all constraint parameters, but the observed solution is apparently not a candidate to be optimal for the forward problem: the reason for this may be that we have failed to take into account a model of uncertainty the decision maker incorporated into her decision-making process. Given the growing adoption of robust optimization in both the research and practitioner community, it will increasingly be the case that robustly optimized decisions may be observed in a variety of settings and IO models will be taking such solutions as input. Accordingly, we consider the problem of recovering unspecified parameters of an uncertainty set, given an observed solution and nominal estimates of all constraint coefficients for a linear optimization problem. In particular, we will solve this problem for two different uncertainty sets, interval uncertainty (Ben-Tal and Nemirovski, 2000) and cardinality constrained uncertainty (Bertsimas and Sim, 2004). In both cases, the robust problem is itself a linear optimization problem, and hence imputing unspecified uncertainty set parameters is a special case of the general problem considered in the first part of this thesis.

For each of the three forward problems we have introduced, we consider two IO problem variants distinguished by whether or not it is possible to guarantee a priori that there exist parameters making the observed solution exactly optimal. In inverse linear optimization models that impute a cost vector only, the candidacy of an observed solution to be optimal depends on whether or not the solution is on the boundary of the feasible set. However, because in this work we are imputing parameters that determine the feasible set, the candidacy of the observed solution to be optimal depends on the extent to which the unspecified parameters are allowed to change the feasible set. Limitations on this ability to change the feasible set may arise from external constraints on the unspecified parameters, motivated by the application domain. Accordingly, we will propose two alternative IO models for each forward model: the first will require the observed solution to be exactly optimal, and the second will minimize suboptimality, in case optimality cannot be achieved exactly.

In this work, we study three different linear optimization problems as forward problems, each with a different set of constraint parameters to be imputed. For each forward problem, we make two contributions:

1. We derive a tractable solution method for the nonconvex IO problem of minimally perturbing prior estimates of constraint parameters, such that an observed solution is exactly optimal for some nonzero cost vector. When the forward problem is the general linear programming problem or a robust linear program with interval uncertainty, the method requires solving a finite number of convex optimization problems (linear when the extent of perturbation is measured by a linear objective function.) When the forward problem is a robust linear program with cardinality constrained uncertainty, the method requires

solving a finite number of linear optimization problems.

2. We derive a tractable solution method for the nonconvex IO problem of imputing duality gap-minimizing constraint parameters, subject to external (application-motivated) constraints on the parameters to be imputed. When the forward problem is the general linear programming problem, the method requires solving a finite number of optimization problems which are linear whenever the external constraints are linear. When the forward problem is a robust linear optimization problem, the method requires solving a single mixed integer optimization problem, which is mixed integer linear whenever the external constraints are linear.

1.1 Related literature

In this section, we provide an overview of related literature in two areas: inverse optimization and robust optimization.

1.1.1 Inverse optimization

Ahuja and Orlin (2001) formalized the classical approach to inverse linear optimization, which assumes that a single observed solution is on the boundary of a known feasible region, and minimally perturbs a prior cost vector such that the observed solution is optimal. Chan et al. (2014) generalized this approach by allowing that the observed solution may not be a candidate to be an optimal solution, and instead impute a cost vector that minimizes the duality gap. Chan et al. (2017) further generalized this approach by proposing a model that minimizes a general error function, and showing that the duality gap-minimizing model can be derived as a special case.

Recently, some methods have been developed to impute both the right-hand-side and cost vectors for a linear optimization problem. This problem is in general nonconvex, and the authors who have addressed this problem have either derived an approximate solution, or exploited characteristics of their particular problem instance to derive a globally optimal solution. In the former group of papers, Dempe and Lohse (2006) propose an IO model that minimally perturbs an observed solution such that there exist right-hand-side and cost vectors making it exactly optimal, and they derive a local optimality condition for this model. Saez-Gallego et al. (2016) and Saez-Gallego and Morales (2017) also formulate models to recover these two vectors, and apply the methods to inverse problems in electricity markets: the first paper minimizes the sum of perturbations of the observed solutions such that they are exactly optimal, whereas the second paper minimizes the sum of duality gaps for the observed solutions. Both models are solved using approximate solution methods that first estimate the right-hand-side vector, and then fix its value to linearize the IO model.

Several papers exploit problem-specific characteristics to impute globally optimal right-hand-side and cost vectors. Černỳ and Hladík (2016) assume that the observed solution is optimal with respect to the prior estimates of these vectors, and determine the maximum distance that each vector can be perturbed in the direction of a given perturbation vector. Chow and Recker (2012) reduce the complexity of the IO problem by exploiting the structure of their forward problem, the household activity pattern problem. We also note two papers that impute right-hand-side coefficients under the assumption that the cost vector is known: Güler and Hamacher (2010) consider the problem of minimally perturbing edge capacities in the minimum cost flow problem such that an observed solution is optimal, and they solve this problem using an optimality condition specific to the minimum cost flow problem. Xu et al. (2016) assume that multiple observed solutions in the feasible set $\mathbf{A}\mathbf{x} \leq \mathbf{b}$ are all exactly optimal, and therefore the optimality conditions will be satisfied by a right-hand-side vector in which each component has the minimum possible value such that all observed solutions are feasible.

Few papers have addressed the problem of imputing the unspecified parameters of the coefficient matrix, and those that do exploit specific problem characteristics or make assumptions to derive a tractable problem. Birge et al. (2017) assume partial access to both the primal and dual solutions, to eliminate bilinearities in the IO problem. Brucker and Shakhlevich (2009) make use of the necessary and sufficient optimality conditions for their forward problem, the minimax lateness scheduling problem.

Chassein and Goerigk (2016) propose IO models that recover parameters of the uncertainty set for a robust optimization problem, but their work is distinct from ours in several respects. Their forward problem is a discrete optimization problem in which only the cost vector is subject to uncertainty, and they consider interval uncertainty but not cardinality constrained uncertainty. Their inverse problem assumes that the feasible set is known and the observed solution is optimal with respect to the nominal cost vector, and seeks to determine the greatest degree of uncertainty in the cost vector such that the nominal optimal solution remains optimal for the robust problem. Hence their IO models do not recover constraint parameters for the forward problem.

Although we have focused on IO models for which the forward problem is linear, we note that IO models have also been proposed for forward problems that are nonlinear, including conic (Iyengar and Kang, 2005), convex (Keshavarz et al., 2011), and discrete (Heuberger, 2004, Schaefer, 2009). IO models have furthermore been proposed for problems where parameters of the forward problem must be imputed from multiple observed solutions (Aswani et al., 2015, Bertsimas et al., 2015, Keshavarz et al., 2011).

1.1.2 Robust optimization

Robust optimization solves the problem of optimization under parameter uncertainty by requiring a feasible solution to satisfy the constraints for all possible realizations of uncertain parameters. In this thesis we are concerned with an uncertain linear program $\min\{\mathbf{c}^\intercal\mathbf{x} \colon \mathbf{A}\mathbf{x} \geq \mathbf{b}\}$, where the matrix $\mathbf{A}$ contains nominal estimates of uncertain quantities. The corresponding *robust counterpart* $\min\{\mathbf{c}^\intercal\mathbf{x}\colon \tilde{\mathbf{A}}\mathbf{x} \geq \mathbf{b},\ \forall\tilde{\mathbf{A}} \in \mathcal{U}\}$ minimizes the same objective function over the *robust feasible region*, which requires a solution to satisfy the constraints for all possible realizations of $\tilde{\mathbf{A}}$ in the *uncertainty set* $\mathcal{U}$, which includes the nominal estimate $\mathbf{A}$. The robust feasible region is thus a subset of the nominal feasible region. For any given uncertainty set, some transformations may be required to derive a tractable optimization problem that is equivalent to the robust counterpart.

The first robust optimization model was proposed by Soyster (1973), who defined an uncertainty set where the columns of the constraint matrix are elements of independent convex sets. He then showed that the robust counterpart is equivalent to a linear program in which each constraint coefficient is obtained by taking the maximum value over the convex set defining the allowable realizations of that coefficient's respective column. This approach was then applied by Ben-Tal and Nemirovski (2000) to a special case of Soyster's uncertainty set referred to as interval uncertainty, in which each and every constraint coefficient is allowed to vary independently within a symmetric interval centered on the nominal estimate (see also Ben-Tal et al. (2009, p. 19)).

The interval uncertainty approach can be considered over-conservative because the robust counterpart effectively assumes that all uncertain coefficients may take their worst-case values simultaneously, leading to the exclusion of good solutions that may be feasible with high probability. Accordingly, several authors proposed ellipsoidal uncertainty sets that allow the user to better tune the degree of conservatism. The simplest ellipsoidal uncertainty set (Ben-Tal et al., 2009, p. 19) allows that for a given constraint, each coefficient may vary within some fraction of its interval of uncertainty, and the vector of fractions must be within a Euclidean ball of some fixed radius, where the radius controls the degree of conservatism. A more sophisticated variant intersects simple ellipsoidal with interval uncertainty (Ben-Tal and Nemirovski, 2000).

A disadvantage of the ellipsoidal uncertainty set is that the resulting robust counterpart is a second order conic program and hence more computationally demanding than the original uncertain linear program. To preserve the advantages of the ellipsoidal uncertainty set while also preserving the complexity of the original problem, Bertsimas and Sim (2004) proposed the cardinality constrained uncertainty set, which is similar to the ellipsoidal uncertainty set, but requires the vector of fractions to be within a ball of the L_1 norm rather than the Euclidean norm. It can furthermore be shown that the cardinality constrained uncertainty set is a special instance of the polyhedral uncertainty set, in which each row of the constraint matrix is allowed to vary within a polyhedron, and for which the robust counterpart is also a linear program (Bertsimas et al., 2011).

Several other approaches have been proposed in the literature, including norm uncertainty (Bertsimas et al., 2004), distributional robustness (Delage and Ye, 2010), and data-driven uncertainty (Bertsimas et al., 2017). Moreover, while we have focused on uncertain linear programs, robust counterparts have also been proposed for nonlinear optimization problems, including quadratic (Ben-Tal et al., 2002) and discrete (Kouvelis and Yu, 1997). Applications of robust optimization have included finance (Goldfarb and Iyengar, 2003), revenue management (Ball and Queyranne, 2009), supply chain management (Bertsimas and Thiele, 2006), medicine (Unkelbach et al., 2007), and energy systems (Jiang et al., 2012). For fuller surveys on robust optimization, we refer the reader to Bertsimas et al. (2011), Gabrel et al. (2014), and the reference book by Ben-Tal et al. (2009).

1.2 Organization and notation

The remainder of this thesis is organized as follows. In Chapter 2, we propose IO models and solution methods to impute left-hand-side coefficients for the general linear programming problem. First we consider a model that requires the imputed parameters to satisfy strong duality, second we consider a model that minimizes a duality gap, and then we provide numerical examples to illustrate the two approaches. In Chapter 3, we propose IO approaches to impute uncertainty set parameters for the robust linear optimization problem under interval uncertainty, and cardinality constrained uncertainty. For each of the two cases we again propose strong duality and duality gap-minimizing variants, and numerical examples. In Chapter 4, we provide concluding remarks. Proofs which do not appear in the body of the thesis are contained in the appendix.

We define the following notation. Let $\mathbf{e}$ be the vector of all ones. Let $\mathbf{e}_i$ be the unit vector with i-th coordinate equal to 1. Let $\mathbf{a}_i$ be the i-th row of $\mathbf{A}$. If we have a set of vectors with common index but differing lengths (e.g., $\boldsymbol{\alpha}_i$ for all $i \in I$), we will sometimes abuse notation and use $\boldsymbol{\alpha}$ to denote the set of vectors $\{\boldsymbol{\alpha}_i\}_{i\in I}$. In some optimization models, we will abuse notation and write $\mathbf{A}$ as a variable, although we are minimizing over the vectors $\{\mathbf{a}_i\}_{i\in I}$ rather than a matrix $\mathbf{A}$. We define $\text{sgn}(x) = 1$ if $x \geq 0$ and -1 otherwise.

Chapter 2

Inverse linear optimization

In this chapter, we consider the general linear optimization problem

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \sum_{j \in J} c_j x_j \\ \text{subject to} \quad & \sum_{j \in J} a_{ij} x_j \geq b_i, \quad \forall i \in I. \end{aligned} \tag{2.1}$$

Given $\mathbf{b}$ and an observed solution $\hat{\mathbf{x}}$, the IO problem aims to identify a constraint matrix $\mathbf{A}$ that makes the observed solution $\hat{\mathbf{x}}$ optimal for some nonzero cost vector. In particular, we consider two variants of this problem. The goal of the first problem is to minimally perturb prior constraint vectors $\hat{\mathbf{a}}_i$ such that $\hat{\mathbf{x}}$ is exactly optimal for some nonzero cost vector. The goal of the second problem is to identify constraint vectors $\mathbf{a}_i$ in some predefined set $\boldsymbol{\Omega}$ such that there exists a nonzero cost vector minimizing the duality gap. The set $\boldsymbol{\Omega}$ corresponds to external constraints on the constraint coefficients, motivated by the application domain, and if these constraints are sufficiently restrictive then a duality gap of zero (i.e., exact optimality) will not be possible.

We note here that the problem of recovering both $\mathbf{A}$ and $\mathbf{b}$ can be subsumed by the problem of recovering $\mathbf{A}$ only. To show this, we first note that problem (2.1) is equivalent to

$$\begin{aligned} \underset{\mathbf{x},q}{\text{minimize}} \quad & \mathbf{c}^\intercal \mathbf{x} + hq \\ \text{subject to} \quad & \begin{pmatrix} \mathbf{A} & \mathbf{b} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ q \end{pmatrix} \geq \mathbf{0}, \\ & q = -1. \end{aligned} \tag{2.2}$$

The auxiliary variable q can be assigned an arbitrary coefficient h in the objective function because the value of q is fixed to -1, and thus the particular value of its coefficient in the objective function is inconsequential. Let our forward problem be $\min_{\mathbf{x}}\{\bar{\mathbf{c}}^\intercal \mathbf{x} \colon \bar{\mathbf{A}}\mathbf{x} \geq \bar{\mathbf{b}}\}$, with $\bar{\mathbf{A}} = \begin{pmatrix} \mathbf{A} & \mathbf{b} \end{pmatrix}$, $\bar{\mathbf{b}} = \mathbf{0}$, $\bar{\mathbf{c}} = (\mathbf{c}, h)$, and $\bar{\mathbf{x}} = (\hat{\mathbf{x}}, -1)$. The IO problem recovers $\bar{\mathbf{A}}$, and hence it effectively recovers $\mathbf{A}$ and $\mathbf{b}$.

The remainder of this chapter is organized as follows. Sections 2.1 and 2.2 introduce the strong duality and duality gap problems respectively, and describe tractable solution approaches. Subsection 2.1.1 will demonstrate that the strong duality formulation may produce trivial solutions in some situations. Section 2.3 provides numerical examples to illustrate the solutions to the IO problems, including examples with trivial solutions.

2.1 Strong duality

We assume $\hat{\mathbf{x}} \neq \mathbf{0}$ (otherwise, unless $b_i = 0$ for some $i \in I$, it would not be possible for any constraint to be active, and hence it would not be possible to achieve strong duality.) We assume that prior constraint vectors $\hat{\mathbf{a}}_i \neq \mathbf{0}$ are given for all $i \in I$. Let $\boldsymbol{\pi}$ be the dual vector associated with the constraints of the forward problem (2.1). The following formulation minimizes the weighted deviations of the vectors $\mathbf{a}_i$ from $\hat{\mathbf{a}}_i$, while enforcing strong duality, and primal and dual feasibility:

$$\begin{aligned}
\underset{\mathbf{A},\mathbf{c},\boldsymbol{\pi}}{\text{minimize}} \quad & \sum_{i\in I} \xi_i \|\mathbf{a}_i - \hat{\mathbf{a}}_i\| && \text{(2.3a)}\\
\text{subject to} \quad & \sum_{j\in J} c_j \hat{x}_j - \sum_{i\in I} b_i \pi_i = 0, && \text{(2.3b)}\\
& \sum_{j\in J} a_{ij}\hat{x}_j \geq b_i, \quad \forall i \in I, && \text{(2.3c)}\\
& \sum_{i\in I} \pi_i = 1, && \text{(2.3d)}\\
& \sum_{i\in I} a_{ij}\pi_i = c_j, \quad \forall j \in J, && \text{(2.3e)}\\
& \pi_i \geq 0, \quad \forall i \in I. && \text{(2.3f)}
\end{aligned}$$

In the objective function (2.3a), $\|\cdot\|$ is an arbitrary norm, and $\boldsymbol{\xi}$ is a vector of real-valued weights that is user-tunable. Unlike previous IO approaches that minimize deviation of $\mathbf{c}$ from some prior $\hat{\mathbf{c}}$, we do not include such an objective since our goal is to determine a constraint matrix $\mathbf{A}$ that makes $\hat{\mathbf{x}}$ optimal; the variable $\mathbf{c}$ is simply needed to ensure $\hat{\mathbf{x}}$ is optimal with respect to *some* cost vector.

Constraints (2.3b), (2.3c), and (2.3e)-(2.3f) represent strong duality, primal feasibility, and dual feasibility, respectively. Notice that the strong duality and dual feasibility constraints can trivially be satisfied by $(\mathbf{c}, \boldsymbol{\pi}) = (\mathbf{0}, \mathbf{0})$. Formulation (2.3) would then only require $\mathbf{A}$ to ensure primal feasibility, which is insufficient to guarantee $\hat{\mathbf{x}}$ is optimal with respect to some nonzero $\mathbf{c}$. Accordingly, constraint (2.3d) is a normalization constraint that prevents $\boldsymbol{\pi} = \mathbf{0}$ from being feasible, and requires $\mathbf{c}$ to be in the convex hull of $\{\mathbf{a}_i\}_{i\in I}$. This set of feasible cost vectors may still include $\mathbf{c} = \mathbf{0}$, but whether $\mathbf{c} = \mathbf{0}$ will be optimal depends on the choice of norm, and on the problem data. We will comment further on this possibility in Section 2.1.1.

Before proceeding with the solution of formulation (2.3), we briefly note that it has a feasible solution.

Proposition 1. *Formulation (2.3) is feasible.*

All constraints of formulation (2.3) are linear except for the dual feasibility constraint (2.3e), which is bilinear in $\mathbf{A}$ and the dual vector $\boldsymbol{\pi}$. Nevertheless we can determine an efficient solution method. To do so, we first show that the constraints of formulation (2.3) effectively formalize the geometric intuition that an optimal solution for a linear program must be on the boundary of the feasible region.

Lemma 1. *Every feasible solution for formulation (2.3) satisfies*

$$\sum_{j \in J} a_{\hat{i}j} \hat{x}_j = b_{\hat{i}}, \quad \text{for some } \hat{i} \in I, \tag{2.4a}$$

$$\sum_{j \in J} a_{ij} \hat{x}_j \geq b_i, \quad \forall i \in I. \tag{2.4b}$$

Conversely, for every $\mathbf{A}$ satisfying (2.4), there exists $(\mathbf{c}, \boldsymbol{\pi})$ such that $(\mathbf{A}, \mathbf{c}, \boldsymbol{\pi})$ is feasible for formulation (2.3).

Proof: To prove the first statement, we assume $\sum_{j \in J} a_{ij} \hat{x}_j > b_i$ for all $i \in I$ and derive a contradiction. Substituting (2.3e) into (2.3b), we get

$$\sum_{i \in I} \pi_i \sum_{j \in J} a_{ij} \hat{x}_j = \sum_{i \in I} \pi_i b_i. \tag{2.5}$$

Constraint (2.3d) ensures that $\bar{I} := \{i \in I \colon \pi_i > 0\} \neq \varnothing$. Since we have assumed $\sum_{j \in J} a_{ij} \hat{x}_j > b_i$ for all $i \in I$, we have

$$\pi_i \sum_{j \in J} a_{ij} \hat{x}_j > \pi_i b_i, \quad \forall i \in \bar{I},$$

and since $\pi_i \geq 0$ for all $i \in I$,

$$\sum_{i \in I} \pi_i \sum_{j \in J} a_{ij} \hat{x}_j > \sum_{i \in I} \pi_i b_i,$$

which contradicts equation (2.5).

To prove the second statement, let $\mathbf{A}$ satisfy (2.4), $\hat{i}$ be defined by (2.4a), $\mathbf{c} = \mathbf{a}_{\hat{i}}$ and $\boldsymbol{\pi} = \mathbf{e}_{\hat{i}}$. This solution is feasible for formulation (2.3). □

Lemma 1 allows us to characterize an optimal solution for formulation (2.3) and devise an efficient solution method:

Theorem 1. *For all $i \in I$, let*

$$f_i = \min_{\mathbf{a}_i} \left\{ \xi_i \|\mathbf{a}_i - \hat{\mathbf{a}}_i\| : \sum_{j \in J} a_{ij} \hat{x}_j = b_i \right\}, \tag{2.6}$$

$$g_i = \min_{\mathbf{a}_i} \left\{ \xi_i \|\mathbf{a}_i - \hat{\mathbf{a}}_i\| : \sum_{j \in J} a_{ij} \hat{x}_j \geq b_i \right\}, \tag{2.7}$$

and let $\mathbf{a}_i^f$ and $\mathbf{a}_i^g$ be optimal solutions for (2.6) and (2.7), respectively. Let $i^* \in \arg\min_{i \in I}\{f_i - g_i\}$. Then the optimal value of formulation (2.3) is $f_{i^*} + \sum_{i \neq i^*, i \in I} g_i$, and there exists an optimal solution of (2.3) with

$$\mathbf{a}_i = \begin{cases} \mathbf{a}_i^f & \text{if } i = i^*, \\ \mathbf{a}_i^g & \text{if } i \neq i^*, i \in I, \end{cases} \tag{2.8}$$

$$\mathbf{c} = \mathbf{a}_{i^*}. \tag{2.9}$$

Remark 1. Theorem 1 shows that an optimal solution to the nonconvex inverse problem (2.3) can be found by solving $2|I|$ convex problems (linear with appropriate choice of $\|\cdot\|$).

Proof: By Lemma 1, solving formulation (2.3) is equivalent to solving the following optimization problem for all $\hat{i} \in I$, and taking the minimum over all $|I|$ optimal values:

$$\begin{aligned} \underset{\mathbf{A}}{\text{minimize}} \quad & \sum_{i \in I} \xi_i \|\mathbf{a}_i - \hat{\mathbf{a}}_i\| \\ \text{subject to} \quad & \sum_{j \in J} a_{\hat{i}j} \hat{x}_j = b_{\hat{i}}, \\ & \sum_{j \in J} a_{ij} \hat{x}_j \geq b_i, \quad \forall i \in I. \end{aligned} \tag{2.10}$$

Suppose we fix some $\hat{i} \in I$. Since formulation (2.10) is separable by i, the optimal value of the $\hat{i}$-th formulation (2.10) is $f_{\hat{i}} + \sum_{i \neq \hat{i}, i \in I} g_i$. Therefore, the optimal value of formulation (2.3) is

$$\min_{\hat{i} \in I} \left\{ f_{\hat{i}} + \sum_{i \neq \hat{i}, i \in I} g_i \right\}.$$

Clearly, the optimal index i^* must satisfy $i^* \in \arg\min_{i \in I}\{f_i - g_i\}$. An optimal $\mathbf{A}$ is derived from the optimal solutions of (2.6) and (2.7),

$$\mathbf{a}_i = \begin{cases} \mathbf{a}_i^f & \text{if } i = i^*, \\ \mathbf{a}_i^g & \text{if } i \neq i^*, i \in I, \end{cases}$$

and an optimal cost vector is $\mathbf{c} = \mathbf{a}_{i^*}$. □

Theorem 1 can be interpreted as follows. For all $i \in I$, f_i is the minimal value of the i-th term in objective function (2.3a) such that constraint i is rendered active. Similarly, g_i is the minimal value for constraint i to be rendered feasible; clearly, $g_i \neq 0$ only if $\hat{\mathbf{x}}$ is infeasible with respect to $\hat{\mathbf{a}}_i$. For $\hat{\mathbf{x}}$ to be optimal for the forward problem, some constraint i^* must have $\mathbf{a}_{i^*}$ set such that $\hat{\mathbf{x}}$ is on the boundary. The optimal choice of this constraint is the one that requires the minimal additional increase in $\xi_i \|\mathbf{a}_i - \hat{\mathbf{a}}_i\|$ for the constraint to be active rather than merely feasible, i.e., $i^* \in \arg\min\{f_i - g_i\}$. To satisfy the optimality conditions, the cost vector is set perpendicular to this active constraint.

Because problems (2.6) and (2.7) are the projection of a point onto a hyperplane and halfspace respectively, they have analytical solutions when the projection uses the Euclidean norm (Boyd and Vandenberghe, 2004, p. 398). We omit the proof of this result, which is straightforward.

Corollary 1. *Let the norm in objective function (2.3a) be the Euclidean norm. Then for all $i \in I$,*

$$f_i = \xi_i \left\| \frac{\hat{\mathbf{a}}_i^\intercal \hat{\mathbf{x}} - b_i}{\|\hat{\mathbf{x}}\|_2^2} \hat{\mathbf{x}} \right\|_2, \tag{2.11}$$

$$\mathbf{a}_i^f = \hat{\mathbf{a}}_i - \frac{\hat{\mathbf{a}}_i^\intercal \hat{\mathbf{x}} - b_i}{\|\hat{\mathbf{x}}\|_2^2} \hat{\mathbf{x}}, \tag{2.12}$$

$$g_i = \begin{cases} \xi_i \left\| \frac{\hat{\mathbf{a}}_i^\intercal \hat{\mathbf{x}} - b_i}{\|\hat{\mathbf{x}}\|_2^2} \hat{\mathbf{x}} \right\|_2 & \text{if } \hat{\mathbf{a}}_i^\intercal \hat{\mathbf{x}} < b_i, \\ 0 & \text{otherwise,} \end{cases} \tag{2.13}$$

$$\mathbf{a}_i^g = \begin{cases} \hat{\mathbf{a}}_i - \frac{\hat{\mathbf{a}}_i^\intercal \hat{\mathbf{x}} - b_i}{\|\hat{\mathbf{x}}\|_2^2} \hat{\mathbf{x}} & \text{if } \hat{\mathbf{a}}_i^\intercal \hat{\mathbf{x}} < b_i, \\ \hat{\mathbf{a}}_i & \text{otherwise.} \end{cases} \tag{2.14}$$

2.1.1 Possibility of trivial solutions

In general, it is possible that the optimal solutions for formulation (2.3) described in Theorem 1 are trivial, by which we mean that $\mathbf{c} = \mathbf{0}$ and/or $\mathbf{a}_i = \mathbf{0}$ for any $i \in I$. The conditions on the problem data $(\mathbf{A}, \mathbf{b}, \hat{\mathbf{x}})$ that cause this issue depend on the choice of norm in the objective function (2.3a). Only for the Euclidean norm will we be able to characterize the problem data under which this issue occurs. For the remainder of this section, we assume the norm in (2.3a) is the Euclidean norm.

To determine when (2.8)-(2.9) give a trivial solution, we must first determine when $\mathbf{a}_i^f = \mathbf{0}$ or $\mathbf{a}_i^g = \mathbf{0}$. The following two results establish necessary and sufficient conditions on $\hat{\mathbf{a}}_i, b_i$, and $\hat{\mathbf{x}}$ for these trivial solutions to occur:

Proposition 2. *Let the norm in objective function (2.3a) be the Euclidean norm. For all $i \in I$, $\mathbf{a}_i^f = \mathbf{0}$ if and only if $b_i = 0$ and $\hat{\mathbf{x}} = \delta_i \hat{\mathbf{a}}_i$ for some $\delta_i \in \mathbb{R}$.*

Proof: ($\Rightarrow$) Since $\mathbf{a}_i = \mathbf{a}_i^f = \mathbf{0}$ must satisfy $\mathbf{a}_i^\intercal \hat{\mathbf{x}} = b_i$, we get $b_i = 0$. Using equation (2.12),

$\mathbf{a}_i^f = \mathbf{0}$ and $b_i = 0$ imply $\hat{\mathbf{a}}_i = \frac{\hat{\mathbf{a}}_i^\intercal \hat{\mathbf{x}}}{\|\hat{\mathbf{x}}\|_2^2}\hat{\mathbf{x}}$. Since $\hat{\mathbf{a}}_i \neq \mathbf{0}$ by assumption, $\hat{\mathbf{a}}_i^\intercal \hat{\mathbf{x}} \neq 0$, and we can deduce $\hat{\mathbf{x}} = \delta_i \hat{\mathbf{a}}_i$ with $\delta_i = \frac{\|\hat{\mathbf{x}}\|_2^2}{\hat{\mathbf{a}}_i^\intercal \hat{\mathbf{x}}}$.

($\Leftarrow$) If $b_i = 0$ and $\hat{\mathbf{x}} = \delta_i \hat{\mathbf{a}}_i$, then it is easy to see from equation (2.12) that $\mathbf{a}_i^f = \mathbf{0}$. □

Proposition 3. *Let the norm in objective function (2.3a) be the Euclidean norm. For all $i \in I$, $\mathbf{a}_i^g = \mathbf{0}$ if and only if $\hat{\mathbf{a}}_i^\intercal \hat{\mathbf{x}} < b_i$, $b_i = 0$ and $\hat{\mathbf{x}} = \delta_i \hat{\mathbf{a}}_i$ for some $\delta_i \in \mathbb{R}$.*

Proof: ($\Rightarrow$) Since $\hat{\mathbf{a}}_i \neq \mathbf{0}$ by assumption, $\mathbf{a}_i^g = \mathbf{0}$ and the cases in equation (2.14) imply that $\hat{\mathbf{a}}_i^\intercal \hat{\mathbf{x}} < b_i$. In the case that $\hat{\mathbf{a}}_i^\intercal \hat{\mathbf{x}} < b_i$, the Euclidean projection of $\hat{\mathbf{a}}_i$ onto the halfspace $\hat{\mathbf{x}}^\intercal \mathbf{a}_i \geq b_i$ is equivalent to projection onto the boundary of the halfspace, $\hat{\mathbf{x}}^\intercal \mathbf{a}_i = b_i$. The latter equation must be satisfied by $\mathbf{a}_i = \mathbf{a}_i^g = \mathbf{0}$, which implies that $b_i = 0$. As in the proof of Proposition (2), equation (2.14), $\mathbf{a}_i^g = \mathbf{0}$, and $b_i = 0$ imply $\hat{\mathbf{x}} = \delta_i \hat{\mathbf{a}}_i$ with $\delta_i = \frac{\|\hat{\mathbf{x}}\|_2^2}{\hat{\mathbf{a}}_i^\intercal \hat{\mathbf{x}}}$.

($\Leftarrow$) If $\hat{\mathbf{a}}_i^\intercal \hat{\mathbf{x}} < b_i$, $b_i = 0$ and $\hat{\mathbf{x}} = \delta_i \hat{\mathbf{a}}_i$, then it is easy to see from equation (2.14) that $\mathbf{a}_i^g = \mathbf{0}$. □

The geometric interpretation of Proposition 2 is that $\mathbf{a}_i^f = \mathbf{0}$ occurs if and only if the prior vector $\hat{\mathbf{a}}_i$ is parallel to $\hat{\mathbf{x}}$, and the boundary of the constraint intersects the origin. The interpretation of Proposition 3 is similar, except that $\mathbf{a}_i^g = \mathbf{0}$ additionally requires $\hat{\mathbf{x}}$ to be infeasible with respect to $\hat{\mathbf{a}}_i$.

Propositions 2 and 3 characterize the problem data for which problems (2.6) and (2.7) have trivial solutions, but we are interested in when formulation (2.3) has a trivial solution. There are two slightly different ways this can occur. First, if $\mathbf{a}_i^f = \mathbf{0}$ for all $i \in I^* := \arg\min_{i \in I}\{f_i - g_i\}$, then $\mathbf{a}_{i^*} = \mathbf{0}$ and $\mathbf{c} = \mathbf{0}$ (it is necessary that $\mathbf{a}_i^f = \mathbf{0}$ *for all* $i \in I^*$ because if there exists any $i \in I^*$ such that $\mathbf{a}_i^f \neq \mathbf{0}$, then that index can be set as i^*.) Second, for any $i \neq i^*$, if $\mathbf{a}_i^g = \mathbf{0}$ then $\mathbf{a}_i = \mathbf{0}$. In both scenarios, there are an infinite number of nonzero vectors $\mathbf{a}_i$ which could take the place of $\mathbf{a}_{i^*}$ or $\mathbf{a}_i^g$, but it is unclear that there is any reasonable general approach to choosing among them. In the first scenario, we may alternatively choose $i^* \in I \setminus I^*$ such that that $\mathbf{a}_{i^*}^f \neq \mathbf{0}$ (in general, if $|I| > 1$, we may assume that there exists $i \in I$ such that $\mathbf{a}_i^f \neq \mathbf{0}$ because if $\mathbf{a}_i^f = \mathbf{0}$ for all $i \in I$, then all $(\hat{\mathbf{a}}_i, b_i)$ would be linearly dependent.) Regardless of how we do so, if we take some ad hoc approach to choosing a nontrivial solution to formulation (2.3), such a solution would be suboptimal, which suggests that the IO problem is somehow ill-posed for such problem data.

2.2 Duality gap minimization

In this section we propose an alternative IO model that can be used when it is not clear a priori that strong duality can be achieved exactly, such as when there are external constraints on $\mathbf{A}$ motivated by the application domain of the problem. In this case, we consider a model variant which minimizes the duality gap, subject to some constraints $\mathbf{A} \in \mathbf{\Omega}$ and the remaining constraints from formulation (2.3):

$$\begin{aligned}
\underset{\mathbf{A},\mathbf{c},\boldsymbol{\pi}}{\text{minimize}} \quad & \sum_{j\in J} c_j \hat{x}_j - \sum_{i\in I} b_i \pi_i && (2.15a)\\
\text{subject to} \quad & \mathbf{A} \in \boldsymbol{\Omega}, && (2.15b)\\
& (2.3c) - (2.3f). && (2.15c)
\end{aligned}$$

Whereas the previous IO model was feasible regardless of the problem data, the feasibility of formulation (2.15) is determined by whether or not $\boldsymbol{\Omega}$ allows for primal feasibility of the forward problem. We omit the proof of this result, which is straightforward.

Proposition 4. *Formulation (2.15) is feasible if and only if there exists $\mathbf{A} \in \boldsymbol{\Omega}$ such that $\mathbf{a}_i^\intercal \hat{\mathbf{x}} \geq b_i$ for all $i \in I$.*

Formulation (2.15) is nonconvex for the same reason as formulation (2.3), but the inclusion of constraints $\mathbf{A} \in \boldsymbol{\Omega}$ will require a different solution method:

Theorem 2. *For all $i \in I$, let*

$$t_i = \min_{\mathbf{A}} \left\{ \sum_{j\in J} a_{ij}\hat{x}_j - b_i \colon \mathbf{A} \in \boldsymbol{\Omega}, \mathbf{A}\hat{\mathbf{x}} \geq \mathbf{b} \right\}, \tag{2.16}$$

and let $\underline{\mathbf{A}}^i$ be an optimal solution for (2.16). Let $i^* \in \arg\min_{i\in I}\{t_i\}$. Then the optimal value of formulation (2.15) is t_{i^*}, and an optimal solution $(\mathbf{A}, \mathbf{c}, \boldsymbol{\pi})$ is

$$\mathbf{a}_i = \underline{\mathbf{a}}_i^{i^*}, \quad \forall i \in I, \tag{2.17}$$

$$\mathbf{c} = \underline{\mathbf{a}}_{i^*}^{i^*}, \tag{2.18}$$

$$\boldsymbol{\pi} = \mathbf{e}_{i^*}. \tag{2.19}$$

Remark 2. Theorem 2 shows that an optimal solution to the nonconvex inverse problem (2.15) can be found by solving $|I|$ optimization problems which are linear whenever the set $\boldsymbol{\Omega}$ is linear.

Proof: Substituting (2.3e) into the objective function (2.15a), we get the problem

$$\begin{aligned}
\underset{\mathbf{A},\boldsymbol{\pi}}{\text{minimize}} \quad & \sum_{i\in I} \pi_i \left(\sum_{j\in J} a_{ij}\hat{x}_j - b_i \right)\\
\text{subject to} \quad & \mathbf{A} \in \boldsymbol{\Omega},\ \mathbf{A}\hat{\mathbf{x}} \geq \mathbf{b},\\
& \mathbf{e}^\intercal \boldsymbol{\pi} = 1,\ \boldsymbol{\pi} \geq \mathbf{0}.
\end{aligned} \tag{2.20}$$

For a given feasible $\mathbf{A}$, it is clear that an optimal $\boldsymbol{\pi} = \mathbf{e}_{i^*}$, where $i^* \in \arg\min_{i\in I}\{\sum_{j\in J} a_{ij}\hat{x}_j - b_i\}$. Problem (2.20) is therefore equivalent to $\min_{i\in I}\left\{\min_{\mathbf{A}}\left\{\sum_{j\in J} a_{ij}\hat{x}_j - b_i \colon \mathbf{A} \in \boldsymbol{\Omega}, \mathbf{A}\hat{\mathbf{x}} \geq \mathbf{b}\right\}\right\}$.

By definition, $\underline{\mathbf{A}}^i$ is an optimal solution for the inner problem, and the optimal value of the outer problem is $\min_{i \in I}\{t_i\}$. Finally, $\boldsymbol{\pi} = \mathbf{e}_{i^*}$ and (2.3e) imply that $\mathbf{c} = \underline{\mathbf{a}}_{i^*}^{i^*}$. □

Theorem 2 and its proof can be interpreted as follows. For all $i \in I$, t_i is the minimum achievable surplus for constraint i, while respecting primal feasibility and the constraints $\mathbf{A} \in \boldsymbol{\Omega}$. Because of the normalization constraint (2.3d), the duality gap is equal to a convex combination of the surpluses of the constraints of the forward problem. The minimum possible duality gap will therefore equal the surplus of some constraint i^*, and the optimal choice of this constraint is the one with the minimum possible surplus, i.e., $i^* \in \arg\min_{i \in I}\{t_i\}$. The constraint vectors are then chosen such that the surplus of constraint i^* equals t_{i^*}, and the cost vector is set perpendicular to constraint i^*.

In Sections 2.1 and 2.2, we have derived tractable solution approaches for the nonconvex IO models (2.3) and (2.15), which recover a constraint matrix $\mathbf{A}$. The former model minimally perturbs prior estimated parameters such that there exists a nonzero cost vector rendering the observed solution exactly optimal; the latter model identifies parameters from a predefined set $\boldsymbol{\Omega}$ such that there exists a cost vector minimizing the duality gap. The choice of which model to use thus depends on whether the application domain motivates constraints of the form $\mathbf{A} \in \boldsymbol{\Omega}$. If the model (2.15) is found to have an optimal value of zero, then this implies that the observed solution was in fact exactly optimal with respect to some $\mathbf{A} \in \boldsymbol{\Omega}$, in which case a reasonable next step would be to attempt to solve model (2.3) with the addition of the constraints $\mathbf{A} \in \boldsymbol{\Omega}$. Although not shown here, this variant of the IO problem is also solvable, following similar reasoning as in the proof of Theorem 2.

2.3 Numerical examples

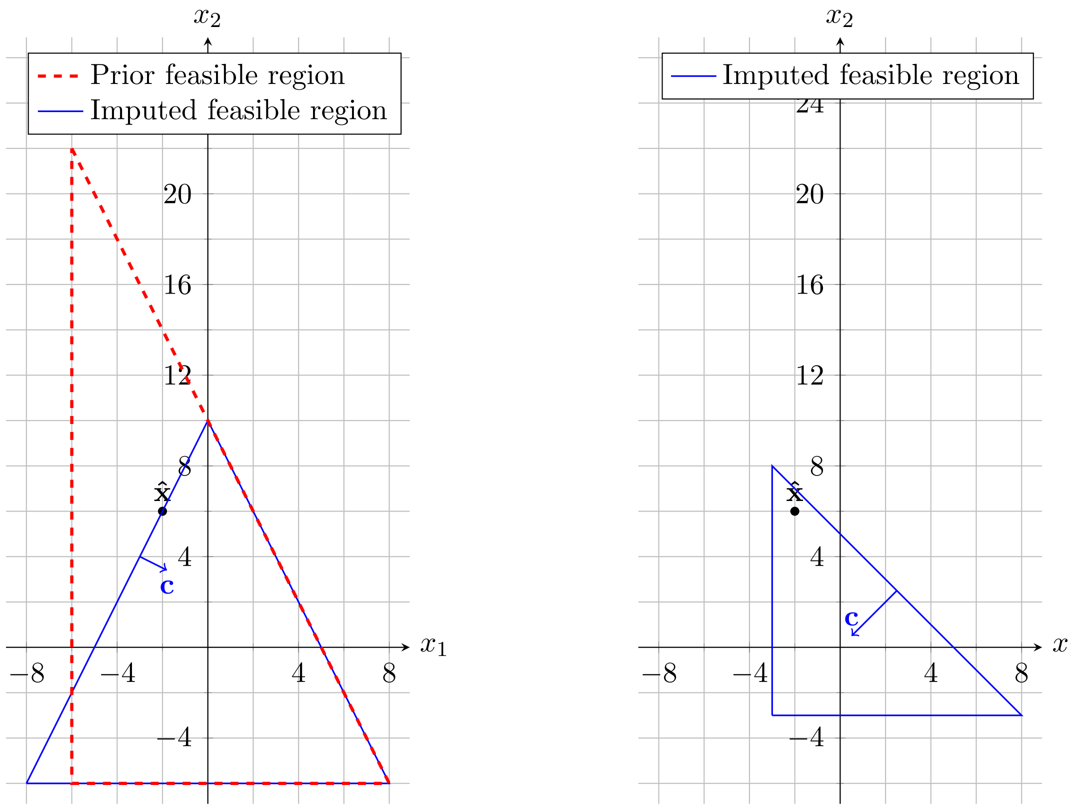
In this section, we give numerical examples to illustrate how the optimal inverse solutions are found for formulations (2.3) and (2.15), and the geometric characteristics of these solutions.

2.3.1 Strong duality

We give three examples illustrating the solution of formulation (2.3). For all examples in this subsection, we let the norm in the objective function (2.3a) be the Euclidean norm, and we let $\boldsymbol{\xi} = \mathbf{e}$ for simplicity. For our first example, let the observed solution be $\hat{\mathbf{x}} = (-2, 0.6)$, and let the remaining problem data be

$$\hat{\mathbf{A}} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -2 & -1 \end{pmatrix}, \ \mathbf{b} = \begin{pmatrix} -6 \\ -6 \\ -10 \end{pmatrix}.$$

The prior feasible region defined by $(\hat{\mathbf{A}}, \mathbf{b})$ is shown in Figure 2.1a. We next apply Theorem 1. Since $\hat{\mathbf{x}}$ is feasible with respect to $(\hat{\mathbf{A}}, \mathbf{b})$, we find that $g_i = 0$ for all $i \in I$. We also find that $\mathbf{f} = (0.63, 1.90, 1.26)$, thus $i^* = 1$. This means that while the other two constraints will remain



(a) Strong duality. Because the observed solution is an interior point of the prior feasible region, the optimal solution of the IO model (2.3) only needs to adjust a single constraint such that it is rendered active, and sets the cost vector perpendicular to this constraint.

(b) Duality gap minimization. The constraints on the unspecified coefficients do not allow for a feasible region with the observed solution on its boundary. The IO model (2.15) instead minimizes the surplus of a single constraint, and sets the cost vector perpendicular to this constraint.

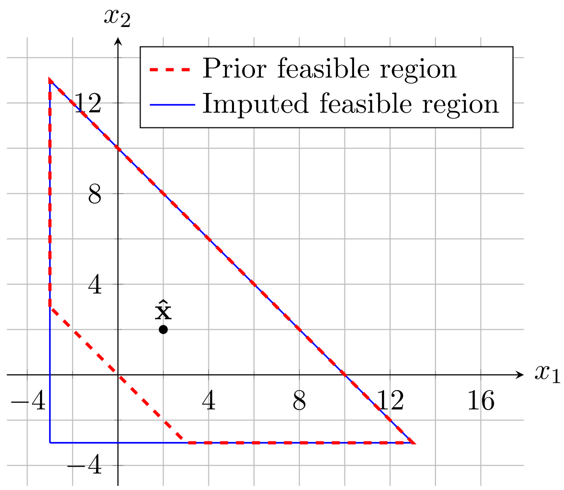
Figure 2.1: Numerical examples of the IO models. Both examples share the same observed solution.

at their prior settings, the first constraint will be adjusted such that $\mathbf{a}_1 = \mathbf{a}_1^f = (1.2, -6)$, rendering the constraint active. Additionally, we set the cost vector perpendicular to the first constraint, as shown in Figure 2.1a.

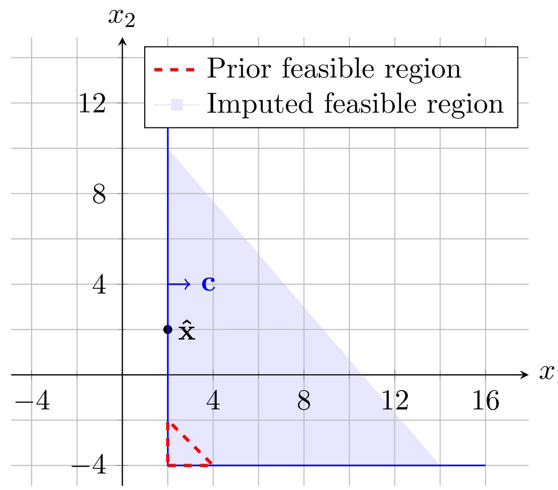
Next, we give two examples to illustrate that formulation (2.3) can have a trivial solution in some situations. Consider a problem in which the observed solution is $\hat{\mathbf{x}} = (2, 2)$, and our remaining problem data is

$$\hat{\mathbf{A}} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ -1 & -1 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} -3 \\ -3 \\ 0 \\ -10 \end{pmatrix}.$$

The prior feasible region is illustrated in Figure 2.2a. Applying Theorem 1, we find $\mathbf{g} = \mathbf{0}$ and



(a) Trivial cost vector. The observed solution is an interior point of the prior feasible region, so the IO model is expected to adjust a single constraint such that it becomes active. The IO model achieves this artificially by setting all coefficients of one constraint equal to zero, effectively eliminating the constraint and implying a zero cost vector.



(b) Elimination of prior infeasible constraint. The observed solution is infeasible with respect to the one of the prior constraints, and active with respect to another. The IO model is expected to adjust the infeasible constraint to achieve feasibility, but it does so by setting all constraint coefficients to zero, thus eliminating the constraint altogether.

Figure 2.2: Numerical examples in which formulation (2.3) produces trivial solutions.

$\mathbf{f} = (1.77, 1.77, 1.41, 2.12)$, thus $i^* = 3$. However, $b_3 = 0$ and $\hat{\mathbf{x}} = 2\hat{\mathbf{a}}_3$, thus $\mathbf{a}_3 = \mathbf{a}_3^f = (0, 0)$ and $\mathbf{c} = \mathbf{0}$ as per Proposition 2. This solution effectively means that the imputed feasible region is obtained from the prior feasible region by eliminating the third constraint, as shown in Figure 2.2a. The observed $\hat{\mathbf{x}}$ remains an interior point, thus is not optimal with respect to any nonzero cost vector.

Next, consider a problem using the same $\hat{\mathbf{x}} = (2, 2)$, but with the following constraint data:

$$\hat{\mathbf{A}} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 2 \\ -4 \\ 0 \end{pmatrix}.$$

This example differs from our previous examples in two respects. First, $\hat{\mathbf{x}}$ sits on the boundary of the first constraint of the prior feasible region. As we expect from this scenario, $i^* = 1$ and $\mathbf{c} = \hat{\mathbf{a}}_1 = (1, 0)$. Second, $\hat{\mathbf{x}}$ is infeasible with respect to $\hat{\mathbf{a}}_3$, and thus the third constraint must be adjusted such that $\hat{\mathbf{x}}$ becomes feasible. However, $b_3 = 0$ and $\hat{\mathbf{x}} = -2\hat{\mathbf{a}}_3$, thus $\mathbf{a}_3^g = (0, 0)$ as per Proposition 3. In other words, the third constraint is eliminated and the feasible region is rendered unbounded, as shown in Figure 2.2b. This solution is meaningful insofar as $\hat{\mathbf{x}}$ does lie on the boundary of the imputed feasible region and is therefore optimal, but it appears unreasonable to claim that the best way to shift the third constraint to achieve feasibility is simply to eliminate it entirely; we can imagine nonzero candidates for $\mathbf{a}_3$ that achieve feasibility and re-shape the feasible region less drastically.

2.3.2 Duality gap minimization

Finally, we give an example illustrating the solution of formulation (2.15). Let the observed solution be $\hat{\mathbf{x}} = (-2, 6)$, let $\mathbf{b} = (-6, -6, -10)$, and define the following constraints on $\mathbf{A}$:

$$\begin{aligned}\mathbf{\Omega} := \{&0.5 \leq a_{11} \leq 1.5,\\ &0.5 \leq a_{22} \leq 1.5,\\ &a_{12} = 0,\\ &a_{21} = 0,\\ &a_{31} \leq -1.5,\\ &-2 \leq a_{32} \leq -0.5,\\ &a_{31} + 2a_{22} \leq -1\}.\end{aligned}$$

It is easy to check that there exists $\mathbf{A} \in \mathbf{\Omega}$ such that $\mathbf{A}\hat{\mathbf{x}} \geq \mathbf{b}$, and hence formulation (2.15) is feasible. Applying Theorem 2, we find that $\mathbf{t} = (3, 9, 2)$ and hence $i^* = 3$. The duality gap-minimizing constraint matrix is then

$$\underline{\mathbf{A}}^3 = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \\ -2 & -2 \end{pmatrix},$$

and the optimal cost vector is $\mathbf{c} = \underline{\mathbf{a}}^3 = (-2, -2)$. These results are illustrated in Figure 2.1b. In contrast to the example in Figure 2.1a, $\hat{\mathbf{x}}$ is an interior point of the imputed feasible region due to the constraints $\mathbf{A} \in \mathbf{\Omega}$. To minimize the duality gap, the cost vector is set perpendicular to the constraint with the minimum surplus.

Chapter 3

Inverse robust linear optimization

In this chapter, we consider the robust linear optimization problem $\min_{\mathbf{x}}\{\mathbf{c}^\intercal\mathbf{x} \colon \tilde{\mathbf{A}}\mathbf{x} \geq \mathbf{b},\ \forall \tilde{\mathbf{A}} \in \mathcal{U}\}$, for two basic types of uncertainty set $\mathcal{U}$: interval uncertainty and cardinality constrained uncertainty. Given $\mathbf{b}$, the nominal estimate $\mathbf{A}$, and an observed solution $\hat{\mathbf{x}}$, the IO problem aims to recover unspecified parameters of the uncertainty set such that $\hat{\mathbf{x}}$ is optimal for some nonzero cost vector. As in Chapter 2, we consider two variants of this problem: the first induces a zero duality gap, and the second minimizes the duality gap in case it may not be possible to make $\hat{\mathbf{x}}$ exactly optimal. We have restricted our forward problem to only include uncertainty on the left-hand-side constraint matrix, but using similar reasoning as in the discussion of formulation (2.2), we can show that the problem remains general enough to account for uncertainty on all constraint coefficients.

The remainder of this chapter is organized as follows. Sections 3.1 and 3.2 propose IO models for interval uncertainty and cardinality constrained uncertainty, respectively. These two sections are conceptually similar to Chapter 2, so we will draw parallels to previous results and omit details wherever we believe they would be redundant. Section 3.3 provides numerical examples that illustrate how the geometries of the two types of robust feasible region are controlled by the inverse solutions.

3.1 Interval uncertainty

In this section, we consider a robust linear optimization problem with interval uncertainty. Let $J_i \subseteq J$ index the coefficients in the i-th row of $\mathbf{A}$ that are subject to uncertainty, and let $\mathbf{a}_i, b_i, \boldsymbol{\alpha}_i$ be given for all $i \in I$. Following is the robust problem:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \sum_{j \in J} c_j x_j \\ \text{subject to} \quad & \sum_{j \in J_i} \tilde{a}_{ij} x_j + \sum_{j \notin J_i} a_{ij} x_j \geq b_i, \quad \forall \tilde{a}_{ij} \in [a_{ij} - \alpha_{ij}, a_{ij} + \alpha_{ij}], i \in I. \end{aligned} \tag{3.1}$$

The constraints can be written as $\sum_{j \in J} a_{ij} x_j - \sum_{j \in J_i} \alpha_{ij} |x_j| \geq b_i, \; \forall i \in I$, which leads to the following linearization (Ben-Tal and Nemirovski, 2000):

$$\begin{aligned} \underset{\mathbf{x},\mathbf{u}}{\text{minimize}} \quad & \sum_{j \in J} c_j x_j && \text{(3.2a)} \\ \text{subject to} \quad & \alpha_{ij} x_j + u_{ij} \geq 0, \quad \forall j \in J_i, i \in I, && \text{(3.2b)} \\ & -\alpha_{ij} x_j + u_{ij} \geq 0, \quad \forall j \in J_i, i \in I, && \text{(3.2c)} \\ & \sum_{j \in J} a_{ij} x_j - \sum_{j \in J_i} u_{ij} \geq b_i, \quad \forall i \in I. && \text{(3.2d)} \end{aligned}$$

Given $\mathbf{a}_i, b_i$ and J_i for all $i \in I$, and a feasible $\hat{\mathbf{x}}$ for the nominal problem (i.e., formulation (3.1) with $J_i = \varnothing$ for all i), the goal of the IO problem is to determine nonnegative parameters $\boldsymbol{\alpha}_i$ for all $i \in I$ defining the uncertainty set. For simplicity, we assume that every row has at least one coefficient that is subject to uncertainty (if we did not make this assumption, we would define $\hat{I} := \{i \in I \colon J_i \neq \varnothing\}$ and replace I with $\hat{I}$ in relevant places throughout the following development). Additionally, we assume that there is some $i \in I$ and $j \in J_i$ such that $\hat{x}_j \neq 0$ (otherwise, all α_{ij} would be multiplied by zero and modifying $\boldsymbol{\alpha}$ would not change the robust feasible region).

3.1.1 Strong duality

Let $\lambda_{ij}, \mu_{ij}, \pi_i$ be the dual variables corresponding to constraints (3.2b)-(3.2d), respectively. The following formulation minimizes the weighted deviation of the uncertainty set parameters $\boldsymbol{\alpha}_i$ from given values $\hat{\boldsymbol{\alpha}}_i$ while enforcing strong duality, and primal and dual feasibility:

$$\begin{aligned} \underset{\boldsymbol{\alpha},\mathbf{c},\mathbf{u},\boldsymbol{\pi},\boldsymbol{\lambda},\boldsymbol{\mu}}{\text{minimize}} \quad & \sum_{i \in I} \xi_i \|\boldsymbol{\alpha}_i - \hat{\boldsymbol{\alpha}}_i\| && \text{(3.3a)} \\ \text{subject to} \quad & \sum_{j \in J} c_j \hat{x}_j - \sum_{i \in I} b_i \pi_i = 0, && \text{(3.3b)} \\ & \alpha_{ij} \hat{x}_j + u_{ij} \geq 0, \quad \forall j \in J_i, i \in I, && \text{(3.3c)} \\ & -\alpha_{ij} \hat{x}_j + u_{ij} \geq 0, \quad \forall j \in J_i, i \in I, && \text{(3.3d)} \\ & \sum_{j \in J} a_{ij} \hat{x}_j - \sum_{j \in J_i} u_{ij} \geq b_i, \quad \forall i \in I, && \text{(3.3e)} \\ & \alpha_{ij} \geq 0, \quad \forall j \in J_i, i \in I, && \text{(3.3f)} \\ & \sum_{i \in I} \pi_i = 1, && \text{(3.3g)} \\ & \sum_{i \in I} a_{ij} \pi_i + \sum_{i \in I \colon j \in J_i} \alpha_{ij} (\lambda_{ij} - \mu_{ij}) = c_j, \quad \forall j \in J, && \text{(3.3h)} \\ & \pi_i = \lambda_{ij} + \mu_{ij}, \quad \forall j \in J_i, i \in I, && \text{(3.3i)} \\ & \pi_i, \lambda_{ij}, \mu_{ij} \geq 0, \quad \forall j \in J_i, i \in I. && \text{(3.3j)} \end{aligned}$$

Formulation (3.3) is constructed in a conceptually similar manner as formulation (2.3). Constraints (3.3b), (3.3c)-(3.3e), and (3.3h)-(3.3j) represent strong duality, primal feasibility, and dual feasibility, respectively. To prevent the trivial solution $(\mathbf{c}, \boldsymbol{\pi}) = (\mathbf{0}, \mathbf{0})$ from being feasible, we again include the normalization constraint (3.3g). All constraints of formulation (3.3) are linear except for the bilinear dual feasibility constraint (3.3h), but nevertheless we will be able to determine an efficient solution method.

First, we show that feasibility of (3.3) is entirely determined by feasibility of $\hat{\mathbf{x}}$ with respect to the nominal problem.

Proposition 5. *Formulation (3.3) is feasible if and only if* $\sum_{j\in J} a_{ij}\hat{x}_j \geq b_i$ *for all* $i \in I$.

The geometric intuition underlying Proposition 5 is twofold: the robust feasible region is a subset of the nominal feasible region for any choice of $\boldsymbol{\alpha}$, and $\hat{\mathbf{x}}$ must lie on the boundary of the robust feasible region in order to be optimal. Hence if $\hat{\mathbf{x}}$ is feasible for the nominal problem, it is possible to set $\boldsymbol{\alpha}$ that shrinks the feasible region such that $\hat{\mathbf{x}}$ renders some constraint active. And conversely, if $\hat{\mathbf{x}}$ is not feasible for the nominal problem, then there is no way to grow the feasible region such that $\hat{\mathbf{x}}$ lies on the boundary, or is even feasible. Whereas Proposition 5 formalizes this intuition into a condition that can be used to check whether the IO problem is feasible, the following Lemma formalizes this intuition in a way that will be useful for our solution method (cf. Lemma 1):

Lemma 2. *Every feasible solution for formulation (3.3) satisfies*

$$\sum_{j\in J} a_{\hat{i}j}\hat{x}_j - \sum_{j\in J_{\hat{i}}} \alpha_{\hat{i}j}|\hat{x}_j| = b_{\hat{i}}, \quad \textit{for some } \hat{i} \in I, \tag{3.4a}$$

$$\sum_{j\in J} a_{ij}\hat{x}_j - \sum_{j\in J_i} \alpha_{ij}|\hat{x}_j| \geq b_i, \quad \forall i \in I, \tag{3.4b}$$

$$\alpha_{ij} \geq 0, \quad \forall j \in J_i, i \in I. \tag{3.4c}$$

Conversely, for every $\boldsymbol{\alpha}$ *satisfying (3.4), there exists* $(\mathbf{c}, \mathbf{u}, \boldsymbol{\pi}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ *such that* $(\boldsymbol{\alpha}, \mathbf{c}, \mathbf{u}, \boldsymbol{\pi}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ *is feasible for formulation (3.3).*

Proof: To prove the first statement, we first note that constraints (3.3c)-(3.3f) imply (3.4b). To complete the proof, we assume $\sum_{j\in J} a_{ij}\hat{x}_j - \sum_{j\in J_i} \alpha_{ij}|\hat{x}_j| > b_i$ for all $i \in I$ and derive a contradiction. Substituting (3.3h) into (3.3b), we get

$$\sum_{i\in I} \pi_i \sum_{j\in J} a_{ij}\hat{x}_j + \sum_{i\in I}\sum_{j\in J_i} (\lambda_{ij} - \mu_{ij})\alpha_{ij}\hat{x}_j = \sum_{i\in I} \pi_i b_i.$$

Let $s_{ij} = \lambda_{ij} - \mu_{ij}$ for all $j \in J_i, i \in I$. By Lemma 5 (see appendix), $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ satisfies (3.3i) and (3.3j) if and only if $s_{ij} \in [-\pi_i, \pi_i]$ for all $j \in J_i, i \in I$. Thus, the feasible region of (3.3) is equivalent to

$$\sum_{i\in I}\pi_i\sum_{j\in J}a_{ij}\hat{x}_j+\sum_{i\in I}\sum_{j\in J_i}s_{ij}\alpha_{ij}\hat{x}_j=\sum_{i\in I}\pi_i b_i, \tag{3.5a}$$

$$-\pi_i\le s_{ij}\le\pi_i,\quad \forall j\in J_i, i\in I, \tag{3.5b}$$

$$\mathbf{e}^\mathsf{T}\boldsymbol{\pi}=1, \boldsymbol{\pi}\ge\mathbf{0}, \tag{3.5c}$$

$$(3.3c)-(3.3f). \tag{3.5d}$$

Constraint (3.5c) ensures that $\bar{I} := \{i\in I\colon \pi_i>0\}\neq\varnothing$. Since we have assumed $\sum_{j\in J}a_{ij}\hat{x}_j-\sum_{j\in J_i}\alpha_{ij}|\hat{x}_j|>b_i$ for all $i\in I$, we have

$$\pi_i\sum_{j\in J}a_{ij}\hat{x}_j-\pi_i\sum_{j\in J_i}\alpha_{ij}|\hat{x}_j|>\pi_i b_i,\quad \forall i\in\bar{I}.$$

For all $j\in J_i, i\in I$, $s_{ij}\in[-\pi_i,\pi_i]$ implies that $\sum_{j\in J_i}s_{ij}\alpha_{ij}\hat{x}_j\ge-\pi_i\sum_{j\in J_i}\alpha_{ij}|\hat{x}_j|$, and therefore

$$\pi_i\sum_{j\in J}a_{ij}\hat{x}_j+\sum_{j\in J_i}s_{ij}\alpha_{ij}\hat{x}_j>\pi_i b_i,\quad \forall i\in\bar{I}.$$

Since $s_{ij}=0$ if $\pi_i=0$,

$$\sum_{i\in I}\pi_i\sum_{j\in J}a_{ij}\hat{x}_j+\sum_{i\in I}\sum_{j\in J_i}s_{ij}\alpha_{ij}\hat{x}_j>\sum_{i\in I}\pi_i b_i,$$

which contradicts constraint (3.5a).

To prove the second statement, let $\boldsymbol{\alpha}$ satisfy (3.4), $\hat{\imath}$ be defined by (3.4a), and $(\mathbf{c},\mathbf{u},\boldsymbol{\pi},\boldsymbol{\lambda},\boldsymbol{\mu})$ be defined as in the proof of Proposition 5. This solution is feasible for formulation (3.3).

□

Lemma 2 allows us to characterize an optimal solution to (3.3) and devise an efficient solution method (cf. Theorem 1):

Theorem 3. *For all $i\in I$, let*

$$f_i=\min_{\boldsymbol{\alpha}_i\ge\mathbf{0}}\left\{\xi_i\|\boldsymbol{\alpha}_i-\hat{\boldsymbol{\alpha}}_i\|\colon \sum_{j\in J}a_{ij}\hat{x}_j-\sum_{j\in J_i}\alpha_{ij}|\hat{x}_j|=b_i\right\}, \tag{3.6}$$

$$g_i=\min_{\boldsymbol{\alpha}_i\ge\mathbf{0}}\left\{\xi_i\|\boldsymbol{\alpha}_i-\hat{\boldsymbol{\alpha}}_i\|\colon \sum_{j\in J}a_{ij}\hat{x}_j-\sum_{j\in J_i}\alpha_{ij}|\hat{x}_j|\ge b_i\right\}, \tag{3.7}$$

and let $\boldsymbol{\alpha}_i^f$ *and* $\boldsymbol{\alpha}_i^g$ *be optimal solutions for* (3.6) *and* (3.7)*, respectively. Let*

$$c_j^i = \begin{cases} a_{ij} - \operatorname{sgn}(\hat{x}_j)\alpha_{ij}^f & \text{if } j \in J_i, \\ a_{ij} & \text{if } j \in J \setminus J_i, \end{cases} \quad \forall i \in I, \tag{3.8}$$

$$i^* \in \arg\min_{i \in I} \{f_i - g_i\}. \tag{3.9}$$

Then the optimal value of formulation (3.3) *is* $f_{i^*} + \sum_{i \neq i^*, i \in I} g_i$*, and there exists an optimal solution of* (3.3) *with*

$$\boldsymbol{\alpha}_i = \begin{cases} \boldsymbol{\alpha}_i^f & \text{if } i = i^*, \\ \boldsymbol{\alpha}_i^g & \text{if } i \neq i^*, i \in I, \end{cases} \tag{3.10}$$

$$\mathbf{c} = \mathbf{c}^{i^*}. \tag{3.11}$$

Remark 3. Theorem 3 shows that an optimal solution to the nonconvex inverse problem (3.3) can be found by solving $2|I|$ convex problems (linear with appropriate choice of $\|\cdot\|$).

Proof: By Lemma 2, solving formulation (3.3) is equivalent to solving the following optimization problem for all $\hat{i} \in I$, and taking the minimum over all $|I|$ optimal values:

$$\begin{aligned} \underset{\boldsymbol{\alpha}}{\text{minimize}} \quad & \sum_{i \in I} \xi_i \|\boldsymbol{\alpha}_i - \hat{\boldsymbol{\alpha}}_i\| \\ \text{subject to} \quad & \sum_{j \in J} a_{\hat{i}j}\hat{x}_j - \sum_{j \in J_{\hat{i}}} \alpha_{\hat{i}j}|\hat{x}_j| = b_{\hat{i}} \\ & \sum_{j \in J} a_{ij}\hat{x}_j - \sum_{j \in J_i} \alpha_{ij}|\hat{x}_j| \geq b_i, \quad \forall i \in I, \\ & \alpha_{ij} \geq 0, \quad \forall j \in J_i, i \in I. \end{aligned} \tag{3.12}$$

Suppose we fix some $\hat{i} \in I$. Since formulation (3.12) is separable by i, the optimal value of the $\hat{i}$-th formulation (3.12) is $f_{\hat{i}} + \sum_{i \neq \hat{i}, i \in I} g_i$. Therefore, the optimal value of formulation (3.3) is

$$\min_{\hat{i} \in I} \left\{ f_{\hat{i}} + \sum_{i \neq \hat{i}, i \in I} g_i \right\}.$$

Clearly, the optimal index i^* must satisfy $i^* \in \arg\min_{i \in I}\{f_i - g_i\}$. An optimal $\boldsymbol{\alpha}$ is derived from the optimal solutions of (3.6) and (3.7) for i^*,

$$\boldsymbol{\alpha}_i = \begin{cases} \boldsymbol{\alpha}_i^f & \text{if } i = i^*, \\ \boldsymbol{\alpha}_i^g & \text{if } i \neq i^*, i \in I, \end{cases} \tag{3.13}$$

and the optimal cost vector is $\mathbf{c} = \mathbf{c}^{i^*}$, where the structure of $\mathbf{c}^i$ is derived in the proof of Proposition 5. □

The interpretation of Theorem 3 is conceptually identical to the interpretation of Theorem

1. The primary difference between the two results is that although both set the cost vector perpendicular to the active constraint, equation (3.11) more specifically sets the cost vector perpendicular to the part of constraint i^* that is contained in the same orthant as $\hat{\mathbf{x}}$. Note that the vector perpendicular to a robust constraint changes as the constraint crosses into different orthants (cf. Figure 3.1). The optimal cost vector $\mathbf{c}^{i^*}$ captures this geometric situation.

3.1.2 Duality gap minimization

As in Chapter 2, we propose an alternative model that minimizes the duality gap, subject to some constraints $\boldsymbol{\alpha} \in \boldsymbol{\Omega}$ and the remaining constraints from formulation (3.3):

$$\underset{\boldsymbol{\alpha},\mathbf{c},\mathbf{u},\boldsymbol{\pi},\boldsymbol{\lambda},\boldsymbol{\mu}}{\text{minimize}} \quad \sum_{j\in J} c_j \hat{x}_j - \sum_{i\in I} b_i \pi_i \tag{3.14a}$$

$$\text{subject to} \quad \boldsymbol{\alpha} \in \boldsymbol{\Omega}, \tag{3.14b}$$

$$(3.3c) - (3.3j). \tag{3.14c}$$

The feasibility of $\hat{\mathbf{x}}$ with respect to the nominal problem remains a necessary condition for the feasibility of (3.14). However, it is not sufficient anymore because the constraints on $\boldsymbol{\alpha}$ may result in a robust feasible region that excludes $\hat{\mathbf{x}}$. The feasibility of (3.14) is thus determined by whether or not $\boldsymbol{\Omega}$ allows for primal feasibility of the robust optimization problem. We omit the proof of this result since it is very similar to the proof of Proposition 5.

Proposition 6. *Formulation (3.14) is feasible if and only if there exists nonnegative $\boldsymbol{\alpha} \in \boldsymbol{\Omega}$ such that $\sum_{j\in J} a_{ij}\hat{x}_j - \sum_{j\in J_i} \alpha_{ij}|\hat{x}_j| \geq b_i$ for all $i \in I$.*

Formulation (3.14) is again nonconvex due to constraint (3.3h), but we were unable to find a general solution of the form in Theorem 3 because of the constraints on $\boldsymbol{\alpha}$. However, we will show that optimal solutions for (3.14) can be found by solving the following mixed integer optimization model, which is mixed integer linear when the constraints on $\boldsymbol{\alpha}$ are linear:

$$\underset{\boldsymbol{\alpha},\mathbf{u},\boldsymbol{\pi},t}{\text{minimize}} \quad t \tag{3.15a}$$

$$\text{subject to} \quad \boldsymbol{\alpha} \in \boldsymbol{\Omega}, \tag{3.15b}$$

$$(3.3c) - (3.3f), \tag{3.15c}$$

$$\sum_{i\in I} \pi_i = 1, \tag{3.15d}$$

$$t \geq \sum_{j\in J} a_{ij}\hat{x}_j - \sum_{j\in J_i} \alpha_{ij}|\hat{x}_j| - b_i - M(1-\pi_i), \quad \forall i \in I, \tag{3.15e}$$

$$\pi_i \in \{0,1\}, \quad \forall i \in I. \tag{3.15f}$$

Formulation (3.15) can be interpreted as follows. Constraints (3.15b) and (3.15c) are retained from formulation (3.14), requiring that external constraints on $\boldsymbol{\alpha}$ and primal feasibility be satisfied. The duality gap is represented by the auxiliary variable t, and as a result of the normalization constraint, the duality gap is equal to the surplus of a single constraint $i \in I$. The choice of this constraint is encoded in the binary vector $\boldsymbol{\pi}$ and the optimal choice i^* corresponds to the constraint with the minimum surplus.

Although formulation (3.15) does not explicitly include the variable $\mathbf{c}$, it can be determined post-optimization using the equation

$$c_j = \sum_{i \in I} a_{ij} \pi_i - \sum_{i \in I:\, j \in J_i} \alpha_{ij} \operatorname{sgn}(\hat{x}_j) \pi_i, \ \forall j \in J, \tag{3.16}$$

which is derived from equation (3.3h) and by letting $(\lambda_{ij} - \mu_{ij}) = -\operatorname{sgn}(\hat{x}_j)\pi_i$, for all $j \in J_i, i \in I$ (cf. proof of Theorem 4 below). The geometric interpretation of this solution is the same as in formulation (3.3): if $\boldsymbol{\pi} = \mathbf{e}_{i^*}$, then the cost vector is set perpendicular to the part of constraint i^* that is contained in the same orthant as $\hat{\mathbf{x}}$.

We now formally characterize and prove the correspondence between formulations (3.14) and (3.15):

Theorem 4. *Let $M \geq \max_{i \in I}\{\sum_{j \in J} a_{ij} \hat{x}_j - b_i\}$. Formulations (3.14) and (3.15) have the same optimal objective value, and a solution $\boldsymbol{\alpha}$ is optimal for formulation (3.14) if and only if it is optimal for formulation (3.15).*

Proof: First, we eliminate $\mathbf{c}$ by substituting the dual feasibility constraint (3.3h) into the objective function (3.14a). The resulting model has an objective function that is bilinear in variables whose corresponding feasible sets $P = \{(\boldsymbol{\alpha}, \mathbf{u}) \colon \boldsymbol{\alpha} \in \boldsymbol{\Omega}, (3.3c) - (3.3f)\}$ and $D = \{(\boldsymbol{\pi}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \colon (3.3g), (3.3i) - (3.3j)\}$ are disjoint:

$$\underset{\substack{(\boldsymbol{\alpha},\mathbf{u}) \in P,\\ (\boldsymbol{\pi},\boldsymbol{\lambda},\boldsymbol{\mu}) \in D}}{\text{minimize}} \quad \sum_{i \in I} \sum_{j \in J} a_{ij} \pi_i \hat{x}_j + \sum_{i \in I} \sum_{j \in J_i} \alpha_{ij} (\lambda_{ij} - \mu_{ij}) \hat{x}_j - \sum_{i \in I} b_i \pi_i. \tag{3.17}$$

Since D is a bounded polyhedron and disjoint from P, an optimal solution to (3.17) exists among the vertices of D (Horst et al., 2000, Proposition 3.1). The constraints $\sum_{i \in I} \pi_i = 1$ and $\pi_i \geq 0$ for all $i \in I$ imply that a vertex of D will satisfy $\pi_{\hat{i}} = 1$ for some $\hat{i} \in I$, and $\pi_i = 0$ for all $i \in I, i \neq \hat{i}$. So it suffices to consider binary π_i. Let $s_{ij} = \lambda_{ij} - \mu_{ij}$ for all $j \in J_i, i \in I$. By Lemma 5 (see appendix), constraint (3.3i) and nonnegativity of $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ are equivalent to $s_{ij} \in [-\pi_i, \pi_i]$ for all $j \in J_i, i \in I$. Thus, formulation (3.17) is equivalent to

$$\begin{aligned}
\underset{(\boldsymbol{\alpha},\mathbf{u}) \in P, \boldsymbol{\pi}, \mathbf{s}}{\text{minimize}} \quad & \sum_{i \in I} \left(\sum_{j \in J} a_{ij} \hat{x}_j - b_i \right) \pi_i + \sum_{i \in I} \sum_{j \in J_i} \alpha_{ij} \hat{x}_j s_{ij} && \text{(3.18a)}\\
\text{subject to} \quad & -\pi_i \leq s_{ij} \leq \pi_i, \quad \forall j \in J_i, i \in I, && \text{(3.18b)}\\
& \sum_{i \in I} \pi_i = 1, && \text{(3.18c)}\\
& \pi_i \in \{0, 1\}, \quad \forall i \in I. && \text{(3.18d)}
\end{aligned}$$

By inspection, we see that for a given $(\boldsymbol{\alpha}, \mathbf{u}, \boldsymbol{\pi})$, an optimal $\mathbf{s}$ satisfies $s_{ij} = -\operatorname{sgn}(\hat{x}_j)\pi_i$. This fact allows us to eliminate $\mathbf{s}$:

$$\begin{aligned} \underset{(\boldsymbol{\alpha},\mathbf{u})\in P,\boldsymbol{\pi}}{\text{minimize}} \quad & \sum_{i\in I}\left(\sum_{j\in J} a_{ij}\hat{x}_j - \sum_{j\in J_i}\alpha_{ij}|\hat{x}_j| - b_i\right)\pi_i && (3.19a)\\ \text{subject to} \quad & \sum_{i\in I}\pi_i = 1, && (3.19b)\\ & \pi_i \in \{0,1\}, \quad \forall i \in I. && (3.19c) \end{aligned}$$

It is clear that given $(\boldsymbol{\alpha}, \mathbf{u})$, the optimal value of (3.19) is

$$\min_{i\in I}\left\{\sum_{j\in J} a_{ij}\hat{x}_j - \sum_{j\in J_i}\alpha_{ij}|\hat{x}_j| - b_i\right\}, \tag{3.20}$$

and an optimal $\boldsymbol{\pi}$ equals $\mathbf{e}_{i^*}$ where

$$i^* \in \arg\min_{i\in I}\left\{\sum_{j\in J} a_{ij}\hat{x}_j - \sum_{j\in J_i}\alpha_{ij}|\hat{x}_j| - b_i\right\}.$$

Since the objective in formulation (3.15) minimizes t, we can ensure the optimal t equals the expression in (3.20) by using the constraint (3.15e) if M is sufficiently large (the constraint will be active for i^* and inactive for other i). To show that $M = \max_{i\in I}\{\sum_{j\in J} a_{ij}\hat{x}_j - b_i\}$ is sufficient, we substitute it into the right-hand side of constraint (3.15e). For $i \neq i^*$, the resulting expression is nonpositive, due to the nonnegativity of $\boldsymbol{\alpha}_i$:

$$\left(\sum_{j\in J} a_{ij}\hat{x}_j - b_i\right) - \max_{k\in I}\left\{\sum_{j\in J} a_{kj}\hat{x}_j - b_k\right\} - \sum_{j\in J_i}\alpha_{ij}|\hat{x}_j| \le 0, \quad \forall i \neq i^*, i \in I. \tag{3.21}$$

For $i = i^*$, the resulting expression is nonnegative due to constraints (3.3c)-(3.3e):

$$0 \le \left(\sum_{j\in J} a_{i^*j}\hat{x}_j - b_{i^*}\right) - \sum_{j\in J_{i^*}}\alpha_{i^*j}|\hat{x}_j|. \tag{3.22}$$

Therefore the optimal t for formulation (3.15) equals the right-hand-side of (3.22), which equals the optimal objective value for formulation (3.14). In the process of deriving formulation (3.15) from (3.14), we have not manipulated $\boldsymbol{\alpha}$. That is, all steps in this proof are equivalences as far as $\boldsymbol{\alpha}$ is concerned. Hence $\boldsymbol{\alpha}$ is optimal for (3.15) if and only if it is optimal for (3.14). □

In this section, we have derived tractable solution approaches for two nonconvex IO models, (3.3) and (3.14), which recover interval uncertainty parameters $\boldsymbol{\alpha}$. Analagous to the models in Chapter 2, the choice of which model to use depends primarily on whether the application domain motivates constraints of the form $\boldsymbol{\alpha} \in \boldsymbol{\Omega}$, which we have only included in the latter model. However, although we have not shown it here, it is possible to solve a variant of formulation (3.3) with the addition of $\boldsymbol{\alpha} \in \boldsymbol{\Omega}$.

3.2 Cardinality constrained uncertainty

In this section, we consider a robust linear optimization problem with a cardinality constrained uncertainty set (Bertsimas and Sim, 2004), assuming a nearly identical setup as in the previous section. For each constraint $i \in I$, this uncertainty set bounds the number of uncertain coefficients $\tilde{a}_{ij}$ that can deviate from their nominal value a_{ij} within the range $[a_{ij} - \alpha_{ij}, a_{ij} + \alpha_{ij}]$, for all $j \in J_i$, using a budget parameter Γ_i. In particular, $\lfloor \Gamma_i \rfloor$ coefficients can take any value within their uncertain intervals, and up to one coefficient can change by at most $(\Gamma_i - \lfloor \Gamma_i \rfloor)\alpha_{ij}$.

$$\underset{\mathbf{x}}{\text{minimize}} \quad \sum_{j \in J} c_j x_j \tag{3.23a}$$

$$\text{subject to} \quad \sum_{j \in J} a_{ij} x_j - \max_{\substack{\{S_i \cup \{t_i\} :\, S_i \subseteq J_i, \\ |S_i| = \lfloor \Gamma_i \rfloor, t_i \in J_i \setminus S_i\}}} \left\{ \sum_{j \in S_i} \alpha_{ij} |x_j| + (\Gamma_i - \lfloor \Gamma_i \rfloor) \alpha_{it_i} |x_{t_i}| \right\} \geq b_i, \quad \forall i \in I. \tag{3.23b}$$

We refer to the embedded maximization problem in the constraint (3.23b) as the *protection function*. When $\Gamma_i = |J_i|$, the protection function equals $\sum_{j \in J_i} \alpha_{ij} |\hat{x}_j|$ and (3.23b) becomes equivalent to the corresponding constraint of the robust linear program with interval uncertainty. Constraint (3.23b) can be linearized to yield the equivalent robust counterpart (Bertsimas and Sim, 2004):

$$\underset{\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{u}}{\text{minimize}} \quad \sum_{j \in J} c_j x_j \tag{3.24a}$$

$$\text{subject to} \quad \alpha_{ij} x_j + u_{ij} \geq 0, \quad \forall j \in J_i, i \in I, \tag{3.24b}$$

$$-\alpha_{ij} x_j + u_{ij} \geq 0, \quad \forall j \in J_i, i \in I, \tag{3.24c}$$

$$y_{ij} + z_i - u_{ij} \geq 0, \quad \forall j \in J_i, i \in I, \tag{3.24d}$$

$$\sum_{j \in J} a_{ij} x_j - \sum_{j \in J_i} y_{ij} - \Gamma_i z_i \geq b_i, \quad \forall i \in I, \tag{3.24e}$$

$$y_{ij}, z_i \geq 0, \quad \forall j \in J_i, i \in I. \tag{3.24f}$$

Given $\mathbf{a}_i, b_i, J_i$ and $\boldsymbol{\alpha}_i$ for all $i \in I$, and a feasible $\hat{\mathbf{x}}$ for the nominal problem, our IO problem aims to determine parameters $\Gamma_i \in [0, |J_i|]$ for all $i \in I$ such that $\hat{\mathbf{x}}$ is optimal for some nonzero cost vector. Note the slight difference from the interval uncertainty case: here, $\boldsymbol{\alpha}_i$ is fixed as opposed to variable, and the new parameter Γ_i is the primary variable in the inverse problem that determines the uncertainty set.

As in the interval uncertainty case, we propose two IO models: the first requires the optimality conditions to be satisfied exactly, while the second minimizes the duality gap. Previously, in the case of interval uncertainty, the first model identified uncertainty set parameters such that some constraint of the robust problem was rendered active, while the second model identified uncertainty set parameters such that the surplus for a single constraint was minimized. In the case of cardinality constrained uncertainty, the two approaches have the same interpretation. However, due to the upper bound on Γ_i, it may be the case that a constraint $i \in I$ cannot be rendered active at $\hat{\mathbf{x}}$ for any feasible choice of Γ_i. Therefore, we define the set $\hat{I} := \{i \in I \colon \sum_{j\in J} a_{ij}\hat{x}_j - b_i \leq \sum_{j\in J_i} \alpha_{ij}|\hat{x}_j|\} \subseteq I$ to identify the constraints that can be made active by a feasible choice of Γ_i. That is, constraints in $\hat{I}$ have a nominal surplus less than or equal to the maximum possible value of the protection function (achieved when $\Gamma_i = |J_i|$). The set $\hat{I}$ depends on the given $\hat{\mathbf{x}}$ and may thus be more accurately denoted by $\hat{I}(\hat{\mathbf{x}})$, however, we use $\hat{I}$ for simplicity.

For $i \in \hat{I}$, we also need to determine the value of Γ_i that will render the corresponding constraint active. First, let j^i_k index the k-th largest element in the set $\{\alpha_{ij}|\hat{x}_j|\}_{j\in J_i}$, for all $k = 1, \ldots, |J_i|, i \in I$. Then for all $i \in \hat{I}$, let $\Gamma_i = \underline{\Gamma}_i$ satisfy

$$\sum_{j\in J} a_{ij}\hat{x}_j - b_i = \sum_{k=1}^{\lfloor\Gamma_i\rfloor} \alpha_{ij^i_k}|\hat{x}_{j^i_k}| + (\Gamma_i - \lfloor\Gamma_i\rfloor)\alpha_{ij^i_{\lceil\Gamma_i\rceil}}|\hat{x}_{j^i_{\lceil\Gamma_i\rceil}}|. \tag{3.25}$$

In other words, $\underline{\Gamma}_i \in [0, |J_i|]$ is a budget parameter such that the nominal surplus of constraint i equals the value of the protection function, thereby rendering constraint i active. For each $i \in \hat{I}$, $\underline{\Gamma}_i$ can be computed as the optimal value of the following linear optimization problem:

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & \sum_{j\in J_i} w_j \\ \text{subject to} \quad & \sum_{j\in J_i} \alpha_{ij}|\hat{x}_j|w_j = \sum_{j\in J} a_{ij}\hat{x}_j - b_i, \\ & 0 \leq w_j \leq 1, \quad \forall j \in J_i. \end{aligned} \tag{3.26}$$

Unless the data $\mathbf{a}_i, b_i, \boldsymbol{\alpha}_i, \hat{\mathbf{x}}$ meet some specific conditions, there exists a unique $\underline{\Gamma}_i$ that satisfies equation (3.25). To see this, first notice that the right-hand side of equation (3.25) is strictly increasing in Γ_i if $\alpha_{ij}|\hat{x}_j| > 0$ for all $j \in J_i$. If there exists any $j \in J_i$ such that $\alpha_{ij}|\hat{x}_j| = 0$, and we let v_i be the number of such indices, then the right-hand side of equation (3.25) is strictly increasing for $\Gamma_i \in [0, |J_i| - v_i]$ and constant for $\Gamma_i \in [|J_i| - v_i, |J_i|]$, at which point the expression equals $\sum_{j\in J_i} \alpha_{ij}\hat{x}_j$, the maximum possible value of the protection function. We can see that there are multiple Γ_i satisfying equation (3.25) if $v_i > 0$ and $\sum_{j\in J} a_{ij}\hat{x}_j - b_i = \sum_{j\in J_i} \alpha_{ij}\hat{x}_j$.

For simplicity, we assume that there exists a unique $\underline{\Gamma}_i$ that satisfies equation (3.25). Under this assumption, constraint i will be infeasible for $\Gamma_i > \underline{\Gamma}_i$, and will have positive surplus for $\Gamma_i < \underline{\Gamma}_i$. If we did not make this assumption, then equation (3.25) would be satisfied by $\Gamma_i \in [\underline{\Gamma}_i, \overline{\Gamma}_i]$, where

$$\overline{\Gamma}_i = \begin{cases} |J_i| & \text{if } \sum_{j\in J} a_{ij}\hat{x}_j - b_i = \sum_{j\in J_i} \alpha_{ij}|\hat{x}_j|, \\ \underline{\Gamma}_i & \text{otherwise,} \end{cases} \tag{3.27}$$

and correspondingly, constraint i would be infeasible for $\Gamma_i > \overline{\Gamma}_i$, and would have positive surplus for $\Gamma_i < \underline{\Gamma}_i$. The results in the remainder of this section would change by requiring $\Gamma_i \in [0, \overline{\Gamma}_i]$ wherever we currently have $\Gamma_i \in [0, \underline{\Gamma}_i]$, and $\Gamma_i \in [\underline{\Gamma}_i, \overline{\Gamma}_i]$ wherever we currently have $\Gamma_i = \underline{\Gamma}_i$.

3.2.1 Strong duality

Let $\lambda_{ij}, \mu_{ij}, \varphi_{ij}, \pi_i$ be the dual variables corresponding to constraints (3.24b)-(3.24e), respectively. The following formulation minimizes the deviation of $\mathbf{\Gamma}$ from given values $\hat{\mathbf{\Gamma}}$. We assume without loss of generality that $\hat{\Gamma}_i \in [0, |J_i|]$ for all $i \in I$, since any $\hat{\Gamma}_i$ outside the interval can be moved to the closest end point of the interval without changing the solution.

$$\underset{\substack{\mathbf{\Gamma},\mathbf{c},\mathbf{u},\mathbf{y},\mathbf{z},\\ \boldsymbol{\pi},\boldsymbol{\varphi},\boldsymbol{\lambda},\boldsymbol{\mu}}}{\text{minimize}} \quad \|\mathbf{\Gamma} - \hat{\mathbf{\Gamma}}\| \tag{3.28a}$$

$$\text{subject to} \quad \sum_{j\in J} c_j\hat{x}_j - \sum_{i\in I} b_i\pi_i = 0, \tag{3.28b}$$

$$\alpha_{ij}\hat{x}_j + u_{ij} \geq 0, \quad \forall j \in J_i, i \in I, \tag{3.28c}$$

$$-\alpha_{ij}\hat{x}_j + u_{ij} \geq 0, \quad \forall j \in J_i, i \in I, \tag{3.28d}$$

$$y_{ij} + z_i \geq u_{ij}, \quad \forall j \in J_i, i \in I, \tag{3.28e}$$

$$\sum_{j\in J} a_{ij}\hat{x}_j - \sum_{j\in J_i} y_{ij} - \Gamma_i z_i \geq b_i, \quad \forall i \in I, \tag{3.28f}$$

$$y_{ij}, z_i \geq 0, \quad \forall j \in J_i, i \in I, \tag{3.28g}$$

$$0 \leq \Gamma_i \leq |J_i|, \quad \forall i \in I, \tag{3.28h}$$

$$\sum_{i\in I} \pi_i = 1, \tag{3.28i}$$

$$\sum_{i\in I} a_{ij}\pi_i + \sum_{i\in I:\, j\in J_i} \alpha_{ij}(\lambda_{ij} - \mu_{ij}) = c_j, \quad \forall j \in J, \tag{3.28j}$$

$$\varphi_{ij} \leq \pi_i, \quad \forall j \in J_i, i \in I, \tag{3.28k}$$

$$\varphi_{ij} = \lambda_{ij} + \mu_{ij}, \quad \forall j \in J_i, i \in I, \tag{3.28l}$$

$$\sum_{j\in J_i} \varphi_{ij} \leq \Gamma_i\pi_i, \quad \forall i \in I, \tag{3.28m}$$

$$\pi_i, \varphi_{ij}, \lambda_{ij}, \mu_{ij} \geq 0, \quad \forall j \in J_i, i \in I. \tag{3.28n}$$

The construction of (3.28) parallels that of (3.3). Constraints (3.28b), (3.28c)-(3.28g), and (3.28j)-(3.28n) represent strong duality, primal feasibility and dual feasibility, respectively. We use the same normalization constraint (3.28i) to prevent the trivial solution $(\mathbf{c}, \boldsymbol{\pi}) = (\mathbf{0}, \mathbf{0})$ from being optimal.

First, we characterize the feasibility of (3.28).

Proposition 7. *Formulation (3.28) is feasible if and only if* $\sum_{j\in J} a_{ij}\hat{x}_j \geq b_i$ *for all* $i \in I$*, and* $\hat{I} \neq \varnothing$*.*

The first condition ($\sum_{j\in J} a_{ij}\hat{x}_j \geq b_i,\ \forall i \in I$) is required for primal feasibility of the robust problem (3.24) to be satisfied. The second condition ($\hat{I} \neq \varnothing$) is required for strong duality to be satisfied. For the rest of this section, we assume that (3.28) is feasible.

There are many similarities but some important differences between the inverse cardinality constrained robust problem (3.28) and the inverse interval uncertainty problem (3.3). Most importantly, while both formulations have bilinear constraints ((3.28f) and (3.28m) in (3.28) and (3.3h) in (3.3)), the structure of these constraints is different, and therefore different analysis and solution methods are required. First, we present a result that enables us to tractably deal with the bilinearity in (3.28f). For convenience, we define

$$\boldsymbol{\Theta} = \{\boldsymbol{\Gamma} : \Gamma_i \in [0, \underline{\Gamma}_i], i \in \hat{I}; \Gamma_i \in [0, |J_i|], i \in I \setminus \hat{I}\},$$

which will be used in several results below.

Lemma 3. *If* $(\boldsymbol{\Gamma}, \mathbf{u}, \mathbf{y}, \mathbf{z})$ *satisfies constraints (3.28c)-(3.28h), then* $\boldsymbol{\Gamma} \in \boldsymbol{\Theta}$*. Conversely, if* $\boldsymbol{\Gamma} \in \boldsymbol{\Theta}$*, then there exists* $(\mathbf{u}, \mathbf{y}, \mathbf{z})$ *such that* $(\boldsymbol{\Gamma}, \mathbf{u}, \mathbf{y}, \mathbf{z})$ *satisfies constraints (3.28c)-(3.28h).*

Proof: To prove the first statement, suppose to the contrary that $\Gamma_i \in (\underline{\Gamma}_i, |J_i|]$ for some $i \in \hat{I}$, i.e., $\boldsymbol{\Gamma} \notin \boldsymbol{\Theta}$ but Γ_i does satisfy constraint (3.28h). We will show that there is no $(\mathbf{u}, \mathbf{y}, \mathbf{z})$ that satisfies constraints (3.28c)-(3.28g). In particular, we will show that any $(\mathbf{u}, \mathbf{y}, \mathbf{z})$ that satisfies (3.28c)-(3.28e) and (3.28g) will never satisfy (3.28f).

Consider the following linear optimization problem:

$$\begin{aligned} \underset{\mathbf{u}_i, \mathbf{y}_i, z_i}{\text{minimize}} \quad & \sum_{j\in J_i} y_{ij} + \Gamma_i z_i \\ \text{subject to} \quad & y_{ij} + z_i \geq u_{ij}, \quad \forall j \in J_i, \\ & -u_{ij} \leq \alpha_{ij}\hat{x}_j \leq u_{ij}, \quad j \in J_i, \\ & y_{ij}, z_i \geq 0, \quad \forall j \in J_i. \end{aligned} \tag{3.29}$$

By Lemma 6 (see appendix), an optimal solution to this problem is

$$\begin{aligned} u_{ij}^* &= \alpha_{ij}|\hat{x}_j|, \quad \forall j \in J_i, \\ y_{ij}^* &= \max\{u_{ij}^* - z_i^*, 0\}, \quad \forall j \in J_i, \\ z_i^* &= \alpha_{ij^i_{\lceil \Gamma_i \rceil}}|\hat{x}_{j^i_{\lceil \Gamma_i \rceil}}|, \end{aligned} \tag{3.30}$$

which satisfies (3.28c)-(3.28e) and (3.28g) since the constraints of (3.29) are identical to (3.28c)-(3.28e) and (3.28g). However, $(\mathbf{u}_i^*, \mathbf{y}_i^*, z_i^*)$ does not satisfy (3.28f), as shown below. Because $(\mathbf{u}_i^*, \mathbf{y}_i^*, z_i^*)$ yields the smallest possible value of $\sum_{j\in J_i} y_{ij} + \Gamma_i z_i$, there cannot be any other feasible solution for (3.28c)-(3.28e) and (3.28g) which will satisfy (3.28f).

For completeness, we substitute $(\mathbf{u}_i^*, \mathbf{y}_i^*, z_i^*)$ into the left-hand-side of constraint (3.28f) to show that the constraint will not be satisfied:

$$\begin{aligned}
& \sum_{j\in J} a_{ij}\hat{x}_j - \left(\sum_{j\in J_i} \max\left\{ \alpha_{ij}|\hat{x}_j| - \alpha_{ij^i_{\lceil\Gamma_i\rceil}}|\hat{x}_{j^i_{\lceil\Gamma_i\rceil}}|, 0 \right\} + \Gamma_i \alpha_{ij^i_{\lceil\Gamma_i\rceil}}|\hat{x}_{j^i_{\lceil\Gamma_i\rceil}}| \right) \\
\Leftrightarrow\ & \sum_{j\in J} a_{ij}\hat{x}_j - \sum_{k=1}^{\lfloor\Gamma_i\rfloor} \left(\alpha_{ij^i_k}|\hat{x}_{j^i_k}| - \alpha_{ij^i_{\lceil\Gamma_i\rceil}}|\hat{x}_{j^i_{\lceil\Gamma_i\rceil}}| \right) - \Gamma_i \alpha_{ij^i_{\lceil\Gamma_i\rceil}}|\hat{x}_{j^i_{\lceil\Gamma_i\rceil}}| \\
\Leftrightarrow\ & \sum_{j\in J} a_{ij}\hat{x}_j - \sum_{k=1}^{\lfloor\Gamma_i\rfloor} \alpha_{ij^i_k}|\hat{x}_{j^i_k}| - (\Gamma_i - \lfloor\Gamma_i\rfloor)\alpha_{ij^i_{\lceil\Gamma_i\rceil}}|\hat{x}_{j^i_{\lceil\Gamma_i\rceil}}|.
\end{aligned} \tag{3.31}$$

Because $\Gamma_i = \underline{\Gamma}_i$ satisfies (3.25) we know that

$$\sum_{j\in J} a_{ij}\hat{x}_j - \sum_{k=1}^{\lfloor\underline{\Gamma}_i\rfloor} \alpha_{ij^i_k}|\hat{x}_{j^i_k}| - (\underline{\Gamma}_i - \lfloor\underline{\Gamma}_i\rfloor)\alpha_{ij^i_{\lceil\underline{\Gamma}_i\rceil}}|\hat{x}_{j^i_{\lceil\underline{\Gamma}_i\rceil}}| = b_i, \tag{3.32}$$

and since $\Gamma_i > \underline{\Gamma}_i$ by assumption,

$$\begin{aligned}
& \sum_{j\in J} a_{ij}\hat{x}_j - \sum_{k=1}^{\lfloor\Gamma_i\rfloor} \alpha_{ij^i_k}|\hat{x}_{j^i_k}| - (\Gamma_i - \lfloor\Gamma_i\rfloor)\alpha_{ij^i_{\lceil\Gamma_i\rceil}}|\hat{x}_{j^i_{\lceil\Gamma_i\rceil}}| \\
=\ & \sum_{j\in J} a_{ij}\hat{x}_j - \sum_{k=1}^{\lfloor\underline{\Gamma}_i\rfloor} \alpha_{ij^i_k}|\hat{x}_{j^i_k}| - \sum_{k=\lfloor\underline{\Gamma}_i\rfloor+1}^{\lfloor\Gamma_i\rfloor} \alpha_{ij^i_k}|\hat{x}_{j^i_k}| - (\Gamma_i - \lfloor\Gamma_i\rfloor)\alpha_{ij^i_{\lceil\Gamma_i\rceil}}|\hat{x}_{j^i_{\lceil\Gamma_i\rceil}}| \\
<\ & b_i,
\end{aligned}$$

that is, $(\mathbf{u}_i^*, \mathbf{y}_i^*, z_i^*)$ does not satisfy (3.28f).

To prove the second statement, suppose we are given $\mathbf{\Gamma} \in \mathbf{\Theta}$. Then it is straightforward to check that $(\mathbf{u}_i^*, \mathbf{y}_i^*, z_i^*), i \in I$ from (3.30) satisfies (3.28c)-(3.28h). □

Lemma 3 not only eliminates the bilinear constraint (3.28f) but also the auxiliary variables $\mathbf{u}, \mathbf{y}, \mathbf{z}$, by showing that the bounds $\mathbf{\Gamma} \in \mathbf{\Theta}$ are both necessary and sufficient for primal feasibility to be satisfied. Intuitively, primal feasibility requires for each constraint $i \in I$ that the nominal surplus be greater than or equal to the protection function, which is a function of Γ_i only. As suggested by the definition of $\hat{I}$ and formalized in Lemma 3, the protection functions $i \in \hat{I}$ meet this condition for $\Gamma_i \in [0, \underline{\Gamma}_i]$, and the protection functions $i \in I \setminus \hat{I}$ meet this condition for $\Gamma_i \in [0, |J_i|]$.

Lemma 3 establishes conditions on $\mathbf{\Gamma}$ to ensure the primal feasibility constraints are satisfied in the IO problem. Next, Lemma 4 addresses the strong duality and dual feasibility conditions. In particular, to satisfy optimality, $\mathbf{\Gamma}$ must be chosen such that $\hat{\mathbf{x}}$ lies on the boundary of the robust feasible region, which corresponds to a choice of $\mathbf{\Gamma}$ such that for at least one constraint, the protection function equals the nominal surplus. The next result formalizes this intuition (cf. Lemma 1, Lemma 2):

Lemma 4. *Every feasible solution for formulation* (3.28) *satisfies* $\mathbf{\Gamma} \in \mathbf{\Theta}$ *with* $\Gamma_{\hat{i}} = \underline{\Gamma}_{\hat{i}}$ *for a specific* $\hat{i} \in \hat{I}$. *Conversely, for every* $\mathbf{\Gamma} \in \mathbf{\Theta}$ *satisfying* $\Gamma_{\hat{i}} = \underline{\Gamma}_{\hat{i}}$ *for a specific* $\hat{i} \in \hat{I}$, *there exists* $(\mathbf{c}, \mathbf{u}, \mathbf{y}, \mathbf{z}, \boldsymbol{\pi}, \boldsymbol{\varphi}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ *such that* $(\mathbf{\Gamma}, \mathbf{c}, \mathbf{u}, \mathbf{y}, \mathbf{z}, \boldsymbol{\pi}, \boldsymbol{\varphi}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ *is feasible for formulation* (3.28).

Proof: To prove the first statement, we first note that by Lemma 3, constraints (3.28c)-(3.28h) are equivalent to $\mathbf{\Gamma} \in \mathbf{\Theta}$. So we only need to show that the constraints of (3.28) imply $\Gamma_{\hat{i}} = \underline{\Gamma}_{\hat{i}}$, for some $\hat{i} \in \hat{I}$. If we substitute (3.28j) into (3.28b), let $s_{ij} = \lambda_{ij} - \mu_{ij}$, and use reasoning similar to the proof of Lemma 5 (see appendix), we get

$$\sum_{i \in I} \pi_i \left(\sum_{j \in J} a_{ij} \hat{x}_j - b_i \right) \leq \sum_{i \in I} \sum_{j \in J_i} \alpha_{ij} |\hat{x}_j| \varphi_{ij}. \tag{3.33}$$

For a given $(\mathbf{\Gamma}, \boldsymbol{\pi})$, the constraints applicable to $\boldsymbol{\varphi}$ are (3.28k), (3.28m), and nonnegativity. As such, the maximum value of $\sum_{j \in J_i} \alpha_{ij} |\hat{x}_j| \varphi_{ij}$ over feasible $\boldsymbol{\varphi}_i$, for all $i \in I$, is the optimal value of the following optimization problem:

$$\begin{aligned} \underset{\boldsymbol{\varphi}_i}{\text{maximize}} \quad & \sum_{j \in J_i} \alpha_{ij} |\hat{x}_j| \varphi_{ij} \\ \text{subject to} \quad & 0 \leq \varphi_{ij} \leq \pi_i, \ \forall j \in J_i, \\ & \sum_{j \in J_i} \varphi_{ij} \leq \Gamma_i \pi_i. \end{aligned} \tag{3.34}$$

Formulation (3.34) is an instance of the continuous knapsack problem, so its optimal value is

$$\sum_{k=1}^{\lfloor \Gamma_i \rfloor} \alpha_{ij_k^i} |\hat{x}_{j_k^i}| \pi_i + (\Gamma_i - \lfloor \Gamma_i \rfloor) \alpha_{ij^i_{\lceil \Gamma_i \rceil}} |\hat{x}_{j^i_{\lceil \Gamma_i \rceil}}| \pi_i,$$

and we can conclude that

$$\sum_{i \in I} \pi_i \left(\sum_{j \in J} a_{ij} \hat{x}_j - b_i \right) \leq \sum_{i \in I} \pi_i \left(\sum_{k=1}^{\lfloor \Gamma_i \rfloor} \alpha_{ij_k^i} |\hat{x}_{j_k^i}| + (\Gamma_i - \lfloor \Gamma_i \rfloor) \alpha_{ij^i_{\lceil \Gamma_i \rceil}} |\hat{x}_{j^i_{\lceil \Gamma_i \rceil}}| \right). \tag{3.35}$$

Now, assume to the contrary that $\Gamma_i < \underline{\Gamma}_i$ for all $i \in \hat{I}$. Using this assumption and the fact that $\Gamma_i = \underline{\Gamma}_i$ satisfies (3.25), we can deduce a contradiction with (3.35).

To prove the second statement, let $(\mathbf{c}, \mathbf{u}, \mathbf{y}, \mathbf{z}, \boldsymbol{\pi}, \boldsymbol{\varphi}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ be defined as in the proof of Proposition 7. With the assumed $\mathbf{\Gamma}$, this solution is feasible for formulation (3.28). □

Lemma 4 allows us to easily characterize an optimal solution to (3.28) and devise an efficient solution method (cf. Theorem 1, Theorem 3):

Theorem 5. *Let*

$$f_i = \underline{\Gamma}_i - \hat{\Gamma}_i, \quad \forall i \in \hat{I}, \tag{3.36}$$

$$g_i = \min\left\{\underline{\Gamma}_i - \hat{\Gamma}_i, 0\right\}, \quad \forall i \in \hat{I}, \tag{3.37}$$

$$c_j^i = a_{ij} - \operatorname{sgn}(\hat{x}_j)\bar{\alpha}_{ij}, \quad \forall j \in J, i \in \hat{I}, \tag{3.38}$$

$$i^* \in \underset{i \in \hat{I}}{\arg\min}\{\|\mathbf{g} + (f_i - g_i)\mathbf{e}_i\|\}, \tag{3.39}$$

where

$$\bar{\alpha}_{ij} = \begin{cases} \alpha_{ij} & \text{if } j = j_k^i, k = 1, \dots, \lfloor \Gamma_i \rfloor, \\ \alpha_{ij}(\Gamma_i - \lfloor \Gamma_i \rfloor) & \text{if } j = j^i_{\lfloor \Gamma_i \rfloor + 1}, \\ 0 & \text{otherwise}, \end{cases} \tag{3.40}$$

for all $i \in \hat{I}$. Then the optimal value of formulation (3.28) is $\|\mathbf{g} + (f_{i^*} - g_{i^*})\mathbf{e}_{i^*}\|$, and there exists an optimal solution of (3.28) with

$$\Gamma_i = \begin{cases} \underline{\Gamma}_i & \text{if } i = i^*, \\ \min\{\hat{\Gamma}_i, \underline{\Gamma}_i\} & \text{if } i \neq i^*, i \in \hat{I}, \\ \hat{\Gamma}_i & \text{if } i \in I \setminus \hat{I}, \end{cases} \tag{3.41}$$

$$\mathbf{c} = \mathbf{c}^{i^*}. \tag{3.42}$$

Remark 4. Theorem 5 shows that an optimal solution to the nonconvex inverse problem (3.28) can be found by solving $|\hat{I}|$ linear optimization problems of the form (3.26).

Proof: By Lemma 4, solving formulation (3.28) is equivalent to solving the following optimization problem for all $\hat{\imath} \in \hat{I}$, and taking the minimum over all $|\hat{I}|$ optimal values:

$$\begin{aligned} \underset{\mathbf{\Gamma}}{\text{minimize}} \quad & \|\mathbf{\Gamma} - \hat{\mathbf{\Gamma}}\| \\ \text{subject to} \quad & \Gamma_{\hat{\imath}} = \underline{\Gamma}_{\hat{\imath}}, \\ & 0 \leq \Gamma_i \leq \underline{\Gamma}_i, \quad \forall i \in \hat{I}, \\ & 0 \leq \Gamma_i \leq |J_i|, \quad \forall i \in I \setminus \hat{I}. \end{aligned} \tag{3.43}$$

Suppose we fix some $\hat{\imath} \in \hat{I}$. For all $i \in I$, the variable Γ_i is included only in a term $(\Gamma_i - \hat{\Gamma}_i)$ in the objective function. For $i \in I \setminus \hat{I}$, because $\hat{\Gamma}_i \in [0, |J_i|]$ by assumption, it is clear that the optimal solution is $\Gamma_i = \hat{\Gamma}_i$ and the corresponding term in the objective function equals 0. For $i \neq \hat{\imath}, i \in \hat{I}$, we require $\Gamma_i \in [0, \underline{\Gamma}_i]$, but we will have either $\hat{\Gamma}_i \in [0, \underline{\Gamma}_i]$ or $\hat{\Gamma} \in [\underline{\Gamma}_i, |J_i|]$. Hence the optimal solution is $\Gamma_i = \min\{\hat{\Gamma}_i, \underline{\Gamma}_i\}$, and the corresponding term in the objective function equals g_i. For $i = \hat{\imath}$, we require $\Gamma_i = \hat{\Gamma}_i$ and the corresponding term in the objective function equals f_i. It follows that the optimal value of the $\hat{i}$-th formulation (3.43) is $\|\mathbf{g} + (f_{\hat{i}} - g_{\hat{i}})\mathbf{e}_{\hat{i}}\|$. Thus the optimal value of formulation (3.28) is $\min_{\hat{i}\in\hat{I}} \left\{\|\mathbf{g} + (f_{\hat{i}} - g_{\hat{i}})\mathbf{e}_{\hat{i}}\|\right\}$, and the optimal cost vector is $\mathbf{c} = \mathbf{c}^{i^*}$, where the structure of $\mathbf{c}^i$ is derived in the proof of Proposition 7. □

The interpretation of Theorem 5 is conceptually very similar to the interpretation of Theorem 3. The two results are subtly different because f_i and g_i here correspond to the values of the i-th component of the vector inside the norm in the objective function, whereas in the former result they correspond to the values of the i-th term in the objective function, thus why the expressions for i^* differ. Due to our assumption that there is a unique value of Γ_i that renders a given constraint active, the two results also differ in that f_i and g_i can be determined analytically rather than by auxiliary optimization problems.

3.2.2 Duality gap minimization

Next, we formulate an IO model that minimizes the duality gap while enforcing primal and dual feasibility, and external constraints $\mathbf{\Gamma} \in \mathbf{\Omega}$:

$$\underset{\substack{\mathbf{\Gamma},\mathbf{c},\mathbf{u},\mathbf{y},\mathbf{z},\\ \boldsymbol{\pi},\boldsymbol{\varphi},\boldsymbol{\lambda},\boldsymbol{\mu}}}{\text{minimize}} \quad \sum_{j\in J} c_j \hat{x}_j - \sum_{i\in I} b_i \pi_i \tag{3.44a}$$

$$\text{subject to} \quad \mathbf{\Gamma} \in \mathbf{\Omega}, \tag{3.44b}$$

$$(3.28c) - (3.28n). \tag{3.44c}$$

As in the interval uncertainty case, constraints on the uncertainty set parameters may prevent strong duality from being achieved exactly, thus requiring an inverse model of the form (3.44). However, in the case of cardinality constrained uncertainty there is a second possible motivation, which is that $\hat{I}$ may be empty. By Proposition 7, if $\hat{I} = \varnothing$ then formulation (3.28) will not be feasible. In particular, it will be possible to satisfy both primal and dual feasibility, but not strong duality.

Analogous to formulation (3.14) in the interval uncertainty case, the feasibility of $\hat{\mathbf{x}}$ for the nominal problem, along with extra conditions on $\mathbf{\Gamma}$, are necessary and sufficient conditions for feasibility of the IO model (3.44). We omit the proof of this result, which is straightforward and similar to the proof of Proposition 7.

Proposition 8. *Formulation* (3.44) *is feasible if and only if* $\sum_{j\in J} a_{ij}\hat{x}_j \geq b_i$ *for all* $i \in I$, *and* $\mathbf{\Theta} \cap \mathbf{\Omega} \neq \varnothing$.

Next, we reformulate (3.44), which is bilinear, into an equivalent mixed-integer optimization problem, which is linear whenever the external constraints are linear:

$$\begin{aligned}
\underset{\boldsymbol{\Gamma},\boldsymbol{\pi},\boldsymbol{\varphi}}{\text{minimize}} \quad & \sum_{i\in I}\left(\sum_{j\in J} a_{ij}\hat{x}_j - b_i\right)\pi_i - \sum_{i\in I}\sum_{j\in J_i}\alpha_{ij}|\hat{x}_j|\varphi_{ij} && (3.45a)\\
\text{subject to} \quad & \boldsymbol{\Gamma}\in\boldsymbol{\Omega}, && (3.45b)\\
& \boldsymbol{\Gamma}\in\boldsymbol{\Theta}, && (3.45c)\\
& \sum_{i\in I}\pi_i = 1, && (3.45d)\\
& 0\le\varphi_{ij}\le\pi_i, \quad \forall j\in J_i, i\in I, && (3.45e)\\
& \sum_{j\in J_i}\varphi_{ij}\le\Gamma_i + M(1-\pi_i), \quad \forall i\in I, && (3.45f)\\
& \sum_{j\in J_i}\varphi_{ij}\le M\pi_i, \quad \forall i\in I, && (3.45g)\\
& \pi_i\in\{0,1\}, \quad \forall i\in I. && (3.45h)
\end{aligned}$$

Formulation (3.45) can be interpreted as follows. Constraints (3.45b) are retained from formulation (3.44), and (3.45c) replaces (3.28c)-(3.28h) as per Lemma 3. As a result of the normalization constraint, the duality gap is equal to the surplus of a single constraint $i\in I$; the choice of this constraint is encoded in the binary vector $\boldsymbol{\pi}$, and the optimal solution for $\boldsymbol{\varphi}$ will make $\sum_{i\in I}\sum_{j\in J_i}\alpha_{ij}|\hat{x}_j|\varphi_{ij}$ equal to the protection function of constraint i. The optimal choice i^* will correspond to the constraint with the minimum surplus. Although formulation (3.45) does not include the variable $\mathbf{c}$, the cost vector $\mathbf{c}$ can be determined post-optimization using

$$c_j = \sum_{i\in I} a_{ij}\pi_i - \sum_{i\in I:\, j\in J_i}\alpha_{ij}\,\operatorname{sgn}(\hat{x}_j)\varphi_{ij}, \quad \forall j\in J, \tag{3.46}$$

which is derived from equation (3.28j).

We now formally characterize and prove the correspondence between formulations (3.44) and (3.45) (cf. Theorem 4). When we reformulated the corresponding problem in the case of interval uncertainty, we substituted the bilinear dual feasibility constraint (3.3h) into the objective function to render the dual variables disjoint from the remaining variables, from which we could conclude that there exists an optimal solution with binary $\boldsymbol{\pi}$. In the case of cardinality constrained uncertainty, the same reasoning cannot be applied because there is no possible substitution that will render the dual variables disjoint from the remaining variables. Nevertheless, we will be able to use an alternative line of reasoning to show that there exists an optimal solution to (3.44) with binary $\boldsymbol{\pi}$.

Theorem 6. *Let $M\ge|J|$. Formulations (3.44) and (3.45) have the same optimal objective value, and a solution $\boldsymbol{\Gamma}$ is optimal for formulation (3.44) if and only if it is optimal for formulation (3.45).*

Proof: To prove the first statement, we will eliminate variables from formulation (3.44) until we have an equivalent formulation in the variables $\mathbf{\Gamma}, \boldsymbol{\pi}$ only, at which point we can conclude by inspection that there exists an optimal solution with integer $\boldsymbol{\pi}$. This conclusion is then used to derive an equivalent mixed-integer linear optimization model in the variables $\mathbf{\Gamma}, \boldsymbol{\pi}, \boldsymbol{\varphi}$.

We begin by noting that by Lemma 3, constraints (3.28c)-(3.28h) (and thus variables $\mathbf{u}, \mathbf{y}, \mathbf{z}$) in formulation (3.44) can be replaced by $\mathbf{\Gamma} \in \mathbf{\Theta}$. We now omit the details of several steps that are conceptually similar to steps in the proof of Theorem 4: we eliminate $\mathbf{c}$ by substituting constraint (3.28j) into the objective function of formulation (3.44), we let $s_{ij} = \lambda_{ij} - \mu_{ij}$ for all $i \in I, j \in J_i$, we use Lemma 5 (see appendix) to replace constraints on $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ with constraints on $\mathbf{s}$, and then we eliminate $\mathbf{s}$ by identifying its optimal solution by inspection, giving us the following optimization problem equivalent to (3.44):

$$\begin{aligned}
\underset{\mathbf{\Gamma},\boldsymbol{\pi},\boldsymbol{\varphi}}{\text{minimize}} \quad & \sum_{i\in I}\left(\sum_{j\in J} a_{ij}\hat{x}_j - b_i\right)\pi_i - \sum_{i\in I}\sum_{j\in J_i}\alpha_{ij}|\hat{x}_j|\varphi_{ij} && (3.47a)\\
\text{subject to} \quad & \mathbf{\Gamma} \in \mathbf{\Omega}, && (3.47b)\\
& \mathbf{\Gamma} \in \mathbf{\Theta}, && (3.47c)\\
& \sum_{i\in I}\pi_i = 1, && (3.47d)\\
& 0 \le \varphi_{ij} \le \pi_i, \quad \forall j \in J_i, i \in I, && (3.47e)\\
& \sum_{j\in J_i}\varphi_{ij} \le \Gamma_i\pi_i, \quad \forall i \in I, && (3.47f)\\
& \pi_i \ge 0, \quad \forall i \in I. && (3.47g)
\end{aligned}$$

By inspection we can describe the optimal $\boldsymbol{\varphi}$ for a given value of $(\mathbf{\Gamma}, \boldsymbol{\pi})$. In the objective function, $\boldsymbol{\varphi}$ only appears in the term $-\sum_{i\in I}\sum_{j\in J_i}\alpha_{ij}|\hat{x}_j|\varphi_{ij}$, and the only constraints applicable to $\boldsymbol{\varphi}$ are (3.47e)-(3.47f). This an instance of the continuous knapsack problem, equivalent to (3.34) in the proof of Lemma 4, so an optimal $\boldsymbol{\varphi}$ has the form:

$$\varphi_{ij_k^i} = \begin{cases} \pi_i & \text{if } k = 1,\ldots,\lceil\Gamma_i\rceil - 1,\\ (\Gamma_i - \lceil\Gamma_i\rceil + 1)\pi_i & \text{if } k = \lceil\Gamma_i\rceil,\\ 0 & \text{if } k = \lceil\Gamma_i\rceil + 1,\ldots,|J_i|, \quad \forall j_k^i \in J_i, i \in I. \end{cases} \tag{3.48}$$

Substituting (3.48) into formulation (3.47) we obtain an equivalent optimization problem:

$$\min_{\mathbf{\Gamma},\boldsymbol{\pi}}\left\{\sum_{i\in I}\pi_i l_i(\mathbf{\Gamma})\colon (3.47b) - (3.47d), (3.47g)\right\}, \tag{3.49}$$

where

$$l_i(\mathbf{\Gamma}) = \sum_{j\in J} a_{ij}\hat{x}_j - b_i - \sum_{k=1}^{\lceil\Gamma_i\rceil-1} \alpha_{ij^i_k}|\hat{x}_{j^i_k}| - (\Gamma_i - \lceil\Gamma_i\rceil + 1)\alpha_{ij^i_{\lceil\Gamma_i\rceil}}|\hat{x}_{j^i_{\lceil\Gamma_i\rceil}}|.$$

By inspection, we can see that for a given value of $\mathbf{\Gamma}$, an optimal $\boldsymbol{\pi}$ for formulation (3.49) equals $\mathbf{e}_{i^*}$ where $i^* \in \arg\min_{i\in I} l_i(\mathbf{\Gamma})$. Note that we obtained (3.49) from (3.47) by setting $\boldsymbol{\varphi}$ optimally given any feasible $(\mathbf{\Gamma}, \boldsymbol{\pi})$. Hence any $\boldsymbol{\pi}$ that is optimal for (3.49) must also be optimal for (3.47). This means that, without changing the optimal value, we can restrict $\boldsymbol{\pi}$ to be binary in problem (3.47), and the resulting formulation is identical to formulation (3.45) except that (3.47f) is present in place of (3.45f)-(3.45g). Finally, if we substitute $M = |J|$ into constraints (3.45f)-(3.45g), it is straightforward to show that they are equivalent to (3.47f).

To prove the second statement, we note that in the process of deriving formulation (3.45) from (3.44), all steps are equivalences as far as $\mathbf{\Gamma}$ is concerned. Hence $\mathbf{\Gamma}$ is optimal for (3.45) if and only if it is optimal for (3.44). □

In this section, we have derived tractable solution approaches for two nonconvex IO models, (3.28) and (3.44), which recover the parameter $\mathbf{\Gamma}$ for a cardinality constrained uncertainty set with given $\boldsymbol{\alpha}$. The choice of which model to use depends not only on whether the application domain motivates constraints of the form $\mathbf{\Gamma} \in \mathbf{\Omega}$, but also on whether there exists any constraint in the forward model for which the protection function can equal the nominal surplus, and thereby render the constraint active. We note that formulation (3.28) with the addition of the constraint $\mathbf{\Gamma} \in \mathbf{\Omega}$ can also be shown to be efficiently solvable.

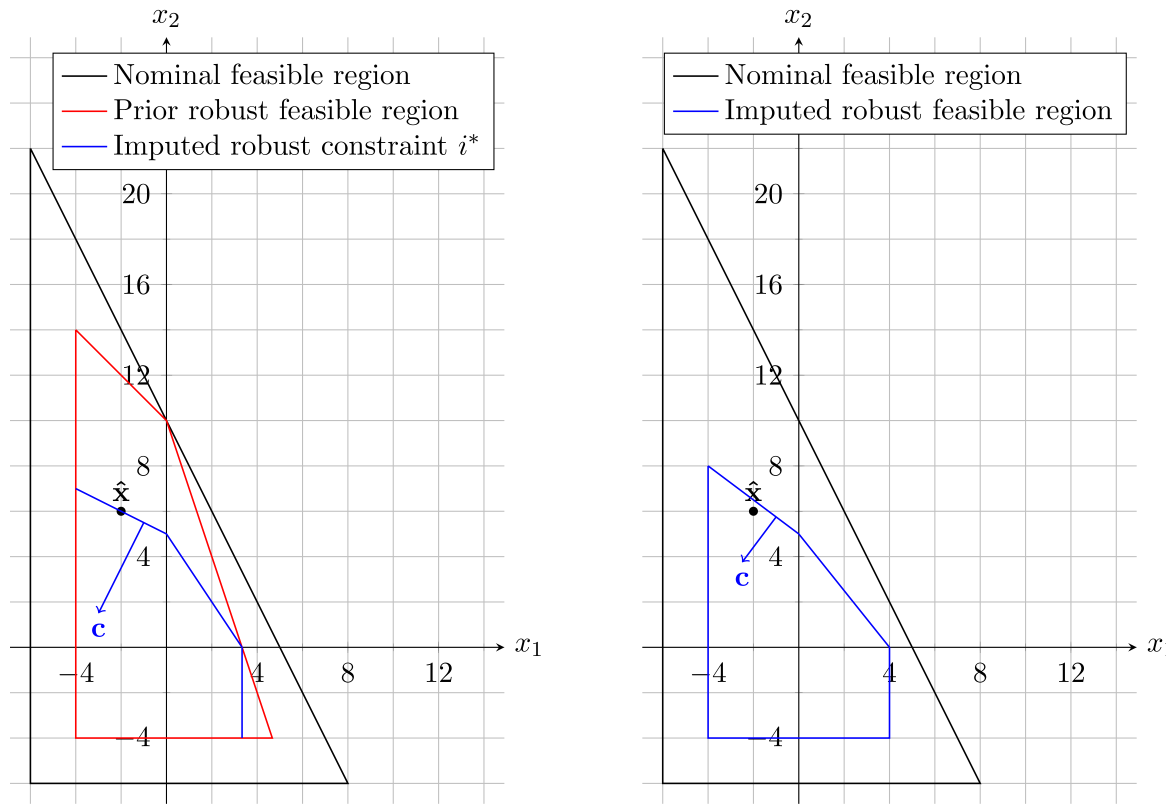
3.3 Numerical examples

In this section, we give numerical examples to illustrate the geometric characteristics of the solutions for the interval uncertainty formulations (3.3) and (3.14), and the cardinality constrained uncertainty formulations (3.28) and (3.44). These examples will demonstrate how the optimal inverse solution is found and how it relates to the geometry of the robust feasible region induced by the uncertainty set parameters. We use the L_1 norm for the objective functions of the strong duality formulations (3.3) and (3.28).

For all examples, let $\hat{\mathbf{x}} = (-2, 6)$ be the observed solution, and let the nominal problem be

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & c_1x_1 + c_2x_2 \\ \text{subject to} \quad & x_1 \geq -6, \\ & x_2 \geq -6, \\ & -2x_1 - x_2 \geq -10. \end{aligned}$$

Let the constraints and variables be indexed by $I = \{1, 2, 3\}$ and $J = \{1, 2\}$ respectively, and



(a) Strong duality. Since the observed solution is an interior point of the prior robust feasible region, the optimal solution of the IO model adjusts a single constraint such that it is rendered active, and sets the cost vector perpendicular to this constraint.

(b) Duality gap minimization. The constraints on the unspecified parameters prevent the feasible region from having the observed solution on its boundary, thus the cost vector is set perpendicular to the constraint with the minimum surplus.

Figure 3.1: Numerical examples of the interval uncertainty IO models. Both examples share the same observed solution and nominal feasible region.

let the coefficients subject to uncertainty be defined by $J_1 = \{1\}, J_2 = \{2\}, J_3 = \{1, 2\}$.

3.3.1 Interval uncertainty

For the strong duality formulation (3.3), let the robust optimization problem have the given prior parameters $\hat{\alpha}_{11} = 0.5, \hat{\alpha}_{22} = 0.5, \hat{\boldsymbol{\alpha}}_3 = (1, 0)$. For simplicity, we use the weight vector $\boldsymbol{\xi} = \mathbf{e}$ in the objective function. The nominal and robust (assuming $\hat{\boldsymbol{\alpha}}$) feasible regions are shown in Figure 3.1a; in particular, the robust counterpart of the third constraint is equivalent to $-2x_1 - x_2 - |x_1| \geq -10$, thus the realization of the constraint depends on the sign of x_1. We note that the IO problem (3.3) is feasible, since $\hat{\mathbf{x}}$ is feasible for the nominal problem (see Proposition 5). To determine the optimal solution, we apply Theorem 3. First, we have $g_i = 0$ for all $i \in I$, meaning that $\hat{\mathbf{x}}$ is feasible for the robust problem with the prior $\hat{\boldsymbol{\alpha}}$. The choice of the constraint to perturb will then be determined by $i^* \in \arg\min_{i \in I} f_i$, and evaluating all f_i (by solving the corresponding linear optimization problem (3.6)) we find $i^* = 3, f_3 = 1, \boldsymbol{\alpha}_3^f = (1, 1)$. Letting

$\boldsymbol{\alpha}_3 = \boldsymbol{\alpha}_3^f$, the robust counterpart of the third constraint becomes $-2x_1 - x_2 - |x_1| - |x_2| \geq -10$, and $\hat{\mathbf{x}}$ is on the boundary of this constraint in the second quadrant. Accordingly, the imputed cost vector $\mathbf{c} = \mathbf{c}^3 = (-1, -2)$ (see equation (3.8)) is perpendicular to the third constraint in the second quadrant.

For the duality gap formulation (3.14), we will require that the imputed $\boldsymbol{\alpha}$ be in the set

$$\boldsymbol{\Omega} := \left\{ \boldsymbol{\alpha} \colon \alpha_{ij} \geq 0.5, \forall i \in I, j \in J_i; \sum_{i \in I} \sum_{j \in J_i} \alpha_{ij} \leq 2.5 \right\}.$$

By Proposition 6, formulation (3.14) is feasible if and only if $\hat{\mathbf{x}}$ is robust feasible with respect to some $\boldsymbol{\alpha} \in \boldsymbol{\Omega}$; in this example, $\hat{\boldsymbol{\alpha}}$ meets this requirement, so the IO problem is feasible. To determine the optimal solution, we solve the equivalent mixed integer linear program (3.15). The nominal and imputed robust feasible regions are shown in Figure 3.1b. For the first and second constraints, the imputed $\boldsymbol{\alpha}_i$ equals $\hat{\boldsymbol{\alpha}}_i$ from the strong duality example. For the third constraint, we impute $\boldsymbol{\alpha}_3 = (0.5, 1)$, which is a "smaller" uncertainty set than the one imputed in the strong duality example, and hence the constraint is relaxed, and has a positive rather than zero surplus. The minimum duality gap is obtained by the cost vector $\mathbf{c} = (-1.5, -2)$ (see equation (3.16)), which is perpendicular to the third constraint in the second quadrant, and hence corresponds to optimal $\boldsymbol{\pi} = \mathbf{e}_3$.

3.3.2 Cardinality constrained uncertainty

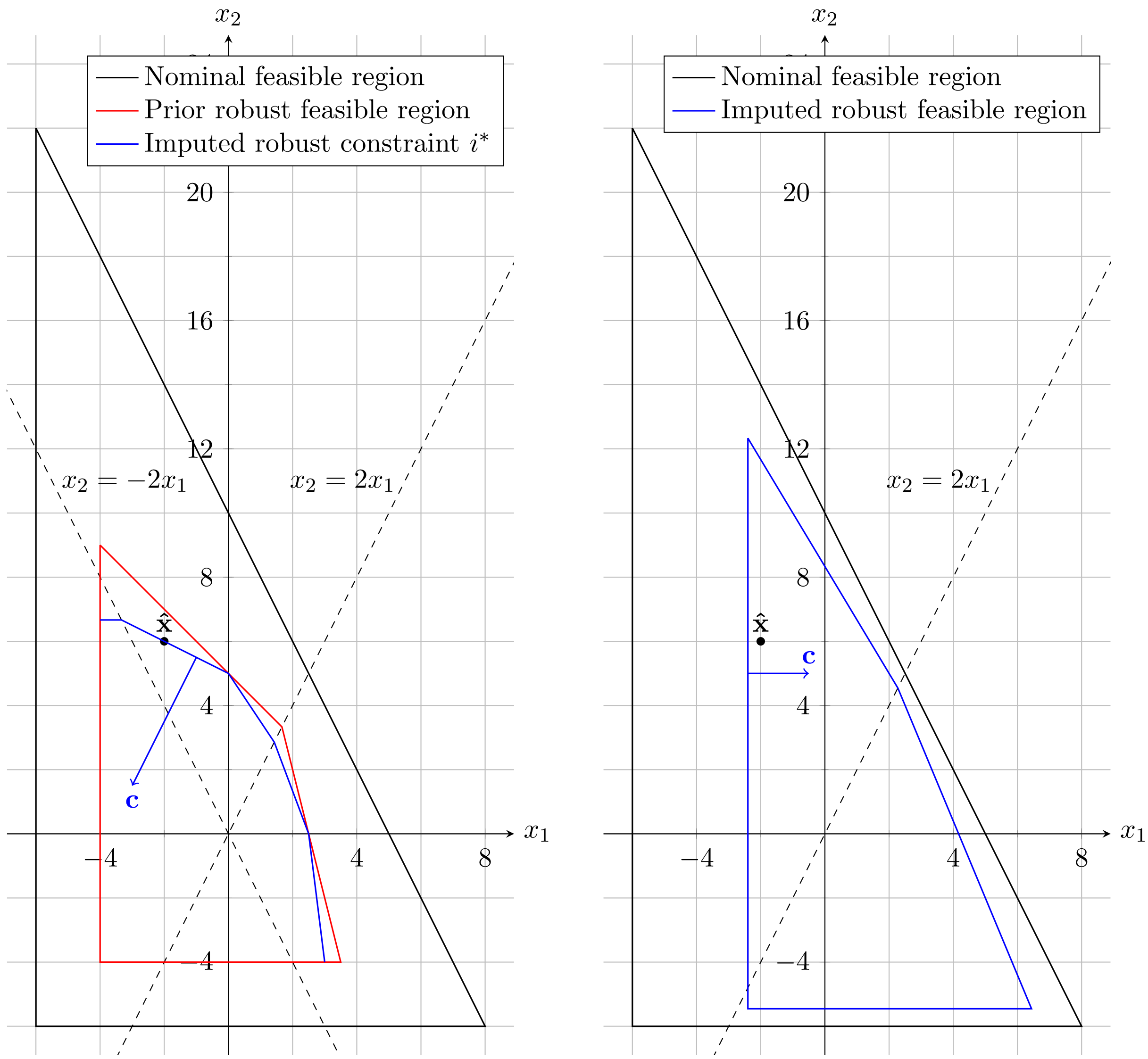
For the strong duality formulation (3.28), we assume fixed parameters $\alpha_{11} = 2.5, \alpha_{22} = 0.5, \boldsymbol{\alpha}_3 = (2, 1)$ and a prior $\hat{\boldsymbol{\Gamma}} = (0.2, 1, 1)$. The nominal and robust feasible regions are shown in Figure 3.2a. In particular, the robust counterpart of the third constraint is equivalent to

$$-2x_1 - x_2 - \max\{2|x_1|, |x_2|\} \geq -10.$$

Thus, the realization of this constraint depends not only the sign of $\mathbf{x}$ (as in the interval uncertainty example), but also the position of $\mathbf{x}$ relative to the lines $x_2 = 2x_1$ and $x_2 = -2x_1$. Accordingly, Figure 3.2a depicts these two lines; wherever they intersect the boundary of the constraint, the constraint changes slope.

First, we verify that constraints 1 and 3 (but not 2) have nominal surplus less than the maximum value of the corresponding protection function, and therefore $\hat{I} = \{1, 3\}$. Solving equation (3.26) results in $\underline{\Gamma}_1 = 0.8, \underline{\Gamma}_3 = 1.5$. Next, we determine that $g_i = 0$ for all $i \in \hat{I}$, meaning that $\hat{\mathbf{x}}$ is feasible with respect to the prior $\hat{\boldsymbol{\Gamma}}$ and the problem reduces to finding $i^* \in \arg\min_{i \in \hat{I}} |f_i|$. Finally, we find the unique solution $i^* = 3$ and $f_3 = 0.5$. Letting $\Gamma_3 = \underline{\Gamma}_3 = 1.5$, the robust counterpart of the third constraint becomes

$$-2x_1 - x_2 - \max\{2|x_1| + 0.5|x_2|, |x_1| + |x_2|\} \geq -10,$$



(a) Strong duality. With a slight adjustment to a single constraint, the observed solution is on the boundary of the feasible region and therefore optimal.

(b) Duality gap minimization. The constraint with the minimum surplus incidentally differs from the constraint which is set active in the strong duality example.

Figure 3.2: Numerical examples of the cardinality constrained uncertainty IO models.

which is piecewise linear with five pieces (four breakpoints defined by the two coordinate axes and the two equations $x_2 = 2x_1$ and $x_2 = -2x_1$). The observed $\hat{\mathbf{x}}$ is on the part of this constraint in the region defined by $x_1 < 0, x_2 \geq 0, x_2 \geq -2x_1$, and the imputed cost vector $\mathbf{c} = \mathbf{c}^3 = (-1, -2)$ is perpendicular to the constraint at $\hat{\mathbf{x}}$.

For the duality gap formulation (3.44), we will require that the imputed $\mathbf{\Gamma}$ be in the set

$$\mathbf{\Omega} := \left\{ \mathbf{\Gamma}\colon \Gamma_i \geq 0.2, \forall i \in I; \sum_{i \in I} \Gamma_i \leq 1 \right\}.$$

Similar to the interval uncertainty duality gap formulation, we can verify the feasibility of the IO problem by finding uncertainty set parameters that meet the condition in Proposition 8 (for brevity, we omit this step). To determine the optimal solution of (3.44), we solve the equivalent mixed integer linear problem (3.45). The nominal and imputed robust feasible regions are shown in Figure 3.2b. The optimal inverse solution has $\mathbf{\Gamma} = (0.6, 0.2, 0.2)$, $\boldsymbol{\pi} = \mathbf{e}_1$ and $\mathbf{c} = (2.5, 0)$ (see equation (3.46)). In other words, the minimum possible surplus for any of the three constraints is obtained by maximizing the degree of uncertainty associated with the first constraint. The duality gap equals the surplus of the first constraint, and the cost vector is perpendicular to the first constraint.

Lastly we note that in both the interval and cardinality-constrained strong duality examples, the problem data were chosen such that all $g_i = 0$, i.e., $\hat{\mathbf{x}}$ was feasible with respect to the prior uncertainty set parameters. If we had any $g_i \neq 0$, then in Figures 3.1 and 3.2, the observed solution $\hat{\mathbf{x}}$ would be within the boundaries of the nominal feasible region but outside the boundaries of the prior robust feasible region. For every constraint i which is infeasible with respect to $\hat{\boldsymbol{\alpha}}_i$ (or $\hat{\Gamma}_i$) and $\hat{\mathbf{x}}$, the minimal perturbation to $\hat{\boldsymbol{\alpha}}_i$ (or $\hat{\Gamma}_i$) would have $\hat{\mathbf{x}}$ sit on the boundary, and hence we could set $i = i^*$.

Chapter 4

Conclusion

In this thesis, we have considered three linear optimization models as forward problems. For each forward problem, we have proposed two different IO approaches that recover unspecified constraint parameters. The first approach minimally perturbs prior estimates of the parameters to be imputed such that the observed solution is exactly optimal, while the second identifies within some predefined set the parameters that minimize the duality gap.

Our three forward models differ in their structure and the parameters to be recovered: the first is a general linear programming problem in which all left-hand-side constraint coefficients are to be recovered, and the latter two are robust linear optimization problems in which constraint coefficients corresponding to parameters of the uncertainty set are to be recovered. In spite of these differences, the key steps in the model construction and solution method for each approach are common to all forward models. For the approach that makes the observed solution exactly optimal, the IO model's constraints are strong duality, primal and dual feasibility, and a normalization constraint that effectively makes the duality gap equal to a convex combination of constraint surpluses of the forward problem. This IO model is nonconvex due to at least one bilinear constraint. The solution method depends on the insight that the constraints are equivalent to requiring the observed solution to be on the boundary of the forward problem's feasible region. Thus the IO model can be solved by computing the minimum objective value associated with making each constraint of the forward model active, and choosing the constraint that yields the minimum objective value. The cost vector is then set perpendicular to this constraint.

For the approach that minimizes the duality gap, the IO model's constraints are primal and dual feasibility, and external constraints on the unspecified parameters. This model is also nonconvex, but can be shown to be equivalent to solving a finite number of optimization problems which are linear whenever the external constraints are linear. The equivalent tractable problems amount to checking the minimum duality gap that can be induced by setting the cost vector perpendicular to each constraint, and then choosing the constraint that yields the minimum objective value.

There are several directions for future work. First, it may be possible to apply our method to the robust linear optimization problem with ellipsoidal uncertainty. As in the cardinality constrained case, the ellipsoidal uncertainty set has a budget parameter for each constraint controlling the extent to which the coefficients in that constraint are allowed to vary within their intervals. However, the robust problem with ellipsoidal uncertainty is a second order conic program, thus an IO model would have to be formulated using the optimality conditions of second order conic programs (Alizadeh and Goldfarb, 2003). Although Iyengar and Kang (2005) have solved the inverse conic programming problem, we do not believe any author has solved this problem when the parameters to be recovered are in the constraints.

Second, it may be possible to generalize our methodology in two ways that classical IO models were also generalized. Our second IO approach minimizes the duality gap, and Chan et al. (2017) showed that for an inverse linear optimization model that recovers the cost vector, the duality gap is a special case of a more general error function. It may be possible that for the inverse linear optimization model that recovers constraint coefficients in addition to a cost vector, the duality gap can also be generalized. Another way that our IO models may be generalizable is to the case of multiple observed solutions, as done by Aswani et al. (2015), Bertsimas et al. (2015) and Keshavarz et al. (2011).

Finally, it may be beneficial to develop IO models that treat the imputation of the cost vector as importantly as the imputation of the constraint parameters. Our IO approaches choose constraint parameters such that the observed solution is optimal (or minimizes the duality gap) with respect to *some* nonzero cost vector, but a shortcoming of this method is that the implied cost vector may not be considered a reasonable fit for the application domain. More nuanced IO approaches would allow for an objective function that minimizes the perturbation of a prior cost vector in addition to the perturbation of prior constraint coefficients; and allow for external constraints on the cost vector in addition to those on the constraint coefficients. This thesis nevertheless serves as the first comprehensive attempt at solving the problems that we have considered, and we believe may serve as a preliminary step toward more general IO approaches.

Bibliography

Ravindra K Ahuja and James B Orlin. Inverse optimization. *Operations Research*, 49(5): 771–783, 2001.

Farid Alizadeh and Donald Goldfarb. Second-order cone programming. *Mathematical Programming*, 95(1):3–51, 2003.

Anil Aswani, Zuo-Jun Max Shen, and Auyon Siddiq. Inverse optimization with noisy data. *arXiv preprint arXiv:1507.03266*, 2015.

Michael O Ball and Maurice Queyranne. Toward robust revenue management: Competitive analysis of online booking. *Operations Research*, 57(4):950–963, 2009.

Aharon Ben-Tal and Arkadi Nemirovski. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming*, 88(3):411–424, 2000.

Aharon Ben-Tal, Arkadi Nemirovski, and Cees Roos. Robust solutions of uncertain quadratic and conic-quadratic problems. *SIAM Journal on Optimization*, 13(2):535–560, 2002.

Aharon Ben-Tal, Laurent El Ghaoui, and Arkadi Nemirovski. *Robust optimization*. Princeton University Press, 2009.

Dimitris Bertsimas and Melvyn Sim. The price of robustness. *Operations Research*, 52(1): 35–53, 2004.

Dimitris Bertsimas and Aurélie Thiele. A robust optimization approach to inventory theory. *Operations Research*, 54(1):150–168, 2006.

Dimitris Bertsimas, Dessislava Pachamanova, and Melvyn Sim. Robust linear optimization under general norms. *Operations Research Letters*, 32(6):510–516, 2004.

Dimitris Bertsimas, David B Brown, and Constantine Caramanis. Theory and applications of robust optimization. *SIAM review*, 53(3):464–501, 2011.

Dimitris Bertsimas, Vishal Gupta, and Ioannis Ch Paschalidis. Data-driven estimation in equilibrium using inverse optimization. *Mathematical Programming*, 153(2):595–633, 2015.

Dimitris Bertsimas, Vishal Gupta, and Nathan Kallus. Data-driven robust optimization. *Mathematical Programming*, 2017. In press.

John R Birge, Ali Hortaçsu, and J Michael Pavlin. Inverse optimization for the recovery of market structure from market outcomes: An application to the MISO electricity market. *Operations Research*, 2017. In press.

Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.

Peter Brucker and Natalia V Shakhlevich. Inverse scheduling with maximum lateness objective. *Journal of Scheduling*, 12(5):475–488, 2009.

Michal Černỳ and Milan Hladík. Inverse optimization: towards the optimal parameter set of inverse LP with interval coefficients. *Central European Journal of Operations Research*, 24 (3):747–762, 2016.

Timothy CY Chan, Tim Craig, Taewoo Lee, and Michael B Sharpe. Generalized inverse multiobjective optimization with application to cancer therapy. *Operations Research*, 62(3): 680–695, 2014.

Timothy CY Chan, Taewoo Lee, and Daria Terekhov. Goodness-of-fit in inverse optimization. *arXiv preprint arXiv:1511.04650*, 2017.

André Chassein and Marc Goerigk. Variable-sized uncertainty and inverse problems in robust optimization. *arXiv preprint arXiv:1606.07380*, 2016.

Joseph YJ Chow and Will W Recker. Inverse optimization with endogenous arrival time constraints to calibrate the household activity pattern problem. *Transportation Research Part B: Methodological*, 46(3):463–479, 2012.

Erick Delage and Yinyu Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, 2010.

S. Dempe and S. Lohse. Inverse linear programming. In A. Seeger, editor, *Recent Advances in Optimization*, pages 19–28. Springer, 2006.

Virginie Gabrel, Cécile Murat, and Aurélie Thiele. Recent advances in robust optimization: An overview. *European Journal of Operational Research*, 235(3):471–483, 2014.

Donald Goldfarb and Garud Iyengar. Robust portfolio selection problems. *Mathematics of Operations Research*, 28(1):1–38, 2003.

Çiğdem Güler and Horst W Hamacher. Capacity inverse minimum cost flow problem. *Journal of Combinatorial Optimization*, 19(1):43–59, 2010.

Clemens Heuberger. Inverse combinatorial optimization: A survey on problems, methods, and results. *Journal of Combinatorial Optimization*, 8(3):329–361, 2004.

Reiner Horst, Panos M Pardalos, and Nguyen Van Thoai. *Introduction to global optimization*, volume 48 of *Nonconvex Optimization and Its Applications*. Springer US, 2nd edition, 2000.

Garud Iyengar and Wanmo Kang. Inverse conic programming with applications. *Operations Research Letters*, 33(3):319–330, 2005.

Ruiwei Jiang, Jianhui Wang, and Yongpei Guan. Robust unit commitment with wind power and pumped storage hydro. *IEEE Transactions on Power Systems*, 27(2):800–810, 2012.

Arezou Keshavarz, Yang Wang, and Stephen Boyd. Imputing a convex objective function. In *Intelligent Control (ISIC), 2011 IEEE International Symposium on*, pages 613–619. IEEE, 2011.

Panos Kouvelis and Gang Yu. *Robust discrete optimization and its applications*, volume 14 of *Nonconvex Optimization and Its Applications*. Springer Science & Business Media, 1997.

Javier Saez-Gallego and Juan Miguel Morales. Short-term forecasting of price-responsive loads using inverse optimization. *IEEE Transactions on Smart Grid*, 2017. In press.

Javier Saez-Gallego, Juan M Morales, Marco Zugno, and Henrik Madsen. A data-driven bidding model for a cluster of price-responsive consumers of electricity. *IEEE Transactions on Power Systems*, 31(6):5001–5011, 2016.

Andrew J Schaefer. Inverse integer programming. *Optimization Letters*, 3(4):483–489, 2009.

Allen L Soyster. Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research*, 21(5):1154–1157, 1973.

Jan Unkelbach, Timothy CY Chan, and Thomas Bortfeld. Accounting for range uncertainties in the optimization of intensity modulated proton therapy. *Physics in Medicine and Biology*, 52(10):2755–2773, 2007.

Zhiwei Xu, Tianhu Deng, Zechun Hu, Yonghua Song, and Jianhui Wang. Data-driven pricing strategy for demand-side resource aggregators. *IEEE Transactions on Smart Grid*, 2016. In press.

Appendix A

Proofs

Proof of Proposition 1: By assumption on $\hat{\mathbf{x}}$, there exists $\hat{j} \in J$ such that $\hat{x}_{\hat{j}} \neq 0$. Then, it is easy to check that the following is a feasible solution for formulation (2.3):

$$\boldsymbol{\pi} = \mathbf{e}_{\hat{i}}, \quad \text{for some } \hat{i} \in I, \tag{A.1}$$

$$a_{ij} = \begin{cases} \frac{b_i}{\hat{x}_j} & \text{if } j = \hat{j}, \\ 0 & \text{otherwise}, \quad \forall i \in I, \end{cases} \tag{A.2}$$

$$\mathbf{c} = \mathbf{a}_{\hat{i}}. \tag{A.3}$$

□

Proof of Proposition 5: ($\Rightarrow$) Assume that $\sum_{j \in J} a_{\hat{i}j}\hat{x}_j < b_{\hat{i}}$ for some $\hat{i} \in I$. Constraints (3.3c) and (3.3d) imply $u_{\hat{i}j} \geq 0$ for all $j \in J_{\hat{i}}$. It follows that $\sum_{j \in J} a_{\hat{i}j}\hat{x}_j - \sum_{j \in J_{\hat{i}}} u_{\hat{i}j} < b_{\hat{i}}$, meaning that the constraint (3.3e) is violated for $\hat{i}$.

($\Leftarrow$) By assumption on $\hat{\mathbf{x}}$, there exists $\hat{i} \in I$ and $\hat{j} \in J_{\hat{i}}$ such that $\hat{x}_{\hat{j}} \neq 0$. Then, it is easy to check that the following is a feasible solution for formulation (3.3):

$$\boldsymbol{\pi} = \mathbf{e}_{\hat{i}}, \tag{A.4}$$

$$(\lambda_{ij}, \mu_{ij}) = \begin{cases} (1,0) & \text{if } j \in J_i \colon \hat{x}_j \leq 0, i = \hat{i}, \\ (0,1) & \text{if } j \in J_i \colon \hat{x}_j > 0, i = \hat{i}, \\ (0,0) & \text{otherwise}, \end{cases} \tag{A.5}$$

$$\alpha_{ij} = \begin{cases} \frac{\sum_{k \in J} a_{ik}\hat{x}_k - b_i}{|\hat{x}_j|} & \text{if } j = \hat{j}, i = \hat{i}, \\ 0 & \text{otherwise}, \end{cases} \tag{A.6}$$

$$u_{ij} = \alpha_{ij}|\hat{x}_j|, \quad \forall j \in J_i, i \in I, \tag{A.7}$$

$$c_j = \begin{cases} a_{\hat{i}j} - \text{sgn}(\hat{x}_j)\alpha_{\hat{i}j} & \text{if } j \in J_{\hat{i}}, \\ a_{\hat{i}j} & \text{if } j \in J \setminus J_{\hat{i}}. \end{cases} \tag{A.8}$$

□

Lemma 5. *Let $s_{ij} = \lambda_{ij} - \mu_{ij}$ for all $j \in J_i, i \in I$. If $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ satisfies (3.3i) and non-negativity, then $s_{ij} \in [-\pi_i, \pi_i]$ for all $j \in J_i, i \in I$. Conversely, if $s_{ij} \in [-\pi_i, \pi_i]$ for all $j \in J_i, i \in I$, then there exists $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ satisfying (3.3i), non-negativity, and $s_{ij} = \lambda_{ij} - \mu_{ij}$ for all $j \in J_i, i \in I$.*

Proof: To prove the first statement, note that since $\lambda_{ij} + \mu_{ij} = \pi_i$ and $\lambda_{ij}, \mu_{ij}, \pi_i \geq 0$, it follows that $\lambda_{ij} \leq \pi_i$, for all $j \in J_i, i \in I$. Since $\mu_{ij} \geq 0$, it further follows that $\lambda_{ij} - \mu_{ij} \leq \pi_i$, i.e., $s_{ij} \leq \pi_i$. The proof of $-\pi_i \leq s_{ij}$ is similar.

To prove the second statement, we will construct $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ satisfying the required conditions. For all $j \in J_i, i \in I$, let $\lambda_{ij} = s_{ij} + \frac{\pi_i - s_{ij}}{2}$, $\mu_{ij} = \frac{\pi_i - s_{ij}}{2}$ if $s_{ij} \geq 0$, and let $\lambda_{ij} = \frac{\pi_i + s_{ij}}{2}$, $\mu_{ij} = -s_{ij} + \frac{\pi_i + s_{ij}}{2}$ otherwise. □

Proof of Proposition 7: (⇒) The proof of this implication is divided into two cases. First, we show that feasibility of (3.28) implies that $\sum_{j \in J} a_{ij}\hat{x}_j \geq b_i$ for all $i \in I$. We have

$$\sum_{j \in J} a_{ij}\hat{x}_j \geq \sum_{j \in J} a_{ij}\hat{x}_j - (\sum_{j \in J_i} y_{ij} + \Gamma_i z_i) \geq b_i, \quad \forall i \in I,$$

where the first inequality is implied by non-negativity of $\boldsymbol{\Gamma}, \mathbf{y}, \mathbf{z}$, and the second inequality is constraint (3.28f).

Second, assume that $\hat{I} = \varnothing$. We will show that formulation (3.28) is infeasible. Since $\hat{I} = \varnothing$, we have

$$0 < \sum_{j \in J} a_{ij}\hat{x}_j - b_i - \sum_{j \in J_i} \alpha_{ij}|\hat{x}_j|, \quad \forall i \in I, \tag{A.9}$$

$$0 < \sum_{j \in J} a_{ij}\hat{x}_j - b_i - \sum_{j \in J_i} \alpha_{ij}\hat{x}_j, \quad \forall i \in I, \tag{A.10}$$

$$\begin{aligned} 0 &< \sum_{i \in I} \pi_i(\sum_{j \in J} a_{ij}\hat{x}_j - b_i - \sum_{j \in J_i} \alpha_{ij}\hat{x}_j) && \text{(A.11)} \\ &\leq \sum_{i \in I} \pi_i(\sum_{j \in J} a_{ij}\hat{x}_j - b_i) - \sum_{i \in I}\sum_{j \in J_i} \varphi_{ij}\alpha_{ij}\hat{x}_j && \text{(A.12)} \\ &\leq \sum_{i \in I} \pi_i(\sum_{j \in J} a_{ij}\hat{x}_j - b_i) + \sum_{i \in I}\sum_{j \in J_i} (\lambda_{ij} - \mu_{ij})\alpha_{ij}\hat{x}_j, && \text{(A.13)} \end{aligned}$$

where the third inequality is implied by $\mathbf{e}^\intercal\boldsymbol{\pi} = 1, \boldsymbol{\pi} \geq \mathbf{0}$; the fourth inequality is implied by constraint (3.28k); and the fifth inequality is implied by $-\varphi_{ij} \leq (\lambda_{ij} - \mu_{ij})$, itself implied by constraints (3.28l) and (3.28n). Now substituting (3.28j) into (3.28b),

$$\sum_{i \in I} \pi_i(\sum_{j \in J} a_{ij}\hat{x}_j - b_i) + \sum_{i \in I}\sum_{j \in J_i} (\lambda_{ij} - \mu_{ij})\alpha_{ij}\hat{x}_j = 0, \tag{A.14}$$

which is a contradiction.

(⇐) Assume that $\sum_{j \in J} a_{ij}\hat{x}_j \geq b_i$ for all $i \in I$ and $\hat{I} \neq \varnothing$. Let $\hat{i} \in \hat{I}$ be an arbitrary index.

Then, it can be checked that the following is a feasible solution for formulation (3.28):

$$
\begin{aligned}
&\boldsymbol{\pi} = \mathbf{e}_{\hat{i}},\\
&\varphi_{ij^i_k} = \begin{cases} \pi_i & \text{if } k = 1,\ldots,\lfloor\Gamma_i\rfloor, i = \hat{i},\\ (\Gamma_i - \lfloor\Gamma_i\rfloor)\pi_i & \text{if } k = \lceil\Gamma_i\rceil, i = \hat{i},\\ 0 & \text{otherwise,}\end{cases}\\
&(\lambda_{ij},\mu_{ij}) = \begin{cases} (\varphi_{ij},0) & \text{if } j \in J_i\colon \hat{x}_j \le 0, i = \hat{i},\\ (0,\varphi_{ij}) & \text{if } j \in J_i\colon \hat{x}_j > 0, i = \hat{i},\\ (0,0) & \text{otherwise,}\end{cases}\\
&\Gamma_i = \begin{cases} \underline{\Gamma}_i & \text{if } i = \hat{i},\\ 0 & \text{if } i \ne \hat{i}, i \in I,\end{cases}\\
&u_{ij} = \alpha_{ij}|\hat{x}_j|, \quad \forall j \in J_i, i \in I,\\
&y_{ij} = \begin{cases} \max\{u_{ij} - z_i, 0\} & \text{if } j \in J_i, i = \hat{i},\\ 0 & \text{otherwise,}\end{cases}\\
&z_i = \begin{cases} \alpha_{ij^i_{\lceil\Gamma_i\rceil}}|\hat{x}_{j^i_{\lceil\Gamma_i\rceil}}| & \text{if } i = \hat{i},\\ \max_{j\in J_i}\{\alpha_{ij}|\hat{x}_j|\} & \text{if } i \ne \hat{i}, i \in I,\end{cases}\\
&c^i_j = \begin{cases} a_{ij} - \operatorname{sgn}(\hat{x}_j)\alpha_{ij} & \text{if } j = j^i_k, k = 1,\ldots,\lfloor\Gamma_i\rfloor,\\ a_{ij} - \operatorname{sgn}(\hat{x}_j)\alpha_{ij}(\Gamma_i - \lfloor\Gamma_i\rfloor) & \text{if } j = j^i_{\lfloor\Gamma_i\rfloor+1},\\ a_{ij} & \text{otherwise,} \quad \forall i \in \hat{I}.\end{cases}
\end{aligned}
$$

□

Lemma 6. *The solution* $(\mathbf{u}^*_i, \mathbf{y}^*_i, z^*_i)$ *defined in* (3.30) *is an optimal solution to* (3.29).

Proof: By inspection, it is clear that an optimal solution has $u^*_{ij} = \alpha_{ij}|\hat{x}_j|$ and $y^*_{ij} = \max\{u^*_{ij} - z_i, 0\}$ for a given z_i. We will prove by contradiction that $z^*_i = \alpha_{ij^i_{\lceil\Gamma_i\rceil}}|\hat{x}_{j^i_{\lceil\Gamma_i\rceil}}|$ and corresponding $\mathbf{y}^*_i$ are optimal. First suppose $\tilde{z}_i \ne z^*_i$ and corresponding $\tilde{\mathbf{y}}_i$ is a better solution, i.e.,

$$
\begin{aligned}
&\sum_{j\in J_i}\tilde{y}_{ij} + \Gamma_i\tilde{z}_i < \sum_{j\in J_i} y^*_{ij} + \Gamma_i z^*_i\\
\Leftrightarrow\quad &\Gamma_i(\tilde{z}_i - z^*_i) < \sum_{j\in J_i}\max\{\alpha_{ij}|\hat{x}_j| - z^*_i, 0\} - \sum_{j\in J_i}\max\{\alpha_{ij}|\hat{x}_j| - \tilde{z}_i, 0\}\\
\Leftrightarrow\quad &\Gamma_i(\tilde{z}_i - z^*_i) < \sum_{j\in J^1_i}\max\{\alpha_{ij}|\hat{x}_j| - z^*_i, 0\} + \sum_{j\in J^2_i}\max\{\alpha_{ij}|\hat{x}_j| - z^*_i, 0\} - \sum_{j\in J^1_i}(\alpha_{ij}|\hat{x}_j| - \tilde{z}_i),
\end{aligned}
\tag{A.15}
$$

where $J^1_i = \{j \in J_i\colon \alpha_{ij}|\hat{x}_j| > \tilde{z}_i\}$ and $J^2_i = \{j \in J_i\colon \tilde{z}_i \ge \alpha_{ij}|\hat{x}_j|\}$. We also define $|J^1_i| = \hat{k}$, meaning that $\hat{k}$ is the number of elements in $\{\alpha_{ij}|\hat{x}_j|\}_{j\in J_i}$ that are strictly greater than $\tilde{z}_i$.

We now distinguish two cases and in each case we will derive a contradiction. First consider the case $\tilde{z}_i > z_i^*$. We note that

$$\alpha_{ij^i_{\hat{k}}}|\hat{x}_{j^i_{\hat{k}}}| > \tilde{z}_i > \alpha_{ij^i_{\lceil \Gamma_i \rceil}}|\hat{x}_{j^i_{\lceil \Gamma_i \rceil}}|, \tag{A.16}$$

where the first inequality is implied by the definitions of J_i^1 and $\hat{k}$, and the second inequality is implied by the case distinction and the definition of z_i^*. Because the inequalities are strict, it follows that $\hat{k} \leq \lceil \Gamma_i \rceil - 1$. Now we rewrite (A.15) as

$$\begin{aligned}
&\Gamma_i(\tilde{z}_i - z_i^*) < \sum_{j \in J_i^1} (\alpha_{ij}|\hat{x}_j| - z_i^*) + \sum_{k=\hat{k}+1}^{\lceil \Gamma_i \rceil - 1} (\alpha_{ij^i_k}|\hat{x}_{j^i_k}| - z_i^*) - \sum_{j \in J_i^1} (\alpha_{ij}|\hat{x}_j| - \tilde{z}_i) \\
\Leftrightarrow\ &\Gamma_i(\tilde{z}_i - z_i^*) < \hat{k}(\tilde{z}_i - z_i^*) + \sum_{k=\hat{k}+1}^{\lceil \Gamma_i \rceil - 1} (\alpha_{ij^i_k}|\hat{x}_{j^i_k}| - z_i^*) \\
\Leftrightarrow\ &(\Gamma_i - \hat{k})(\tilde{z}_i - z_i^*) < \sum_{k=\hat{k}+1}^{\lceil \Gamma_i \rceil - 1} (\alpha_{ij^i_k}|\hat{x}_{j^i_k}| - z_i^*).
\end{aligned} \tag{A.17}$$

We note here that

$$\tilde{z}_i \geq \alpha_{ij^i_k}|\hat{x}_{j^i_k}|,\ \forall k = \hat{k}+1, \ldots, \lceil \Gamma_i \rceil - 1,$$

where the case $k = \hat{k} + 1$ follows from the definitions of $\hat{k}$ and J_i^2, and the case $k = \lceil \Gamma_i \rceil - 1$ follows from $\tilde{z}_i > z_i^* = \alpha_{ij^i_{\lceil \Gamma_i \rceil}}|\hat{x}_{j^i_{\lceil \Gamma_i \rceil}}|$. We use this observation to deduce a final inequality,

$$\sum_{k=\hat{k}+1}^{\lceil \Gamma_i \rceil - 1} (\alpha_{ij^i_k}|\hat{x}_{j^i_k}| - z_i^*) \leq (\lceil \Gamma_i \rceil - 1 - \hat{k})(\tilde{z}_i - z_i^*),$$

which contradicts (A.17), since $\Gamma_i > \lceil \Gamma_i \rceil - 1$.

Second consider the case $z_i^* > \tilde{z}_i$. We can rewrite (A.15) as

$$\begin{aligned}
&\Gamma_i(z_i^* - \tilde{z}_i) > -\sum_{j \in J_i^1} \max\{\alpha_{ij}|\hat{x}_j| - z_i^*, 0\} - \sum_{j \in J_i^2} \max\{\alpha_{ij}|\hat{x}_j| - z_i^*, 0\} + \sum_{j \in J_i^1} (\alpha_{ij}|\hat{x}_j| - \tilde{z}_i) \\
\Leftrightarrow\ &\Gamma_i(z_i^* - \tilde{z}_i) > -\sum_{j \in J_i^1} \max\{\alpha_{ij}|\hat{x}_j| - z_i^*, 0\} + \sum_{j \in J_i^1} (\alpha_{ij}|\hat{x}_j| - \tilde{z}_i).
\end{aligned}$$

Because $z_i^* = \alpha_{ij^i_{\lceil \Gamma_i \rceil}}|\hat{x}_{j^i_{\lceil \Gamma_i \rceil}}|$, we can write

$$\Gamma_i(z_i^* - \tilde{z}_i) > -\sum_{k=1}^{\lceil \Gamma_i \rceil - 1} (\alpha_{ij^i_k}|\hat{x}_{j^i_k}| - z_i^*) + \sum_{k=1}^{\hat{k}} (\alpha_{ij^i_k}|\hat{x}_{j^i_k}| - \tilde{z}_i).$$

Because $z_i^* = \alpha_{ij^i_{\lceil\Gamma_i\rceil}} |\hat{x}_{j^i_{\lceil\Gamma_i\rceil}}| > \tilde{z}_i$ and $\hat{k} = |J_i^1|$ imply that $\lceil\Gamma_i\rceil \leq \hat{k}$, we can write

$$\begin{aligned}
& \Gamma_i(z_i^* - \tilde{z}_i) > \sum_{k=\lceil\Gamma_i\rceil}^{\hat{k}} \alpha_{ij^i_k} |\hat{x}_{j^i_k}| + (\lceil\Gamma_i\rceil - 1) z_i^* - \hat{k}\tilde{z}_i \\
\Leftrightarrow \quad & \Gamma_i z_i^* - \Gamma_i \tilde{z}_i > \sum_{k=\lceil\Gamma_i\rceil}^{\hat{k}} \alpha_{ij^i_k} |\hat{x}_{j^i_k}| + \lceil\Gamma_i\rceil z_i^* - z_i^* - \hat{k}\tilde{z}_i \\
\Leftrightarrow \quad & \hat{k}\tilde{z}_i - \Gamma_i \tilde{z}_i > \sum_{k=\lceil\Gamma_i\rceil}^{\hat{k}} \alpha_{ij^i_k} |\hat{x}_{j^i_k}| + \lceil\Gamma_i\rceil z_i^* - z_i^* - \Gamma_i z_i^*.
\end{aligned}$$

We substitute $z_i^* = \alpha_{ij^i_{\lceil\Gamma_i\rceil}} |\hat{x}_{j^i_{\lceil\Gamma_i\rceil}}|$ and rearrange to obtain

$$(\hat{k} - \Gamma_i)\tilde{z}_i > (\lceil\Gamma_i\rceil - \Gamma_i)\alpha_{ij^i_{\lceil\Gamma_i\rceil}} |\hat{x}_{j^i_{\lceil\Gamma_i\rceil}}| + \sum_{k=\lceil\Gamma_i\rceil+1}^{\hat{k}} \alpha_{ij^i_k} |\hat{x}_{j^i_k}|. \tag{A.18}$$

Because $\lceil\Gamma_i\rceil \leq \hat{k}$ and the definition of J_i^1 imply that $\alpha_{ij^i_k} |\hat{x}_{j^i_k}| > \tilde{z}_i, \ \forall k = \lceil\Gamma_i\rceil, \ldots, \hat{k}$, we can write

$$(\lceil\Gamma_i\rceil - \Gamma_i)\alpha_{ij^i_{\lceil\Gamma_i\rceil}} |\hat{x}_{j^i_{\lceil\Gamma_i\rceil}}| + \sum_{k=\lceil\Gamma_i\rceil+1}^{\hat{k}} \alpha_{ij^i_k} |\hat{x}_{j^i_k}| > (\hat{k} - \Gamma_i)\tilde{z}_i,$$

which contradicts (A.18). □